Birte Snilstveit
Jennifer Stevenson
Daniel Phillips
Martina Vojtkova
Emma Gallagher
Tanja Schmidt
Hannah Jobse
Maisie Geelen
Maria Grazia Pastorello
et al.

Interventions for improving learning outcomes and access to education in low- and middle-income countries

A systematic review

December 2015

Systematic Review 24

Education

International Initiative for Impact Evaluation

About 3ie

The International Initiative for Impact Evaluation (3ie) is an international grant-making NGO promoting evidence-informed development policies and programmes. We are the global leader in funding, producing and synthesising high-quality evidence of what works, for whom, why and at what cost. We believe that better and policy-relevant evidence will make development more effective and improve people's lives.

3ie in-house systematic reviews

3ie systematic reviews appraise and synthesise the available high-quality evidence on the effectiveness of social and economic development interventions in low- and middle-income countries. These reviews conducted in-house by 3ie researchers follow scientifically recognised review methods. This review was peer-reviewed by members of an expert advisory group and three anonymous external reviewers. 3ie is providing leadership in demonstrating rigorous and innovative review methodologies, such as using theory-based approaches suited to inform policy and programming in the dynamic contexts and challenges of low- and middle-income countries.

About this review

This review, *Interventions for improving learning outcomes and access to education in low- and middle- income countries: a systematic review,* is available on the 3ie website.

All content is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or its board of commissioners. Any errors are also the sole responsibility of the authors. 3ie is publishing this report as received from the authors; it has been formatted to 3ie style. Comments or queries should be directed to the corresponding author, Birte Snilstveit, bsnilstveit@3ieimpact.org.

Major funding for this systematic review was provided by UK aid, the Bill & Melinda Gates Foundation and the Hewlett Foundation. A complete listing of all of 3ie's donors can be found on the 3ie website.

Suggested citation: Snilstveit, B, Stevenson, J, Phillips, D, Vojtkova, M, Gallagher, E, Schmidt, T, Jobse, H, Geelen, M, Pastorello, M, and Eyers, J, 2015. *Interventions for improving learning outcomes and access to education in low- and middle- income countries: a systematic review*, *3ie Systematic Review 24*. London: International Initiative for Impact Evaluation (3ie).

3ie systematic review executive editors: Edoardo Masset and Beryl Leach
Managing editor: Deepthy Menon
Technical editor: Radhika Menon
Production manager: Pradeep Singh
Cover design: John F McGill and Akarsh Gupta

Interventions for improving learning outcomes and access to education in low- and middle- income countries: a systematic review

Birte Snilstveit
International Initiative for Impact Evaluation (3ie)

Jennifer Stevenson
3ie

Daniel Phillips
3ie

Martina Vojtkova
NatCen

Emma Gallagher
3ie

Tanja Schmidt
3ie

Hannah Jobse
3ie

Maisie Geelen
3ie

Maria Grazia Pastorello
3ie

John Eyers
3ie

3ie Systematic Review 24

December 2015

International Initiative for Impact Evaluation

Acknowledgements

We are grateful to 3ie colleagues who provided inputs at various stages of the review, particularly Emmanuel Jimenez, Philip Davies, Howard White, Edoardo Masset, Hugh Waddington, Shari Krishnaratne, Radhika Menon and Beryl Leach. Ariel Aloe, Christopher Thompson and Shannon Shisler provided outstanding advice and assistance with data extraction and statistical analysis. Thanks to Josh Polanin for invaluable advice and support just at the right time and to Paul Fenton, Nicholas Rowland and Anna Fox for excellent research assistance. We would also like to thank the members of our advisory board Rukmini Banerji, Chris Berry, Elizabeth King, Juan Miguel Luz, Penelope Bender, Robert Slavin and Stephen Taylor, as well as three external reviewers for their invaluable comments and suggestions. Thanks to the organisers and participants at events where previous versions of this review have been presented. We would like to especially thank Jeff Waage for organising a round table at the London International Development Centre (LIDC) for the presentation of a preliminary draft of the review, as well as all those who participated and provided comments. Special thanks also to Marie Gaarder and colleagues at the World Bank for organising and attending a seminar where an earlier version of the review was presented. Finally, thanks to all the authors who shared unpublished and forthcoming papers and any additional data that helped improve the review.

Summary

Background

Substantial progress has been made in improving access to education in low- and middle income countries (L&MICs). However, the progress has been uneven and several challenges still need to be addressed. In 2013, 69.3 million children of primary school age were still out of school (UNESCO, 2014a). More than half of these children (39.1 million) were in Sub-Saharan Africa (UNESCO, 2014b). Moreover, studies measuring learning outcomes among school children across L&MICs find consistently low levels of learning, with hundreds of millions of children leaving school without basic numeracy and literacy skills (Pritchett, 2013; Robinson, 2011; UNESCO, 2015). Therefore, while education interventions have traditionally focussed on getting children into school, researchers and decision makers are now focusing their attention on efforts to improve learning for all (Pritchett, 2013; Robinson, 2011). This is increasingly reflected in the education policy of major donor agencies (e.g. DFID, 2013; World Bank, 2011) and in the recently adopted Sustainable Development Goals (SDGs). While the Millennium Development Goal on education (Goal 2) was to achieve universal primary education, the SDG on education (Goal 4) is to ensure inclusive and equitable quality education and promote lifelong learning (UNGA, 2015). In this context, evidence on the effects of education interventions is needed for informing decisions about how limited funding can be best used to achieve the ambitious goal of achieving quality education for all children. This comprehensive systematic review synthesises the findings of 238 studies evaluating the effects of a range of different education programmes in 52 L&MICs.

Objectives

The primary objective of this review was to identify, assess and synthesise evidence on the effects of education interventions on children's access to education and learning in L&MICs. We also aimed to assess how education interventions affect different sub-groups of participants and address questions related to context, process and implementation.

To address these objectives, we answered the following questions:

1a) What are the effects of different education interventions on enrolment, attendance, dropout rates, completion and learning outcomes for primary and secondary school children in low- and middle-income countries?
1b) Do the effects differ between sub-groups of participants (according to gender, age, urban or rural location, or socio-economic status)?
2a) What intervention and implementation features are associated with relative success and failure in improving educational outcomes?
2b) What are the contextual barriers to, and facilitators of, the effectiveness of educational interventions?

Study selection criteria

We included studies of primary and secondary school children in mainstream education in L&MICs that were published between 1990 and June 2015. To be included in this systematic review, studies had to use an experimental or quasi-experimental study design and measure school participation (enrolment, attendance, drop-out, completion) or learning outcomes (cognitive skills, maths, language arts and composite score).

We aimed to include a comprehensive range of commonly implemented education interventions designed to address one or more barriers to school participation and learning: school feeding, school-based health, merit-based scholarships, provision of information, reduction of user fees, cash transfers, structured pedagogy, computer assisted learning, remedial education, provision of materials, extra time in school, tracking of students, grade retention, new schools and infrastructure, teacher incentives, teacher hiring, teacher training, diagnostic feedback, school-based management, community-based monitoring, and finally, private-public partnerships.

To address questions 2a and 2b, we also included qualitative studies, descriptive quantitative studies, process evaluations and project documents linked to the interventions that were evaluated in the included experimental and quasi-experimental studies.

Search

We searched through a broad range of electronic academic databases, internet search engines, websites and theses collections. The searches included both general social science sources as well as education specific sources of published and unpublished literature. All searches were updated in June 2015. We screened the titles and abstracts of over 78,000 papers, the majority of which were irrelevant to the topic. We screened the full text of 2042 papers for inclusion. After a final screening by at least two authors, we included 420 papers corresponding to 238 different studies. These studies were assessing 216 different programmes.

Included studies

The included studies cover programmes across 52 L&MICs. This include 59 studies from Sub-Saharan Africa, 38 studies from East Asia & the Pacific, 87 from Latin America & the Caribbean; 51 from South Asia, two from Middle Eastern & North Africa and one from Europe & CIS. Based on reported sample sizes we estimate the studies include data from over 16 million children.

Data collection and analysis

Two independent reviewers assessed the full text papers against the inclusion criteria. Disagreements related to inclusion were resolved through discussion or by a third reviewer, if necessary. We critically appraised the included studies using standard appraisal tools and collected data on programme design, process, implementation and contextual factors. We extracted effect size data from included studies, calculating standardised mean differences (SMDs), standard errors and 95 per cent confidence intervals for all studies. All studies were coded by one reviewer, with a second person checking the accuracy of data extracted, adding additional data as necessary.

We developed a conceptual framework for the review, separating interventions according to the main barriers they aim to address. We used this to organise the analysis in our review and separate interventions into child, household, school, teacher and systems levels.

Studies were then synthesised using random effects meta-analysis, estimating average effects of different education interventions and associated heterogeneity. We also conducted sensitivity analysis and analysis of publication bias. For the synthesis of qualitative evidence (review questions 2a and 2b), we used a thematic approach (Thomas & Harden, 2008), organising themes according to the intervention programme theory. For many intervention areas the analyses we were able to do were however limited by the availability of evidence.

Results

Our results suggest most interventions have an overall positive effect on children who were beneficiaries compared to children who did not receive these interventions, although for some programmes average effects are relatively small. As expected, the intervention type that produces the largest effect differs depending on which outcomes we look at. Among the various interventions that aim to improve access to education, we find that cash transfer programmes have the most substantial and consistent beneficial effects on school participation. Effects range from 0.11 SMD (95% confidence interval (CI), [0.07, 0.15]) for enrolment, to 0.13 SMD, 95% CI [0.08, 0.18] for attendance, with effects on dropout and completion of a similar magnitude. All of these estimates are based on a relatively large number of studies, with no less than 16 comparisons included in any single meta-analysis. While the results are relatively robust, there is still considerable heterogeneity and we observe effects that are both substantially larger and substantially smaller than the average effects. Cash transfers do not however appear to lead to any improvement in learning outcomes. Other interventions that may be promising for improving school participation outcomes include community-based monitoring, new schools and infrastructure and school feeding.

We find that structured pedagogy programmes have the largest and most consistent positive average effects on learning outcomes. Typically, structured pedagogy interventions include development of new content focused on a particular topic, materials for students and teachers, and short term training courses for teachers in delivering the new content. The meta-analysis for language arts outcomes includes effects from eighteen studies. Many of these studies are large scale randomised controlled trials (RCTs), and the meta-analysis shows an overall effect of 0.23 (95% CI, [0.13, 0.34]). The effect on maths test scores is slightly smaller in magnitude at 0.14 (95% CI, [0.08, 0.20]), but it is still the largest and most consistent effect observed for maths test scores across the review. Other interventions that may be promising for improving learning outcomes include merit-based scholarships, school feeding, extra time in school and remedial education.

For several intervention areas the effects are relatively small in magnitude as compared to the results summarised above. We also find zero or small negative effects for some interventions. For instance, the effects of school based management range from -0.01, for language test scores and composite test scores (95% CI [-0.07, 0.05]; 95% CI [-0.10, 0.08]) to 0.05, 95% CI [0.00, 0.09] for completion. Similarly, for de-worming the average effects on education outcomes range from -0.04, 95% CI [-0.11, 0.02] to 0.05 SMD, 95% CI [-0.02, 0.13] for math test scores. For interventions providing materials we find limited, if any, difference between children receiving interventions and those that do not. Finally, the average effects observed for computer assisted learning also lead us to conclude the effects of computer assisted learning on children's learning may not be beneficial in all contexts.

Implications for research

Through this review we have identified a range of programme areas where there are few or no studies. This includes teacher training programmes, remedial education, school- based health programmes (malaria, de-worming, micronutrients), diagnostic feedback, providing information to parents, tracking students by ability, extension of the school day and different approaches to teacher training and hiring. Some of these interventions appear particularly promising and it may be worth focusing new studies in these areas.

While the included studies use rigorous designs to assess the effects of interventions, most studies do not address other questions comprehensively, such as those relating to how and why interventions work or not, and at what cost. Future studies should use mixed-methods study designs to assess the effects of interventions as well as process, implementation and contextual factors that influence final outcomes. This will help explain heterogeneity in effects that can in turn help inform improvements of future programmes. Finally, studies should include information about costs to allow cost-effectiveness analysis.

Not all studies provide clear and comprehensive reporting of methods and results. Studies will be more useful if they clearly describe all main study constructs, report methods in detail and clearly report the statistical information necessary to calculate standardised effect sizes, including sample sizes, standard errors, standard deviations and intracluster correlation coefficients (ICCs). Without clear reporting of what was studied and how, resources used on expensive studies are wasted. Research funders and publishers may consider making it a requirement that researchers follow reporting guidelines such as CONSORT to improve the value of new research.

Implications for policy and practice

There is relatively strong evidence that cash transfer programmes and structured pedagogy interventions are particularly effective in improving school participation and learning outcomes in most contexts. However, ensuring that all children have access to high quality education and gain the knowledge and skills needed to realise the benefits of education is a complex process. Children are faced with multiple barriers to school participation and learning. It may therefore not be surprising that we observe effects of a relatively small magnitude for many interventions addressing barriers in only one sphere.

Depending on the barriers facing children in specific contexts, it may be necessary to intervene across more than one sphere to improve the chances of seeing substantive improvements in one or more outcomes. The main findings of the review offer some support for this. With the possible exception of school feeding and community based monitoring, programmes that improve school participation do not appear to improve learning outcomes and vice versa.

Improving children's school participation through cash transfer programmes may have a limited effect on learning outcomes if the existing curriulum content, materials and teachers available are not of sufficient quality. Similarly, when we observe substantial improvements of learning outcomes, as in the case of structured pedagogy, we do not typically find similar improvements for school participation outcomes.

The findings for many intervention areas are based on few studies and we also observe substantial variability of effects. Average effects that appear small in magnitude often include examples where programmes have had large and substantively important (positive and negative) effects on children's access to education and learning in some contexts. Therefore the average effects should be interpreted with some caution.

Contents

Summary of findings tables

Child level

School Based Health					
Summary of findings					
Outcomes	No of comparisons (sample size)	Overall effect size (95% CI)	Estimated percentile change compared to control group [95% CI]	Heterogeneity of overall effect	Range of effects [95% CI]
Malaria prevention and control					
Learning outcomes					
Math scores	3 (5,243)	0.16 [-0.08, 0.25]	6.36% [-3.19%, 9.87%]	High	From -0.08 [-0.13, -0.02] to 0.62 [0.50, 0.73]
Language scores	3 (5,148)	0.03 [-0.49, 0.55]	1.20% [-18.79%, 20.88%]	High	From -0.26 [-0.32, -0.20] to 0.56 [0.49, 0.62]
Cognitive skills	3 (5,917)	0.03 [-0.05, 0.12]	1.20% [-1.99%, 4.78%]	Medium	From -0.02 [-0.11, 0.06] to 0.11 [0.04, 0.07]
Secondary outcomes					
Nutrition	4 (10, 857)	0.28 [0.04, 0.52]	11.03% [1.60%, 19.85%]	High	From -0.04 [-0.08, 0.01] to 0.50 [0.33, 0.66]

Deworming interventions					
Learning outcomes					
Maths	2 (1,582)	0.05 [-0.02, 0.13]	1.99% [-0.80%, 5.17%]	Low	From 0.00 [-0.19, 0.2] to 0.06 [-0.02, 0.14]
Language Arts	3 (2,499)	-0.04 [-0.11, 0.02]	1.60% [-4.38%, 0.80%]	Low	From -0.07 [-0.15, 0.01] to 0.00 [-0.16, 0.17]
Cognitive skills	3 (8,980)	0.01 [-0.03, 0.05]	0.40% [1.20%, 1.99%]	Low	From 0.00 [-0.06, 0.05] to 0.02 [-0.03
Access outcomes					
Attendance	4 (58,315)	0.04 [-0.13, 0.21]	1.60% [-5.17, 8.32%]	High	From -0.21 [-0.47, 0.06] to 0.25 [0.14, 0.36]
Secondary outcomes					
Nutrition	2 (1619)	-0.26 [-0.43, -0.10]	-10.26% [-16.64%, -3.98%]	Moderate	From -0.32 [-0.40, -0.24] to -0.13 [-0.38, 0.12]

Micronutrient intervention					
Learning outcomes					
Maths	4 (7,755)	0.06 [0.02, 0.10]		Moderate	From 0.00 [-0.06, 0.07] to 0.10 [0.03, 0.16]
Cognitive skills	2 (2650)	0.01 [-0.03, 0.05]	0.40% [-1.20%, 1.99%]	Low	From 0.00 [-0.06, 0.05] to 0.02 [-0.03, 0.08]
Secondary outcomes					
Nutrition	5 (6,726)	0.15 [0.04, 0.26]		High	From -0.32 [-0.40, -0.24] to 0.18 [0.10, 0.26]

School Feeding					
Summary of findings					
Outcomes	No of comparisons (sample size)	Overall effect size [95% CI]	Estimated percentile change compared to control group [95% CI]	Heterogeneity of overall effect	Range of effects [95% CI]
Learning outcomes					
Maths	10 (189,194)	0.10 [0.00, 0.19]	3.98% [0.00%, 8.71%]	High	From -0.11 [-0.21, -0.01] to 0.43 [0.34, 0.51]
Language Arts	8 (185,842)	0.09 [0.01, 0.17]	3.59% [0.40%, 6.75%]	High	From -0.08 [-0.18, 0.02] to 0.28 [0.19, 0.38]
Composite scores	3 (5,846)	0.14 [-0.04, 0.33]	5.57% [-1.60%, 12.93%]	High	From -0.01 [-0.07, 0.05] to 0.35 [0.09, 0.61]
Cognitive skills	7 (6,106)	0.11 [0.00, 0.22]	4.38% [0.00%, 8.71%]	High	From -0.09 [-0.24, 0.07] to 0.37 [0.21, 0.53]
Access outcomes					
Enrolment	7 (2,025,514)	0.14 [-0.05, 0.33]	5.57% [-1.99%, 12.93%]	High	From -0.01 [-0.06, 0.03] to 0.70 [0.63, 0.77]
Attendance	6 (18582)	0.09 [0.03, 0.16]	3.59% [1.20%, 6.36%]	High	From -0.02 [-0.08, 0.05] to 0.21 [0.11, 0.31]
Dropout	3 (182,345)	-0.06 [-0.15, 0.03]	-2.39% [-5.96%, 1.20%]	High	From 0.00 [-0.06, 0.06] to -0.15 [-0.22, -0.09]
Completion	2 (20,365)	-0.01 [-0.03, 0.01]	-0.40% [-1.20%, 0.40%]	Low	From -0.02 [-0.05, 0.02] to -0.01 [-0.03, 0.01]

Merit-based Scholarships					
Summary of findings					
Outcomes	No of comparisons (sample size)	Overall effect size [95% CI]	Estimated percentile change compared to control group [95% CI]	Heterogeneity of overall effect	Range of effects [95% CI]
Learning outcomes					
Maths	10 (9,111)	0.11 [0.03, 0.20]	4.38% [1.20%, 6.75%]	High	From -0.14 [-0.24, -0.04] to 0.32 [0.21, 0.42]
Language Arts	3 (3,016)	0.04 [-0.07, 0.15]	1.60% [-2.79%, 5.96%]	High	From -0.03 [-0.12, 0.06] to 0.16 [0.07, 0.25]
Composite scores	7 (6,913)	0.10 [0.03, 0.17]	3.98% [1.20%, 6.75%]	Moderate	From 0.02 [-0.10, 0.13] to 0.25 [0.14, 0.37]
Access outcomes					
Attendance	4 (4,467)	0.01 [-0.06, 0.08]	3.98% [-2.39%, 3.19%]	Moderate	From 0.07 [0.15, 0.01] to 0.08 [0.01, 0.15]
Dropout	2 (1,133)	0.04 [-0.11, 0.19]	1.60% [-4.38%, 7.53%]	Moderate	From -0.05 [-0.19, 0.10] to 0.11 [0.01, 0.21]
Completion	2 (425)	0.32 [-0.18, 0.46]	12.55% [-7.14%, 17.72%]	Low	From 0.25 [0.06, 0.44] to 0.39 [0.21, 0.57]
Secondary outcomes					
Teacher attendance	2 (1,065)	0.14 [-0.04, 0.32]	5.57% [-1.60%, 12.55%]	High	From 0.04 [-0.08, 0.17] to 0.23 [0.11, 0.34]

Household level

User Fee Reduction					
Summary of findings					
Outcomes	No of comparisons (sample size)	Overall effect size (95% CI)	Estimated percentile change compared to control group [95% CI]	Heterogeneity of overall effect	Range of effects [95% CI]
Learning outcomes					
Highest education attained	2 (8,463)	0.00 [-0.08, 0.09]	0.00% [-3.19%, 3.59%]	High	From - 0.04, [- 0.10, 0.01] to 0.04 [0.01, 0.08]
Access outcomes					
Enrolment	8 (572,323)	0.03 [-0.01, 0.06]	1.2% [-0.40%, 2.39%]	High	From - 0.06 [-0.10, -0.03] to 0.10 [0.07, 0.13]
Attendance	2 (6,808)	0.01 [-0.13, 0.15]	0.4% [-5.17%, 5.96%]	Moderate	From -0.06 [-0.17, 0.05] to 0.08 [- 0.03, 0.19]
Dropout	4 (58,297)	-0.10 [-0.23, 0.02]	-3.98% [-9.10%, 0.80%]	High	From -0.29 [-0.32, - 0.26] to -0.01 [- 0.04, 0.02]
Completion	4 (35,521)	0.02 [-0.10, 0.15]	0.80% [-3.98%, 5.96%]	High	From -0.08 [-0.19, 0.02] to 0.14 [0.03, 0.25]

Cash Transfers					
Summary of findings					
Outcomes	No of comparisons (sample size)	Overall effect size [95% CI]	Estimated percentile change compared to control group [95% CI]	Heterogeneity of overall effect	Range of effects [95% CI]
Learning outcomes					
Maths	14 (17890)	-0.01 [-0.07, 0.05]	0.4% [-2.79%, 1.99%]	High	From -0.42 [-0.55, 0.29] to 0.14 [0.01, 0.27]
Language Arts	14 (21338)	0.00 [-0.04, 0.04]	0.00% [-1.60%, 1.60%}	Moderate	From -0.1 [-0.16, -0.05] to 0.18 [0.06, 0.30
Composite scores	3 (135372)	0.01 [-0.01, 0.03]	0.40% [-0.40%, 1.20%]	High	From -0.01 [-0.02, 0.00] to 0.03 [0.01 , 0.04]
Cognitive skills	2 (2940)	0.07 [-0.11, 0.25]	2.79% [-4.38%, 9.87%]	High	From -0.02 [-0.12, 0.07] to 0.16 [0.1, 0.22]
Access outcomes					
Enrolment	49 (407169)	0.11 [0.07, 0.15]	4.38% [2.79%, 5.96%]	High	From -0.05, [-0.09, -0.01] to 0.72 [0.62, 0.82]
Attendance	38 (267295)	0.13 [0.08, 0.18]	5.17% [3.19%, 7.14%]	High	From -0.05 [-0.10, 0.00] to 0.54 [0.49, 0.59]
Dropout	16 (169938)	-0.12 [-0.17, -0.07]	-4.78% [-6.75%, 2.79%]	High	From -0.38 [-0.49, 0.27] to 0.06 [0.03, 0.09]
Completion	28 (809704)	0.12 [-0.01, 0.22]	4.78% [-0.40%, 8.71%]	High	From -0.25 [-0.35, -0.15] to 0.96 [0.90, 1.01]

School level

Structured pedagogy					
Summary of findings					
Outcomes	No of comparisons (sample size)	Overall effect size [95% CI]	Estimated percentile change compared to control group [95% CI]	Heterogeneity of overall effect	Range of effects [95% CI]
Learning outcomes					
Maths	18 (56,902)	0.14 [0.08, 0.20]	5.57% [3.19%, 7.93%]	High	From -0.09 [-0.15, -0.03]
Language Arts	21 (48,896)	0.23 [0.13, 0.34]	9.10% [5.17%, 13.31%]	High	From -0.14 [-0.23, -0.05] to 0.90 [0.85, 0.94]
Composite scores	3 (16975)	0.06 [0.03, 0.08]	2.39% [1.20%, 3.19%]	Low	From 0.02 [-0.08, 0.11] to 0.08 [0.04, 0.12]
Cognitive skills	2 (4705)	0.01 [-0.04, 0.07]	0.40% [-1.60%, 2.79%]	Low	From 0.01 [-0.07, 0.09] to 0.02 [-0.07, 0.10]
Access outcomes					
Attendance	5 (27776)	0.01 [-0.02, 0.03]	0.40% [-0.80%, 1.20%]	Moderate	From -0.04 [-0.09, 0.00] to 0.04 [0.01, 0.08]
Completion	2 (512)	0.13 [-0.02, 0.28]	5.17% [-7.93%, 11.03%]	Moderate	From 0.06 [-0.05, 0.17] to 0.22 [0.07, 0.36]

Computer-assisted learning					
Summary of findings					
Outcomes	No of comparisons (sample size)	Overall effect size [95% CI]	Estimated percentile change compared to control group [95% CI]	Heterogeneity of overall effect	Range of effects [95% CI]
Learning outcomes					
Maths	19 (38,382)	0.07 [0.02, 0.11]	2.79% [0.80%, 4.38%]	High	From -0.20 [-0.31, -0.08] to 0.19, [0.11, 0.28]
Language Arts	13 (27,616)	0.01 [-0.08, 0.05]	0.40% [-3.19%, 1.99%]	High	From -0.36 [-0.44, -0.27] to 0.13, [0.06, 0.20]
Composite scores	6 (25,117)	0.01 [-0.04, 0.07]	0.40% [-1.60%, 2.79%]	High	From -0.10 [-0.17, -0.03] to 0.11 [0.02, 0.19]
Access outcomes					
Enrolment	2 (7,062)	-0.04 [-0.11, 0.04]	-1.60% [-4.38%, 1.60%]	High	From -0.08 [-0.14, -0.02] to 0.00 [-0.04, 0.04]
Attendance	2 (10,182)	0.04 [0.00, 0.07]	1.60% [0.00%, 2.79%]	Low	From 0.04 [0.00, 0.08] to 0.04[-0.02, 0.09]
Dropout	2 (35,252)	-0.04 [-0.12, 0.04]	-1.60% [-4.78%, 1.60%]	High	From -0.08 [-0.11, -0.06] to 0.00 [-0.02, 0.02]
Completion	2 (34,783)	0.07 [-0.07, 0.22]	2.79% [-2.79%, 8.71%]	High	From 0.00 [-0.02, 0.02] to 0.15 [0.12, 0.18]

Remedial education					
Summary of findings					
Outcomes	No of comparisons (sample size)	Overall effect size (95% CI)	Estimated percentile chaneg compared to control group [95% CI]	Heterogeneity of overall effect	Range of effects
Learning outcomes					
Maths	6 (56,162)	0.19 [-0.05, 0.44]	7.53% [-1.99%, 17.00%]	High	From 0.04 [0.01, 0.08] to 0.81 [0.77, 0.86]
Language Arts	6 (44,710)	0.16 [-0.08, 0.41]	6.36% [-3.19%, 15.91%]	High	From 0.02 [-0.02, 0.05] to 0.78 [0.73, 0.83]
Composite scores	5 (38,150)	0.22 [-0.09, 0.53]	8.71% [-3.59%, 20.19%]	High	From 0.03 [-0.01, 0.08] to 0.85 [0.80, 0.90]

Providing Materials					
Summary of findings					
Outcomes	No of comparisons (sample size)	Overall effect size [95% CI]	Estimated percentile change compared to control group [95% CI]	Heterogeneity of overall effect	Range of effects [95% CI]
Learning outcomes					
Maths	5 (46,387)	-0.02 [-0.06, 0.02]	-0.80% [-2.39%, 0.80%]	High	From -0.10 [-0.14, -0.06] to 0.02 [-0.01, 0.04]
Language Arts	5 (48,037)	0.00 [-0.02, 0.02]	0.00% [-0.80%, 0.80%]	Moderate	From -0.03l [-0.06, 0.00] to 0.03 [0.00, 0.06]
Composite scores	5 (86,242)	0.01 [-0.01, 0.02]	0.40% [-0.40%, 0.80%]	Low	From -0.02 [-0.06, 0.03] to 0.02 [0.00, 0.05]

New Schools and Infrastructure					
Summary of findings					
Outcomes	No of comparisons (sample size)	Overall effect size (95% CI)	Estimated percentile change compared to control group [95% CI]	Heterogeneity of overall effect	Range of effects [95% CI]
Learning outcomes					
Maths: construction of new schools	2 (15,163)	0.19 [-0.15, 0.53]	7.53% [-5.96%, 20.19%]	High	From 0.01 [-0.02, 0.05] to 0.36 [0.29, 0.43]
Language Arts : construction of new schools	2 (15,174)	0.02 [-0.01, 0.05]	0.80% [-0.40%, 1.99%]	Low	From 0.02 [-0.01, 0.05] to 0.03 [-0.05, 0.10]
Access outcomes					
Enrolment: hygiene infrastructure interventions	4 (181,342)	0.11 [0.01, 0.20]	4.38% [0.40%, 7.93%]	High	From 0.03 [0.01, 0.04]) to 0.23 [0.22, 0.24]
Attendance: hygiene infrastructure interventions	2 (7,772)	0.14 [0.05, 0.24]	5.57% [1.99%, 9.48]	High	From 0.09 [0.05, 0.14] to 0.19 [0.14, 0.23]
Enrolment: construction of new schools	2 (15,446)	0.38 [-0.29, 1.04]	14.80% [-11.41%, 35.08%]	High	From 0.04 [0.01, 0.06] to 0.72 [0.64, 0.79]
Attendance: construction of new schools	2 (15,346)	0.08 [-0.04, 0.19]	3.19% [-1.60%, 7.53%]	High	From 0.02 [-0.01, 0.04] to 0.14 [0.10, 0.17]

Extra Time					
Summary of findings					
Outcomes	No of comparisons (sample size)	Overall effect size [95% CI]	Estimated percentile change compared to control group [95% CI]	Heterogeneity of overall effect	Range of effects [95% CI]
Learning outcomes					
Maths	2 (2,977)	0.09 [-0.04, 0.22]	3.59% [-1.60%, 8.71%]	High	From 0.02 [-0.06, 0.11] to 0.16 [0.08, 0.24]
Language Arts	2 (2,977)	0.19 [0.15, 0.24]	7.53% [5.96%, 9.48%]	Low	From 0.19 [0.14, 0.25] to 0.20 [0.11, 0.28]

Tracking					
Summary of findings					
Outcomes	No of comparisons (sample size)	Overall effect size [95% CI]	Estimated percentile change compared to control group [95% CI]	Heterogeneity of overall effect	Range of effects [95% CI]
Learning outcomes					
Maths	2 (11,472)	0.02 [-0.04, 0.08]	0.80% [-1.60%, 3.19%]	High	From -0.01 [-0.05, 0.03] to 0.05 [0.01, 0.09]
Language Arts	2 (11,472)	0.12 [-0.03, 0.27]	4.78% [-1.20%, 10.64%]	High	From 0.06 [0.01, 0.08] to 0.20 [0.13, 0.26]

Teacher Level

Teacher Incentives					
Summary of findings					
Outcomes	No of comparisons (sample size)	Overall effect size [95% CI]	Estimated percentile change compared to control group [95% CI]	Heterogeneity of overall effect	Range of effects [95% CI]
Learning outcomes					
Maths	11 (80,497)	0.08 [0.02, 0.13]	3.19% [0.80%, 5.17%]	High	From -0.02 [-0.08, 0.04] to 0.30 [0.20, 0.40]
Language Arts	7 (74,161)	0.00 [-0.13, 0.12]	0.00% [-5.17%, 4.78%]	High	From -0.45 [-0.64, 0.26] to 0.11 [-0.06, 0.29]
Composite scores	5 (175,668)	0.02 [-0.02, 0.05]	0.80% [-0.80%, 1.99%]	High	From -0.02 [-0.03, -0.01] to 0.10 [0.04, 0.17]
Access outcomes					
Enrolment	2 (17,456)	0.06 [-0.05, 0.16]	2.39% [-1.99%, 6.36%]	Moderate	From 0.02 [0.00, 0.04] to 0.13 [0.01, 0.24]
Attendance	3 (342,981)	0.01 [-0.04, 0.06]	0.40% [-1.60%, 2.39%]	Low	From 0.00 [-0.14, 0.14] to 0.01 [-0.04, 0.06]
Dropout	4 (59,410)	0.00 [-0.01, 0.01]	0.00% [-0.40%, 0.40%]	Low	From -0.01 [-0.07, 0.04] to 0.01 [-0.02, 0.04]
Completion	4 (13,593)	0.03 [0.00, 0.05]	1.20% [0.00%, 1.99%]	Low	From 0.01 [-0.02, 0.04] to 0.07 [0.01, 0.13]

Secondary outcomes					
Teacher attendance	3 (2,125)	0.07 [-0.05, 0.19]	2.79% [-1.99%, 7.53%]	High	From -0.01 [-0.11, 0.08] to 0.19 [-0.09, 0.30]
Teacher performance: classroom management	3 (4,851)	-0.01 [-0.05, 0.03]	-0.40% [-1.99%, 1.20%]	Low	From -0.02 [-0.08, 0.04] to 0.00[-0.05, 0.03]
Teacher performance: use of materials	3 (2,935)	-0.04 [-0.09, 0.02]	-1.60% [-3.59%, 1.20%]	Low	From -0.10 [-0.22, 0.01] to 0.00 [-0.11, 0.11]
Teacher performance: use of assessment in instruction	2 (2,315)	-0.03 [-0.09, 0.03]	-1.20% [-3.59%, 1.20%]	Low	From -0.05 [0.12,0.02] to 0.01 [-0.10, 0.12]
Teacher performance: preparatory sessions	5 (9,297)	0.07 [0.04, 0.10]	2.79% [1.60%, 1.20%]	Low	From 0.03 [-0.08, 0.14] to 0.09 [0.03, 0.16]
Teacher performance: student engagement	3 (4,851)	-0.01 [-0.04, 0.03]	-0.40% [-1.60%, 1.20%]	Low	From -0.02 [-0.08, 0.04] to 0.00 [-0.05, 0.05]

Teacher Hiring					
Summary of findings					
Outcomes	No of comparisons (sample size)	Overall effect size [95% CI]	Estimated percentile change compared to control group [95% CI]	Heterogeneity of overall effect	Range of effects [95% CI]
Learning outcomes					
Maths	2 (12,835)	0.10 [0.00, 0.20]	3.98% [0.00%, 7.93%]	High	From 0.06, 95% CI [0.03, 0.08] to .16, 95% CI [0.09, 0.23]
Language Arts	2 (12,928)	0.06 [0.03, 0.10]	2.39% [1.20%, 3.98%]	Low	From 0.06, 95% CI [0.03, 0.08] to 0.10, 95% CI [0.03, 0.17]
Composite scores	3 (39,252)	0.06 [-0.01, 0.12]	2.39% [-0.40%, 4.78%]	High	From –0.02, 95% CI [-0.01, 0.04] to 0.14, 95% CI [0.14, 0.21]
Access outcomes					
Completion	3 (87,638)	0.04 [0.01, 0.08]	1.60% [0.40%, 3.19%]	High	From 0.00 [-0.05, 0.04] to 0.06 [0.05, 0.08]

Diagnostic Feedback					
Summary of findings					
Outcomes	No of comparisons (sample size)	Overall effect size (95% CI)	Estimated percentile change compared to control group [95% CI]	Heterogeneity of overall effect	Range of effects [95% CI]
Learning outcomes					
Maths	3 (33,606)	0.01 [-0.01, 0.03]	0.40% [-0.40%, 1.20%]	Low	From 0.00 [-0.04, 0.03] to 0.04 [-0.01, 0.09]
Language Arts	3 (33,586)	0.01 [-0.01, 0.05]	0.40% [-0.40%, 1.99%]	Low	From -0.03 [-0.08, 0.02] to 0.02 [-0.01, 0.05]

Systems Level

School-Based Management					
Summary of findings					
Outcomes	No of comparisons (sample size)	Overall effect size [95% CI]	Estimated percentile change compared to control group [95% CI]	Heterogeneity of overall effect	Range of effects [95% CI]
Learning outcomes					
Maths	21 (22,075)	0.01 [-0.02, 0.05]	0.40% [-0.80%, 1.99%]	Moderate	From -0.11 [-0.24, 0.03] to 0.14 [0.07, 0.21]
Language Arts	20 (20,954)	-0.01 [-0.07, 0.05]	-0.4% [-2.79%, 1.99%]	High	From -0.42 [-0.55, - 0.28) to 0.20 (95% CI [0.05, 0.35]
Composite scores	9 (11,949)	-0.01 [-0.10, 0.08]	-0.40% [-3.98%, 3.19%]	High	From -0.34 [-0.54, 0.15] to 0.15 [0.08, 0.21]
Access outcomes					
Enrolment	3 (7,106)	0.01 [-0.04, 0.07]	0.40% [-1.60%, 2.79%]	Low	From -0.01 [-0.10,0.08] to 0.04 [-0.05, 0.12]
Dropout	7 (22,943)	-0.02 [-0.05, 0.01]	-0.80% [-1.99%, 0.00%]	Low	From -0.15 [-0.28, -0.01] to 0.00 [-0.07-0.07]
Completion	8 (3,092,767)	0.05 [0.00, 0.09]	1.99% [0.00%, 3.59%]	High	From -0.09 [-0.21, 0.03] to 0.32 [0.09, 0.55]
Secondary outcomes					
Teacher attendance	4 (663)	-0.01 [-0.26, 0.25]	-0.40% [-10.26%, 9.87%]	Moderate	From -0.20 [-0.49, 0.08] to 0.37 [0.08, 0.67]

Community-based Monitoring					
Summary of findings					
Outcomes	No of comparisons (sample size)	Overall effect size (95% CI)	Estimated percentile change compared to control group [95% CI]	Heterogeneity of overall effect	Range of effects (95% CI)
Learning outcomes					
Maths	9 (52,257)	0.09 [-0.02, 0.2]	3.59% [-0.80%, 7.93%]	High	From -0.39 [-1.07, 0.29] to 0.47, [0.27, 0.67]
Language Arts	9 studies (48,711)	0.12 [0.01, 0.22]	4.78% [0.40%, 8.71%]	High	From -0.05 [-0.14, 0.05] to 0.47 [0.27, 0.66]
Composite scores	7 studies (34,515)	0.10 [-0.01, 0.21]	3.98% [-0.40%, 8.32%]	High	From -0.01 [-0.08,0.07] to 0.48 [0.33, 0.62]
Access outcomes					
Enrolment	12 studies (9,757)	0.17 [0.08, 0.25]	6.75% [3.19%, 9.87%]	Moderate	From -0.08 [-0.22, 0.07] to 0.58 [0.34,0.82]
Attendance	3 studies (3,773)	0.04 [-0.09, 0.18]	1.60% [-3.59%, 7.14%]	High	From -0.06 [-0.13, 0.00] to 0.18 [0.09, 0.26]
Dropout	3 studies (2,280)	0.05 [-0.09, 0.20]	1.99% [-3.59%, 7.93%]	High	From -0.03 [-0.14, 0.08] to 0.22 [0.07, 0.37]
Completion	3 studies (2,656)	0.06 [0.01, 0.12]	2.39% [0.40%, 4.78%]	Low	From 0.01 [-0.10, -0.13] to 0.10 [-0.01,0.13]

Public Private Partnerships					
Summary of findings					
Outcomes	No of comparisons (sample size)	Overall effect size [95% CI]	Estimated percentile change compared to control group [95% CI]	Heterogeneity of overall effect	Range of effects [95% CI]
Learning outcomes					
Maths	7 (111973)	0.05 [0.01, 0.08]	1.99% [0.40%, 3.19%]	High	From -0.02 [-0.06, 0.02] to 0.14 [0.0, 0.21]
Language Arts	7 (108999)	0.04 [0.00, 0.09]	1.60% [0.00%, 3.59%]	High	From -0.04 [-0.08, 0.00] to 0.18 [0.11, 0.24]
Composite scores	4 (17332)	0.07 [-0.07, 0.20]	2.79% [-2.79, 7.93%]	High	From -0.01 [-0.04, 0.02] to 0.20 [0.16, 0.24]
Access outcomes					
Enrolment	7 (32866)	0.19 [0.01, 0.36]	7.53% [0.40%, 14.06%]	Low	From -0.02 [-0.17, 0.13] to 0.61 [0.46, 0.77]
Completion	3 (4305)	0.23 [-0.07, 0.53]	9.10% [-2.79%, 20.19%]	High	From 0.02 [-0.04, 0.08] to 0.53 [0.44, 0.62]

Multilevel Programmes

Multilevel Interventions					
Summary of findings					
Outcomes	No of comparisons (sample size)	Overall effect size [95% CI]	Estimated percentile change compared to control group [95% CI]	Heterogeneity of overall effect	Range of effects [95% CI]
Learning outcomes					
Maths	10 (72,575)	0.16 [-0.17, 0.48]	6.36% [-6.75%, 18.44%]	High	From -0.83 [-0.86, -0.81] to 1.00, [0.95, 1.05]
Language Arts	14 (76,105)	0.04 [-0.17, 0.26]	1.60% [-6.75%, 10.26%]	High	From -1.15 [-1.18, -1.13] to 0.73 [0.69, 0.78]
Composite scores	3 (45,110)	0.02 [-0.08, 0.12]	0.80% [-3.19%, 4.78%]	High	From -0.05 [-0.09, 0.00] to 0.12 [0.10, 0.14]
Access outcomes					
Attendance	3 (120,930)	0.16 [-0.12, 0.44]	6.36% [-4.78%, 17.00%]	High	From 0.01 [0.00, 0.02] to 0.45 [0.33, 0.58]
Dropout	3 (32,270)	0.16 [-0.33, 0.02]	6.36% [-12.93%, 0.80%]	High	From -0.40 [-0.64, -0.16] to -0.04 [-0.06, -0.02]
Completion	4 (53,334)	0.13 [0.04, 0.21]	5.17% [1.60%, 8.32%]	High	From 0.04 [0.01, 0.06] to 0.24 [0.20, 0.29]

List of figures and tables

Appendix figures and tables

1. Background and Objectives

1.1 The issue

There is widespread consensus on the importance of education for human well-being (Glewwe and Kremer, 2005). For instance, Sen argues that education has a ‘direct relevance to the well-being and freedom of people’ as well as an ’indirect role through influencing social change’ and ’economic production‘(1999, p. 296). In addition to the intrinsic value of education in and of itself, research suggests positive relationships between education and economic growth and earnings (Barro, 1991; Duflo, 2000; Psacharopoulos & Patrinos, 2004). This relationship becomes more pronounced in poorer countries (Psacharopoulos, 1985; Mankiw *et al*., 1992). Moreover, various studies have provided evidence of a link between better education systems and other indicators of human development, including health status, maternal and infant mortality, lower population growth and reduced crime (Glewwe, 2013; Hillman & Jenkner, 2004; Hannum & Buchmann, 2003). In other words, individuals with high levels of education are more likely to be employed, generate higher income, overcome economic shocks and maintain healthier families (World Bank, 2011).

Substantial efforts have been made in recent years to improve access to education in low- and middle-income countries (L&MICs). While there has been significant progress, this has been uneven and challenges remain. For instance, the net enrolment rate for children of primary school age increased from 80 to 91 per cent between 1990 and 2015 (UN, 2015). However, improvements in enrolment rates slowed down considerably after 2004 and in 2013, 69.3 million children of primary school age were still out of school (UNESCO, 2014a), more than half of them (39.1 million) in Sub-Saharan Africa (UNESCO, 2014b). While there has been progress in reducing the number of girls excluded from education, from 58 per cent in 1999 to 53 percent in 2010 (UN, 2012), girls are still more likely than boys to miss out on schooling. Girls’ participation rates remain lower than those of boys in 53 developing countries, with disparities particularly pronounced in West Asia and Sub-Saharan Africa (UN, 2015).

Moreover, improved access to education has failed to translate into learning in many countries (Pritchett, 2013). Studies measuring learning outcomes among school children across L&MICs find consistently low levels of learning, with hundreds of millions of children leaving school without basic numeracy and literacy skills (Pritchett, 2013; Robinson, 2011; UNESCO, 2015). For instance, according to the Education for All Global Monitoring report (UNESCO, 2013), around 250 million children in L&MICs cannot read, write, or do basic arithmetic. This number includes over 130 million children who, despite being enrolled in primary school, have not acquired these basic skills. Children will not receive a better education just by virtue of being in school if the conditions that enable learning are not also present (Petrosino et al., 2012; Pritchett, 2013). As Glewwe (2013: 3) argues, ’enrolment is not the final goal of education policy. The ultimate goal is to prepare children for a better life when they are adults.’

Therefore, while education interventions have traditionally focussed on getting children into school, researchers and decision makers are now focusing their attention on efforts to improve learning for all (Pritchett, 2013; Robinson, 2011). This is increasingly reflected in the education policy of major agencies (eg.: DFID, 2013; World Bank, 2011) and in the recently adopted Sustainable Development Goals (SDGs). Whereas the Millennium

Development Goal on education (Goal 2) was to achieve universal primary education, the education SDG calls for inclusive and equitable quality education (UNGA, 2015).

The focus on improved access and quality of education has also been accompanied by significant funding for education programmes. Despite the global economic crisis and regional food crises, domestic government spending on education increased in L&MICs between 1999 and 2011 (UNESCO, 2012). For instance, in low-income countries, the average real annual government spending on education grew at a rate of 7.2 per cent, and at a rate of 5 per cent in Sub-Saharan Africa (ibid). Recent estimates suggest governments in L&MICs spend around US$700 billion on education every year, with estimates suggesting parents contributing a similar amount (Glewwe 2014). Moreover, between 2002 and 2010, aid to education increased by 77 per cent to US$13.5 billion, with the World Bank, the USA and the UK being the largest donors to the sector (UNESCO, 2012). However, the resources available per child are limited and pale in comparison to the resources dedicated to education in high-income countries. For instance, in 2010 countries in North America and Western Europe spent US$7,916 on primary schooling per pupil (constant US$), in contrast to US$134 in Sub-Saharan Africa and US$263 in South and West Asia (UNESCO, 2012).

Evidence on the effects of education interventions can inform decisions about how limited funding can be best used. This paper reports the findings of a comprehensive systematic review of the literature on the effects of education interventions implemented in L&MICs. The next section provides a brief review of existing systematic reviews in the education sector. This is followed by a description of the conceptual framework developed for the review, before the chapter concludes with an outline of our objectives.

1.1.1 Review of existing literature

There is a relatively large literature of experimental and quasi-experimental evaluations assessing the effect of education interventions in L&MICs. Several authors have reviewed this literature in recent years, aiming to synthesise the evidence on the effects of different education interventions (Petrosino *et al*., 2012; Baird *et al*., 2013; Glewwe 2014; McEwan 2015; Morgan *et al.,* 2012; Morgan *et al.,* 2013, among others). A recent Evidence Gap Map provides an overview and appraisal of this literature (Snilstveit *et al., 2015*). The EGM identified 21 systematic reviews of interventions to improve primary and secondary education outcomes in L&MICs.

The majority of existing reviews focus on interventions to improve school enrolment and attendance. For instance, the impact of Conditional Cash Transfers (CCTs) on schooling outcomes has been examined by five systematic reviews (Baird *et al.,* 2013; Bouillon and Tejerina, 2007; Kabeer *et al.,* 2012; Petrosino *et al.,* 2012; Yoong *et al.,* 2012), three of which used meta-analysis (Baird *et al.,* 2013; Kabeer *et al.,* 2012; Petrosino *et al.,* 2012). These reviews all find that CCTs in education contribute to improving enrolment and attendance, though the evidence base on the effects on learning outcomes is limited, with available studies suggesting at best small effects. School vouchers are another popular intervention designed to improve access to education, primarily by enabling children from low-income households to access private education. Three systematic reviews include studies assessing the effects of school vouchers (Bouillon & Tejerina, 2007; Morgan *et al.,* 2013; Petrosino *et al*., 2012), one of which included meta-analysis (Petrosino *et al.,* 2012). Both Bouillon and Tejerina (2007) and Morgan *et al.* (2013) found positive effects of vouchers on attendance and performance, and an increase in private school enrolment among the poorest income groups respectively, while Petrosino *et al.,* (2012) found no effects of school vouchers. Similarly, both Morgan *et al.* (2012) and Petrosino *et al.* (2012) found no effect of interventions reducing or eliminating schooling costs.

School feeding and school-based health interventions are typically implemented to improve both school attendance and learning outcomes, and several systematic reviews have assessed the evidence on the effects of such interventions. Two systematic reviews found school feeding interventions to have positive effects on both attendance and learning outcomes (Kristjansson *et al.,* 2009; Petrosino *et al.,* 2012). Petrosino *et al.* (2012) found generally positive effects of a range of interventions providing health care (de-worming, vitamin A intake, malaria prevention, and menstruation cups), while Taylor-Robinson *et al.* (2012) found positive but weak evidence of the effects of de-worming on school attendance.

Several systematic reviews have assessed interventions targeting teachers, although only one includes meta-analysis. For instance, Orr *et al.* (2013) examined the effect of teacher training and find mixed effects on educational attainment. Guerrero *et al.* (2012) found that teacher monitoring in combination with incentives is effective in tackling teacher absenteeism, although they did not find effects on student achievement. On the other hand, three systematic reviews assessed the evidence on the effects of wage increases and incentives (Carr & Leggatt-Cook, 2011; Bouillon & Tejerina, 2007; Petrosino *et al.,* 2012) and found such interventions can have positive effects on students' attainment. Kingdon *et al.* (2013) reviewed the evidence on the effects of contract teachers, and concluded that contract teachers are more effective at improving student outcomes than teachers with permanent positions.

Finally, Bouillon and Tejerina (2007) reviewed the effect of decentralisation programmes and found that decision-making at the local level can improve performance in schools. The review was conducted some time ago, and subsequent (non-systematic) reviews (Bruns *et al.,* 2011) include additional studies on a broader range of school management interventions.

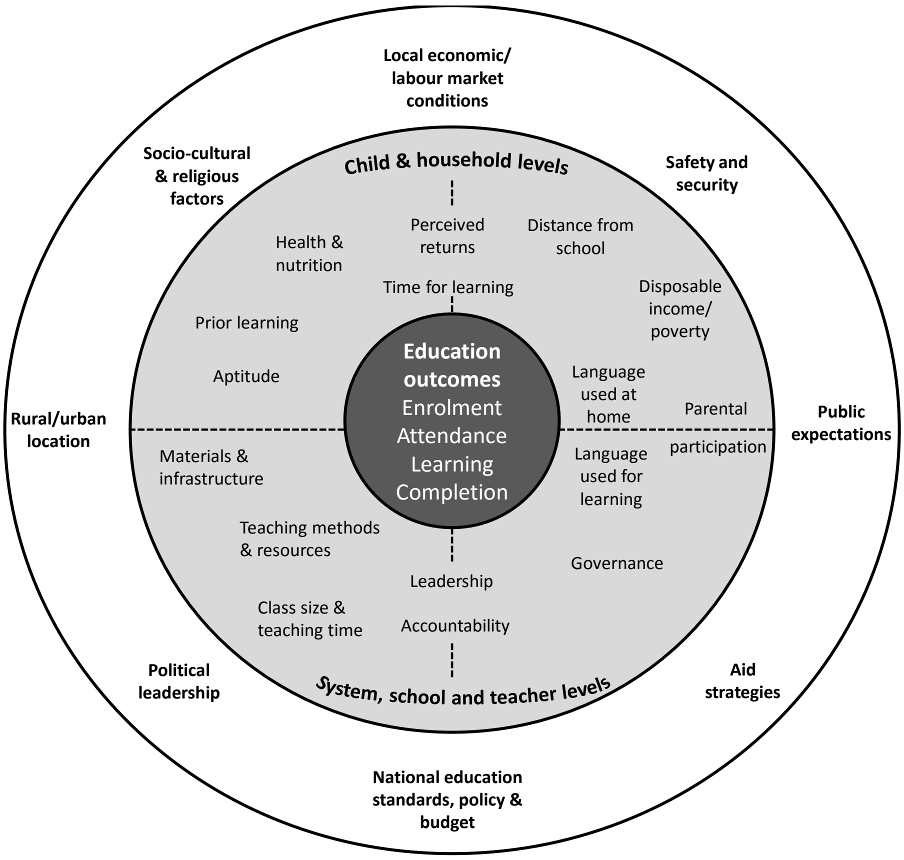
The above summary of findings from existing reviews reveals that while there is an increasing body of evidence on education interventions in L&MICs, existing reviews are scattered across a variety of interventions and outcomes. Moreover, most existing systematic reviews do not use statistical meta-analysis to synthesise findings, resulting in many reviews with mixed or contradictory results. Some reviews focus on enrolment and attendance and others examine only learning outcomes. Few reviews report academic completion outcomes, while only one provides findings on cost effectiveness. Few reviews assess the effects of interventions across a broad set of outcomes and several suffer from methodological shortcomings and rely on searches completed several years ago.

The systematic review and meta-analyses conducted by Petrosino *et al.* (2012), McEwan (2015) and Glewwe *et al.* (2014) are the most comprehensive reviews of education interventions in L&MICs conducted to date. This review expands on these studies in several ways. Firstly, Petrosino *et al.* (2012) only included studies reporting outcomes related to enrolment and attendance, excluding studies that evaluated learning outcomes only. Additionally, the search was conducted in 2009 and studies published after that date are not included. McEwan (2015) provides a meta-analysis of randomised controlled trials of education interventions in primary schools in developing countries. The study does not include any quasi-experimental studies, nor studies assessing access outcomes only. Finally, Glewwe *et al.* (2014) provides a comprehensive review of a broad range of literature, but the results of studies are synthesised using a vote counting approach based on the direction of statistically significant effects. Through its comprehensive search, its inclusion of a broader range of evidence and its rigorous methodology, this review provides a greater depth and breadth of analysis than any systematic review on this topic to date.

1.2 Interventions to improve access to education and learning outcomes

A range of educational interventions are being implemented to promote equitable access to education and ensure that all children can receive a quality education that can provide them with a better life. In this review we have provided an assessment of the effects of a broad range of such programmes. Children's education outcomes in L&MICs are influenced by a number of different factors, as depicted in figure 1a (DFID 2013; Tickly, 2011).

Figure 1 a: Determinants of education



To structure our review we developed a framework for classifying interventions based on the setting or 'level' of the main barrier a specific intervention aims to address. This follows a similar approach to the one adopted by Sherman *et al.* (2002) in their review of the evidence on a range of different crime prevention interventions. Thus, we grouped interventions into five broad categories at the child, household, school, teacher and systems levels. A description of interventions falling under each category is outlined below.

1.2.1 Child-level interventions

We define child-level interventions, or 'child settings', as those interventions targeting children directly, focusing on improving their ability to benefit from schooling or their incentives and motivation for investing time and resources in their own education. Interventions falling under this category include the following:

School feeding programmes typically aim to improve the general health of children, provide a safety net for vulnerable and food insecure families, and improve children's ability to learn (Jomaa *et al.,* 2011). Such interventions fall into two categories: the traditional school feeding programme, where children are provided with meals in school, and take-home ration programmes where children are provided with food in school which they can take home to their family (Lawson, 2012). In many cases the food provided is fortified or supplemented in order to give additional nutritional benefits (Jomaa *et al.,* 2011). School feeding programmes such as these are often targeted towards families and communities that are food insecure or have low incomes. Food for school feeding programmes is procured in a variety of different ways, but recently the focus has been on using local produce.

School-based health programmes include interventions to prevent or treat illness that are delivered to children at school. An example of such an intervention is the de-worming programme in Busia district, Kenya which provided children in schools with free de-worming treatment. The treatment was delivered by nurses and public health workers in local schools and was also combined with a course of worm-prevention classes and provision of wall charts and teacher training on worm prevention (Miguel & Kremer, 2003). Other examples include the provision of micronutrients to children (Kleiman-Weiner *et al.,* 2013) and eye tests, followed by provision of eye-glasses (Ma *et al.,* 2013).

Providing information to children about the potential future benefits of education in terms of income, employment, and social status is thought to increase school participation, enrolment and continuation, where students under-estimate the actual returns to education (Nguyen, 2008). Interventions of this type will typically involve providing information to the students about the future potential returns to schooling. The information can be presented in various ways including teachers or external presenters disseminating statistics about average earnings for each level of education. Other interventions make use of role models, who share their experience of education and current achievements with children, with some programmes using a combination of channels (Nguyen, 2008).

Merit-based scholarships aim to improve learning outcomes by rewarding high performing students with scholarships to continue their study (McEwan, 2013). For example, an intervention in Kenya provided scholarships to girls who performed well in their 6th grade exams. The programme awarded the top 15 per cent of students in the grade with a grant to cover school fees for two years, and also a cash sum which was to be used for school supplies, thereby intending to provide the girls with an incentive to perform well in school (Kremer *et al.,* 2009).

1.2.2 Household-level interventions

We define household level interventions as those initiatives aiming to reduce or remove financial household level barriers to education. These programmes can be delivered by governments, non-governmental organisations, religious organisations or international organisations. Interventions falling under this category include the following:

Cash transfers are typically divided into Unconditional Cash Transfers (UCTs) and Conditional Cash Transfers (CCTs). UCTs provide small cash sums to households to increase their income and the cash transfer is not conditional on any particular behaviour, such as school enrolment or attendance (Baird *et al.,* 2013). CCTs, on the other hand, provide cash sums to households conditional on certain behaviours, such as attending school. The increased household income is supposed to reduce prohibitive costs and any potential benefit to parents of sending their children to work rather than to school.

Programmes reducing or eliminating school user fees aim to improve access to schooling. Direct user fees, including payments for tuition, uniforms, textbooks and parent-teacher association contributions are common in many LMICs (Morgan *et al.,* 2012). Interventions to reduce or eliminate school user fees include removing all or some of these direct costs of schooling, for instance by providing school uniforms for free, or through the elimination of tuition fees, as has been done in many African countries over the last decades (Bentaouet-Kattan, 2006). Tuition fees may be universally removed, rolled out gradually or targeted towards particularly vulnerable groups (Morgan *et al.,* 2012).

1.2.3 School-level interventions

We define education interventions taking place at the school level as those initiatives aiming to improve the quality of the learning environment in schools and classrooms. They include interventions providing physical inputs, more schooling time for students or changes in how teaching is delivered. The education interventions implemented at this level typically fall under the following categories:

Structured pedagogy interventions are designed to improve the content and quality of instruction (Abeberese, Kumler, Linden, 2011; Lucas, *et al.,* 2014). Typically they seek to adapt or improve educational content and/or the methods by which students are taught. The fundamental objective of these programmes is to change existing classroom practices. Often this means developing new curricula and providing teachers with training in delivering new material, usually together with materials for both teachers and children.

Computer Assisted Learning (CAL) interventions use computers, either in the form of lap-tops or computer labs, to aid or support learning and to tailor classes to students' needs (Kremer, 2013). In some cases they are delivered as an integrated package together with new content, new instructional approaches and training for teachers in delivering this material. In other cases the main focus is simply on providing children with access to computers.

Remedial education interventions typically provide additional tutoring to small groups of children designed to provide tailored classes designed to meet students' needs (Banerjee *et al.,* 2007). Some programmes are designed for a particular subject, others target a particular demographic of students.

Extra time programmes aim to provide a longer school day with increased learning time for students. An increase in instructional time aims to increase instructional contact time and ultimately, to improve learning outcomes (NECTL, 2000). Typically, these programmes abolish 'shift' schooling whereby two separate cohorts attend the same school in a given day, one in the morning and one in the afternoon, and expand existing infrastructure so that all children can attend a full school day (Bellei, 2009)

New schools and infrastructure interventions typically include building a school in an area where there was not one previously, or rehabilitating existing facilities. This category may also include providing access to clean water for drinking and washing, safe waste disposal and separate toilets for girls to remove health related barriers to schooling as well as tackle incidents of harassment and humiliation in school toilets (Birdthistle *et al.,* 2011).

Interventions providing materials can assist teachers, facilitate learning and improve educational quality. Such interventions include any intervention providing 'traditional hardware' material such as books, chalkboards or other classroom equipment. For instance, the School Assistance Programme (SAP) funded by the Dutch non-profit organisation International Christelijk Steunfonds (ICS), provided English, Maths and Science text books to primary school children in Kenya (Glewwe *et al.,* 2009).

Grouping students by ability can help provide classes that are more closely tailored to students' needs. Some such interventions allow students with poorer school performances to remain in a grade so that they have additional time to learn material, rather than falling further behind (Chen *et al.,* 2010). Others group students by ability in order to ensure that teaching can be targeted to students' abilities (Duflo *et al.,* 2011).

1.2.4 Teacher-level interventions

We define teacher-level interventions as those interventions targeting teachers directly. These interventions include those designed to hire additional teachers and decrease pupil-teacher ratios. Other interventions are designed to provide teachers with new skills, provide performance-related incentives and increase accountability, or provide teachers with better information about student performance. Interventions falling under this category include the following:

Interventions providing teacher incentives seek to improve the working conditions in schools so that teachers are motivated to come to work and improve their performance. Such interventions take many forms, such as providing direct payments to teachers based on their attendance or based on the achievement of their students (Glewwe *et al.,* 2008). For instance, a programme in India offered teachers a cash bonus linked to their pupils' performance in independent tests (Muralidharan & Sundararaman, 2009). Similarly, a programme in Kenya offered primary-school teachers in-kind rewards based on pupils' exam scores (Glewwe *et al.,* 2010).

Teacher hiring interventions are designed to increase the number and the quality of teachers in schools. Recruiting new and additional staff can help address increased student enrolment rates and high pupil-teacher ratios (Vegas *et al.,* 2013; Kingdon *et al.,* 2012). Smaller class sizes facilitate more targeted tuition and students may receive increased individual attention and opportunities for participation in classes (Banerjee et al., 2007). Other teacher hiring interventions promote the employment of contract teachers instead of permanent civil-service teachers, with the guiding principle that employing teachers on short-term contracts can be economical and can increase incentives for teacher attendance and performance, while ensuring ensure that teachers are qualified and capable (Kingdon *et al.,* 2012). A final type of teacher hiring intervention is designed to introduce new hiring and promotion processes that will increase the quality of new appointees.

Diagnostic feedback interventions set out to provide information to teachers about student achievement so that they can target their teaching more effectively. Typically, this type of intervention relies on the introduction of frequent 'low stakes' tests that allow teachers to monitor students' knowledge and progress and tailor their teaching approach to encourage learning, without subjecting students to stressful high-stakes examinations (Muralidharan and Sundararamen, 2010; Duflo *et al.,* 2015).

Training teachers can help schools improve the quality of instruction and offer more targeted tuition for children that are falling behind. This type of intervention targets the general professional development of teachers in order to build their professional capabilities.

1.2.5 System-level interventions

We define this category of interventions as those aiming to improve education through changes to the education system at either the community, local government and district/state or national level. The interventions taking place at this level are primarily related to the management, governance and financing of education. Because of the nature of these interventions, they are typically implemented by governments, although non-governmental organisations, religious organisations or international organisations may be involved in delivery. The education interventions implemented at this level typically fall under the categories outlined below (drawing on Glewwe and Kremer 2005):

School-based management (SBM) interventions involve de-centralising authority to the school level to improve the quality of school administration and leadership. SBM programmes may involve handing decision-making (for example, on budget, staffing and curriculum development) over to teachers, parents, students or other community members (Barrera-Osorio, 2009). For example, the School Management Initiative in Hong Kong gave school committees authority over staffing and devising the curriculum, as well as some financial matters, aiming to create greater flexibility in school finance, increase accountability, and encourage collaborative decision making (ibid). Committees may also devise school improvement plans and receive funds to finance implementation of these plans. The Education Quality Improvement Project in Cambodia encouraged school committees to identify their school's needs, suggest improvements and then carry out reforms using cash grants from the Ministry of Education (WDR, 2004). These types of intervention may also foster greater accountability to parents or the community, increasing capacity to demand improved services, although they do not always include a participatory component.

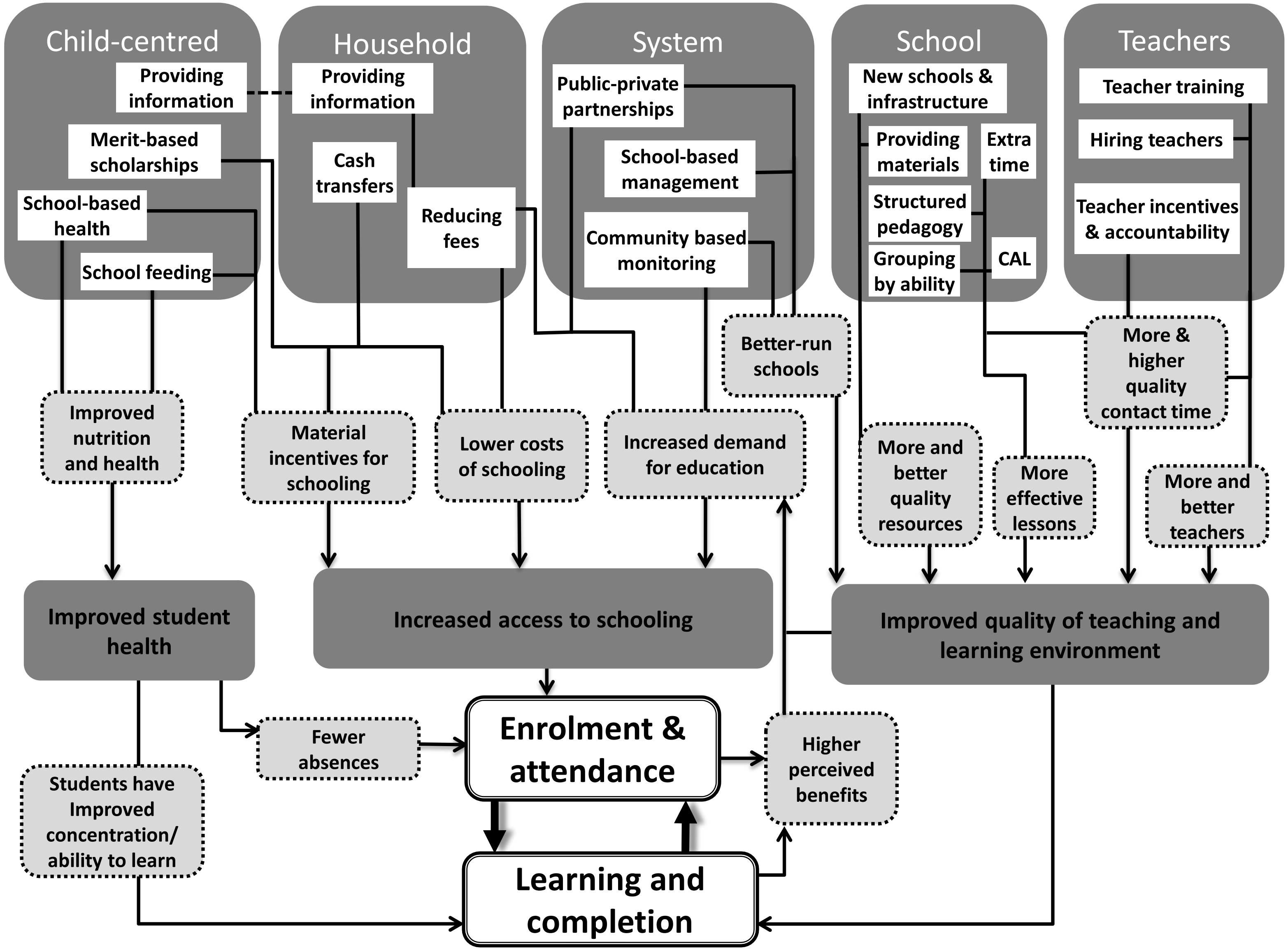
Community based monitoring interventions seek to improve the representation of communities in which service providers, governments, or other public bodies operate (Westhorp *et al.,* 2013). Interventions of this type are used in many sectors, including education, and aim to facilitate increased accountability between service providers and service users (ibid). An example of a community based monitoring intervention in the education sector is the use of a newspaper campaign to provide the public with information on education expenditure in Uganda (Reinikka & Svensson, 2004). In an effort to reduce corruption, the Ugandan government instigated the newspaper campaign, which published information on the amount of funds allocated to each district in both national and local newspapers. This allowed parents, head teachers and others access to information about school grants in their area and to complain if actual amounts received by schools were incorrect or untimely.

Public private partnerships and private provision of schooling may seek to increase parents' and students' choice, provide supply of schooling when there is none, or improve the quality of education provided (Barrera-Osorio *et al.,* 2009). Private schools may be run by profit, or by non-profit or faith-based organisations and there are a range of different mechanisms implemented to facilitate access to private education and school choice for children from poorer households. For instance, school vouchers finance all or most of school tuition fees through payments made by the government to a parent or to a school chosen by the parent, and have been implemented in a range of countries, including Colombia (Morgan *et al.,* 2013). In Pakistan, a programme attempting to induce the creation of private schools was subsidised through a fellowship programme for girls (Alderman *et al.,* 2003).

1.3 How the Interventions Might Work

The interventions included in this review draw on a wide range of programme theories in order to improve access to education and the quality of schooling. To do so, they must also address the wide-ranging set of determinants that are outlined in figure 1b. Our categorisation of these interventions into five levels is designed to help provide a structure for the review as a whole and to provide a framework for understanding the different ways that these interventions are designed to function. We have not provided a detailed discussion of how each intervention may work in this section as this is provided in individual chapters. Rather, we have outlined an overall framework designed to provide the reader with a basic overview of the logic underlying our interventions. Figure 1b categorises our interventions of interest according to our five 'levels' and outlining the main pathways through which they can influence education outcomes.

Figure 1 b: How the interventions might work



The interventions may work through a series of causal pathways either designed to improve school participation, or to facilitate learning by improving teaching, the learning environment, or by improving student health and therefore directly boosting their ability to learn. To pick an example, a child-level intervention such as school feeding aims to provide students with the nutrition they need to learn. Better nourished children are less likely to miss school due to illness, and better attendance can have the knock-on effect of improving learning outcomes (Kristjansson *et al.,* 2009). Such interventions may also provide an incentive for parents to send their children to school.

School-based management interventions are intended to improve the efficiency of school administration and leadership by facilitating innovation and allowing parent power to drive up the quality of schooling (Banerjee *et al.,* 2008). Improving the quality of schooling, and thereby improving learning outcomes may also have an important impact on enrolment and attendance - and vice-versa. Improving the learning environment will likely have a direct effect on learning outcomes, but might also have an indirect effect of pushing up demand by increasing the perceived benefits of schooling. Greater enrolment and attendance may change the student-teacher ratio, or lead to greater competition for limited resources or the inclusion of more children with a lower educational baseline. However, higher enrolment and attendance is likely to increase the absolute number of students completing school or passing exams and may even improve learning and completion through increased competition and gains in efficiency.

1.4 Review objectives

The primary objective of this review is to identify, assess and synthesise evidence on the effects of education interventions on children's access to education and learning in L&MICs. We also aimed to assess how education interventions affect different sub-groups of participants and to address questions relating to context, process and implementation.

To address these objectives, we answered the following questions:

1a) What are the effects of different education interventions on enrolment, attendance, dropout rates, completion and learning outcomes for primary and secondary school age children in low-and middle-income countries?
1b) Do the effects differ between sub-groups of participants (according to gender, age, urban or rural location, or socio-economic status)?
2a) What intervention and implementation features are associated with relative success and failure in improving educational outcomes?
2b) What are the contextual barriers to, and facilitators of, the effectiveness of educational interventions?

1.5 Report outline

The remainder of this report is structured as follows. Chapter 2 describes the methods used in the review. Chapter 3 provides a descriptive overview of the evidence base. It provides the results of the search, describes the characteristics of included studies and summarise the results of the critical appraisal of all included studies. Chapters 4-8 then present the main results for all included interventions, structured according to the conceptual framework outlined above. Chapter 9 reports the results for interventions classified as 'multilevel' interventions. Chapter 10 presents the results of meta-regression analysis and assessment of publication bias, before providing a summary of findings and implications for research, policy and practice.

2. Methodology

The review follows the Campbell and Cochrane Collaborations guidelines to systematic reviewing (Campbell Collaboration, 2015 Hammerstrøm *et al.,* 2010; Higgins & Green, 2011; Shadish & Myers, 2004). We systematically collected and synthesised quantitative evidence from impact evaluations of education interventions to answer review questions 1a and 1b. When sufficient data was available outcomes were synthesised along the causal chain, from intermediate to final outcomes. For the review to be more useful for policy-makers and practitioners, we extended the review of effectiveness (Noyes *et al.,* 2011). This was achieved by collecting quantitative and qualitative evidence on process and implementation, context to address review questions 2a and 2b.

2.1 Criteria for including and excluding studies

We included studies in two phases (See Figure 3, below). To address questions 1a and 1b, we included quantitative impact evaluations of interventions included in the framework presented in chapter 1. To address questions 2a and 2b, studies that passed the criteria for inclusion to address questions 1a and 1b were used as the basis for a second phase to identify and include qualitative studies, project documents, process evaluations and other supplementary data on the programmes examined by the studies included to address questions 1a and 1b.

2.1.1 Types of study designs

To address questions 1a and 1b we included studies that assessed the effects of interventions using experimental and quasi-experimental study designs that allow for causal inference. Specifically, we included:

Studies where participants are randomly assigned to treatment and comparison group (experimental study designs);

Studies where assignment to treatment and comparison group is based on other known allocation rules, including a threshold on a continuous variable (regression discontinuity designs) or exogenous geographical variation in the treatment allocation (natural experiments);

Studies with non-random assignment to treatment and comparison group, provided they include pre-and post-test measures of the outcome variables of interest to ensure equity between groups on the baseline measure, as well as use appropriate methods to control for selection bias and confounding, such as statistical matching (for example, propensity score matching, or covariate matching), regression adjustment (for example, difference-in-differences, and single difference regression analysis, instrumental variables, and 'Heckman' selection models).

Quasi-experimental studies may be subject to bias in their estimate of treatment effects, however, studies have shown that if well conducted quasi-experimental studies can provide un-biased estimates of treatment effects (Cook *et al*., 2008; Shadish, 2011). In setting our inclusion criteria we aimed to incorporate studies that adopt techniques which empirical research suggest are effective in reducing or removing bias. Including a pre-test measure of the outcome and controlling for appropriate covariates in particular have been found to be important in reducing selection bias in quasi-experimental studies (Steiner *et al.*, 2010; Shadish, 2011).

Thus, studies without random allocation to treatment and comparison group that do not include a baseline measure of the outcome variables were excluded. Similarly, studies without random allocation to treatment and comparison group without using matching or other statistical methods to control for selection bias and confounding were also excluded. The selection and measurement of appropriate covariates that are correlated with both the selection or allocation of the treatment and the outcomes is important in reducing selection bias in quasi-experimental studies (Steiner *et al*., 2010). We did not address this issue at the inclusion stage, but did so when assessing the risk of bias in all included studies (details below).

To avoid confounding treatment effects with teacher or school effects studies of any classroom and school level educational interventions, whether they are randomised or not, were excluded if they had less than two teachers or schools in each group. Finally, our interest was is in identifying the evidence on the effects of an intervention implemented as part of a programme under circumstances that approach 'real- world' practice, so- called effectiveness studies. These types of studies stand in contrast to efficacy trials which test an intervention under ideal and controlled conditions in order to maximise the likelihood of observing an effect, if one exists.

Although there exists broad agreement on the type of study design characteristics of effectiveness (pragmatic) trials and efficacy (explanatory) trials, there is currently no validated definition of 'effectiveness studies' (Treweek *et al.,* 2009; Gerthlener *et al.,* 2006; Singal *et al.,* 2014). Furthermore, as Thorpe *et al.* note (2009), the distinction between the two types of trials can be regarded as a continuum rather than a dichotomy as very few trials are purely pragmatic or explanatory. Initially we had planned to include studies on this continuum, and then classify studies according to where on this continuum they were. However as we started reviewing the literature we discovered a large body literature of very small experimental and quasi-experimental studies assessing the effects of very specific techniques, such as use of concept mapping or a specific software, delivered over a short time period. These studies did not evaluate a programme, but were also not explicitly exclude by our existing criteria. Because of the already large scale of our review, the methodological limitations of most of these studies and the difference between these studies and the programme evaluations we were primarily interested in we decided to develop criteria which allowed us to appraise these studies systematically for inclusion in our review.

We developed five criteria to help us distinguish more clearly between efficacy trials and effectiveness studies, drawing on two existing tools (Gartlehner *et al*., 2006; Thorpe *et al.,* 2009). Studies were considered efficacy trials and are excluded if they fulfil at least one of the criteria outlined below:

- Research Objective: Is the study primarily designed to determine to what extent a specific technique, technology, treatment, procedure or service works under ideal condition rather than attempt to answer a question relevant to the roll- out of a large programme?
- Population: Are the participants highly selected and therefore unrepresentative of the general population (Are strict inclusion and exclusion criteria used to enrol a homogenous population which may limit the generalizability of the results? e.g. students that truly have a disease of interest or are more likely to adhere to the treatment)?
- Providers: Is the intervention primarily delivered by the research study team rather than trained laypersons (parents/ teachers/ community members/ NGOs) who don't have extensive expertise?
- Delivery of intervention: Is the intervention delivered with high degree of assurance of delivery of the treatment? (Is the delivery tightly monitored/ supervised by the researcher

following specific protocols; Is adherence to the treatment monitored closely with frequent follow- ups?)

- Delivery of intervention: Are concurrent interventions restricted to the study population in order for a witnessed effect to be attributed to the intervention of interest?

To address question 2a and 2b, we extracted data on relevant evidence from studies included to address questions 1a and 1b. However, the lack of details about the study contexts, beneficiaries and interventions within primary studies can be a barrier for review authors seeking to incorporate this information in their reviews (Herbert & Bø, 2005; Roen *et al.*, 2006). Therefore to address questions 2a and 2b we also included studies and documents linked to the interventions studied in the included impact evaluations AND that met at least one of the following criteria:

- A qualitative study collecting primary data using qualitative methods of data collection and analysis, and report some information on all of the following: the research question, procedures for collecting data, sampling and recruitment, and at least two sample characteristics.
- A descriptive quantitative study collecting primary data using quantitative methods of data collection and descriptive quantitative analysis and report some information on all of the following: the research question, procedures for collecting data, sampling and recruitment, and at least two sample characteristics;
- A process evaluation assessing whether a policy is being implemented as intended and what is felt to be working more or less well, and why (HM Treasury, 2011). Process evaluations may include the collection of qualitative and quantitative data from different stakeholders to cover subjective issues, such as perceptions of intervention success or more objective issues, such as how an intervention was operationalised. They might also be used to collect organisational information;
- A project document providing information about planned, ongoing or completed interventions. They may describe the background and design of an intervention, or the resources available for a project for instance. As such, these documents do not typically include much analysis of primary evidence, but they provide factual information about interventions. The purpose of including them in our review is to ensure we had sufficient information about the context and interventions in included studies.

2.1.2 Type of Participants

The review included interventions targeted at primary school and secondary school[1] age children in mainstream education in L&MICs, as defined by the World Bank at the point in time that an intervention was carried out. We excluded studies focusing on refugees, migrants and orphans only. We also excluded studies focusing on children with special educational needs. All adult education interventions, including those that are university-based were excluded. We also excluded studies from high income countries as the differences with L&MICs, in terms of policy challenges, resources devoted to education systems, state capacity and broader contextual factors, are such that we consider this evidence to be of limited applicability.

[1]Since different age ranges attend primary and secondary school in different countries, we applied national criteria from each relevant country as necessary, noting that in most countries this is 4/5+.

2.1.3 Type of Interventions

We focused on interventions aiming to improve the access and/or quality of primary and secondary education in L&MICs. The outcomes we are interested in are determined by a wide range of factors, meaning that a single intervention is unlikely to be sufficient to address the barriers faced by children, families and education systems across the diversity of contexts covered in this review. We used an intervention typology based on different settings to specify the interventions we included in our review, as outlined in more detail in Chapter 1 above. We do not repeat the descriptions of all interventions here, but Table 2a summarises included interventions according to intervention level.[2]

The intervention inclusion criteria listed here are the same as those listed in our protocol and used for including studies in our review. However, the framework in Chapter 1 and the grouping of studies for the purposes of analysis deviates slightly from this in the following way: The category of Instructional approach, content, time and organisation interventions has been broken out into more specific sub-categories based on the characteristics of included studies (computer assisted learning, pedagogy, extra time, remedial education, grade retention, tracking). Providing information to parents and providing information to children have been merged into one category. Scholarships and allowance have been removed as a standalone category as it became clear that rather than being a distinct intervention category studies initially classified under this category had overlaps with either cash transfers broadly defined and merit based scholarships. We have also included diagnostic feedback as an additional intervention category at the teacher level. Due to the broad scope of our review and lack of existing comprehensive intervention frameworks we anticipated that the intervention classification may change when developing the study protocol and our search strategy and inclusion criteria were developed to be sufficiently broad to allow for such iteration.

Table 2 a: Included Intervention Types

Intervention level	Intervention type
Child level	School feeding programmes
	School-based health programmes
	Providing Information to children
	Merit based scholarships
Household level	Cash transfers
	Scholarships and allowances
	Reducing or eliminating school user fees
	Providing information to parents
School level	Instructional approach, content, time and organisation interventions
	New schools & infrastructure
	Interventions providing materials

[2] The intervention inclusion criteria listed here are the same as those listed in our protocol and used for including studies in our review. The framework in Chapter 1 deviates slightly from this in the following way: The category of Instructional approach, content, time and organisation interventions has been broken out into more specific sub-categories based on the characteristics of included studies (computer assisted learning, pedagogy, extra time, remedial education, grade retention, tracking).

Teacher level	Teacher incentives
	Teacher training
	Hiring additional teachers
System level	School-based management (SBM)
	Community based monitoring (CBM)
	Public private partnerships and private provision of schooling (PPP)

The following interventions do not meet the inclusion criteria and were excluded from the review:

Early childhood development: While ECD is clearly an important part of education, it is a separate sub-component of education and does not directly address primary and secondary education. Moreover, a team at the World Bank is currently working on a systematic review covering all ECD interventions.

Girls' sexual and reproductive health: This is a separate sub-component of education. While important for girls' education, including such interventions would further add to the scope of the review as it would have included a large literature on preventions of HIV and other STDs.

School-based nutrition and health promotion: Such interventions generally have a primary focus on improving knowledge and related health and nutrition behaviour, rather than on improving the primary outcomes of interest in this review. Any impact on education is likely to be indirect and such interventions are therefore excluded from this review.

Interventions teaching physical activities: We focused on interventions promoting key academic subjects such as maths, reading and science, and programmes to promote physical activities are thus excluded.

Distance education: These interventions tend to be focused on further education and adult education and are less common for primary and secondary schools. Distance education can be seen as a separate sub-category of education and may be better reviewed on its own.

Special Educational Needs interventions: The review focused on mainstream education and special education can be seen as a separate sub-component of education and thus, is better dealt with in a review on its own.

Interventions to address disruptive behaviour: While addressing disruptive behaviours and improving discipline are important outcomes, we consider such interventions to be 'second order' interventions and they are beyond the scope of this review.

Microcredit: Microcredit interventions are not primarily about improving education, and any impact on educational outcomes are likely to be indirect, through household income.

Roads and other community wide infrastructure: These are not primarily about improving education, and any impact on educational outcomes may only be incidental.

Community wide health interventions: We have not included community-wide or general health interventions as education is not a primary outcome, and if measured, educational outcomes are incidental.

Interventions extending the school year or duration of primary/ secondary school.

2.1.4 Type of outcomes

To be included, studies had to assess at least one of the education related primary or secondary outcomes described below.

Primary outcomes

Enrolment: defined as the number of students registered for education at the start of primary/secondary education or a given grade year.

Attendance: defined as a measure of the proportion of total school days for which enrolled students are present during the period in which a school is in session. If the only attendance outcome measured was attending an exam we included this as a measure of attendance. If studies only measured absenteeism we included this and reversed the sign (i.e., from a negative to a positive) of the estimate for inclusion in analysis.

Drop-out: defined as the number of children that enrolled in school but at some point in the year ceased to attend (UNESCO, 2005; USAID, 2011).

Completion: defined as the number of students completing primary/secondary education or a given grade. The studies used a range of outcomes to measure whether the intervention results in more children completing a grade. Some studies used 'failure rate' or 'repetition rate' as an outcome variable, and others used pass rate. In the former two cases, a negative effect represents a successful outcome, in the latter a positive effect represent a successful outcome. To be able to combine the studies in a way where they all consistently report the effect on the outcome of interest, we changed the sign for the studies measuring failure or repetition, so that a positive sign represents success. While there were more studies measuring failure than pass rates, the positive sign is a more intuitive representation of a successful outcome in terms of completion rates, so we decided to convert all measures to the positive sign.

Learning: learning is a broad concept and different outcome measures are used to measure children's learning. Existing systematic reviews and meta-analyses with a focus on education quality and learning outcomes adopt different inclusion criteria, from a narrow focus on exam results (Taylor-Robinson *et al.,* 2012) to including outcomes across a range of subjects as long as studies provide a continuously measured outcome (McEvan, 2013).

The Learning Metrics Task Force (LMTF) (2013) proposed a broad framework of seven different domains of learning outcomes that are important for children and youth to master. These range from physical well-being, to literacy and communication, learning approaches and cognition, numeracy and mathematics, and science and technology. The framework reflects a holistic approach to children's learning, and the LMTF argues that learning should not be oversimplified by focusing on only some domains. Nevertheless, it also recognises the challenges involved in measuring outcomes across all of these domains, and that efforts to measure learning outcomes at a global level may have to focus on a more narrow set of measures to be feasible.

It was not feasible to include learning outcomes across all domains in our review and we focused on outcome measures assessing children's learning in a few key domains: (1) maths and language arts (local language and any official language(s) of country), (2) cognitive and problem solving skills, and (3) composite assessment scores from test scores in different subjects or other measures of skills and learning.

Studies reported a range of outcome measures for language skills, including reading, writing, literacy, 'language' (without specifying which language or what was measured) and a specified language (such as ‘Spanish’ or ‘Hindi’ etc.). We included all such measures under an outcome construct called ‘language arts’. If studies reported outcomes for different languages, for instance an official language such as English and a local language such as Swahili, we selected the local language with the aim of selecting the language representing the local language of the largest number of children.

Our preference was for the most comprehensive outcome within a study. Therefore, if a study reported literacy and writing, we used the literacy outcome. If studies only reported disaggregated outcomes (e.g.: different components of literacy) without reporting an aggregate effect size we calculated an average or ‘synthetic’ effect size which we then included in the meta-analysis (Borenstein *et al.,* 2009, Chapter 25). Alternatively, if we had sufficient studies to use robust variance meta-analysis we included all disaggregated outcomes measures in the meta-analysis. We followed the same approach for maths and cognitive outcomes.

Secondary outcomes

These include other education related secondary outcomes, including: (1) teacher attendance: defined as a measure of the proportion of total school days for which teachers are present; and (2) teacher performance: defined as any measure of teachers’ knowledge, practice, motivation or satisfaction (Orr *et al.*, 2013).

We included both measures of teachers present in the school, and teachers present in classroom. If one study reported both of these outcomes, we included the teacher present in classroom in the meta-analysis. For some intervention levels or categories, we identified a very diverse range of teacher performance measures. Given that each study had a different mixture of teacher performance measures, we decided not to pool all of the diverse measures reported in each study into a single index. We created sub-categories of teacher performance outcomes that measure similar concepts:

Classroom management: measures such as a teacher was in control of the class, children were observed sitting within the classroom, and classroom was clean and orderly.

Student engagement and active teaching: teachers observed interacting with students, active teaching at point of observation, teacher calls student by name, teacher addressed questions to student, teacher provided group or individual help, teacher encouraged participation, teacher made children read from textbook.

Using assessment in instruction: teacher assigned homework, teacher gave a test, teacher provided homework guidance, and teacher provided feedback on homework.

Use of materials: teacher used blackboard, teacher used teaching aid, teacher read from textbook.

Preparatory sessions and activities: teacher offered preparatory sessions, teacher provided extra classes/ teaching beyond school hours, teacher provided special preparations for the end of year test.

In addition to the outcomes specified above we also collected data on other secondary and intermediate outcomes if they were reported in studies that met all other inclusion criteria. Relevant secondary outcomes vary for the different interventions included in our review and were identified based on intervention specific programme theories.

Other issues

As noted above, we included qualitative studies, process evaluations and project documents related to the programmes studied in included impact evaluations. From these documents we included descriptive information about programme design and implementation, context and resources, as well as any findings addressing questions 2a and 2b on barriers and facilitators of intervention success or failure.

2.1.5 Type of comparison

To answer question (1), we included studies that used any type of comparison group. As such, a study could use a comparison group that receives no intervention (including wait-list comparisons as part of pipe-line designs), business as usual, or a different form of educational intervention. However, we did not combine studies with business as usual and active comparison groups in the same meta-analysis. Comparisons could be between schools, groups of students or areas such as school districts.

2.1.6 Other criteria for including and excluding studies

We excluded any studies published before 1990. A review of the systematic review literature described in the evidence gap map cited above showed that the earliest cut-off point employed by any review was 1990. Pilot searches carried out to help guide the protocol development returned no studies published before 1990. An overview of study inclusion criteria is provided in Table 2 below.

Table 2 b: Summary of inclusion criteria

Study Characteristic	Inclusion criteria
Population	Primary and secondary school age children in mainstream education in L&MICs
Intervention	Interventions with primary focus on educational outcomes
Comparison	No intervention or different education intervention
Outcomes	Primary outcomes: enrolment, attendance, drop-out, completion, learning (maths, language arts, measures of cognitive and problem solving skills, and composite assessment scores); Secondary outcomes: teacher attendance, teacher performance, intervention specific outcomes
Study Type	1a and 1b: Experimental studies and quasi-experimental studies 2a and 2b: Studies included to address 1a and 1b + qualitative studies, descriptive quantitative studies, process evaluations, project documents linked to interventions studied in included experimental and quasi-experimental studies
Timeframe	Studies published from 1990 onwards – July 2015

2.2 Search strategy

A comprehensive search of the literature for a systematic review on a topic in international development should cover key bibliographic databases, those specific to international development, those specific to social sciences, and specific to the subject of the review

(Waddington *et al.,* 2012). The search strategy was developed in collaboration with an information specialist (JE) and with reference to the guidance in Hammerstrøm *et al.* (2010). In addition we used pearl-harvesting – collecting keywords from studies that meet our inclusion criteria (Sandieson, 2006).

To capture the relevant literature as comprehensively as possible, we developed both a general set of search terms and a series of sub-strategies designed around the typology of educational interventions set out above. An example of these search strategies is included in Appendix 2. The strategy was adapted to fit all the electronic databases included in the search and where appropriate, thesaurus terms were used in addition to natural language terms in those databases where both can be searched. All searches were limited by the L&MICs filter and by year, from 1990 onwards.

To ensure sensitivity, the study methods filter has been excluded from the searches in accordance with Campbell guidelines, with the exception of the general education search where the large number of papers retrieved made scanning impracticable. With this general search, further refinements to reduce numbers were made using the Web of Science Research Areas topics. In the Web of Science search example given in Appendix 2, all results for each of the categories have been combined using the OR Boolean operator to achieve an overall total. Citation searches of included studies were carried out in Web of Science (SSCI & AHCI), Scopus and Google Scholar. The main searches were conducted in January 2014 and updated in June 2015.

2.2.1 Electronic searches

We searched a range of databases and websites, including subject-specific education databases as well as general social science databases.

We searched the following academic databases:

Africa Wide: http://www.ebscohost.com/academic/africa-wide-information

Academic Search Premier: http://www.ebscohost.com/academic/academic-search-premier

Applied Social Science Index and Abstracts (ASSIA): www.csa.com/factsheets/assia-set-c.php

CAB Abstracts

Econlit

Education Resources Information Center (ERIC)

International Bibliography of the Social Sciences (IBSS): http://search.proquest.com/ibss?accountid=149134

PAIS International (Public Administration Information Systems)

PsycInfo

Sociofile/SocIndex

Sociological Abstracts: http://search.proquest.com/socabs

Web of Science: Social Science Citation Index (SSCI) and Arts & Humanities Citation Index (AHCI): http://ip-science.thomsonreuters.com/cgi-bin/jrnlst/jloptions.cgi?PC=SS

Worldwide Political Science Abstracts

We searched the following health databases only using health terms:

Global Health (CABI) (only school feeding and health terms)

Embase (only school feeding and health terms)

Medline (only school feeding and health terms)

We searched the following electronic libraries and registries of impact evaluations:

3ie Systematic Reviews Database

EPPI-Centre Evidence Library

Campbell Library

Cochrane Library (only health terms)

AEA (American Economic Association) RCT Registry

British Library of Development Studies (BLDS): http://blds.ids.ac.uk/

JOLIS (Joint Libraries of the World Bank and IMF): http://external.worldbankimflib.org/external.htm

3ie Register of Impact Evaluation Published Studies: http://www.3ieimpact.org/en/evidence/impact-evaluations/

3ie RIDIE (Registry for International Development Impact Evaluations): http://ridie.3ieimpact.org/

2.2.2 Grey Literature Searching

To ensure maximum coverage of unpublished literature, and reduce the potential for publication bias, we searched the following organisational websites and databases for unpublished grey literature:

Best Evidence Encyclopaedia (BEE)

British Education Index (BEI): http://www.leeds.ac.uk/bei/

DAC Evaluation Resource Centre (DEReC: http://www.oecd.org/derec/publicationsdocuments/

Dissertations & Theses Database (Proquest)

British Library Electronic Theses online Service (EtHOS): http://ethos.bl.uk/Home.do

Networked Digital Library of Theses and Dissertations Index to Theses: http://www.ndltd.org/

Open Grey: http://www.opengrey.eu/search/

ELDIS: http://www.eldis.org/

EVIPNET (Evidence Informed Policy Network) (focus on health, so screening limited to school feeding and school based health interventions)

Global Partnership for Education: http://www.globalpartnership.org

Innovations for Poverty Action (IPA) Database: http://www.poverty-action.org/project-evaluations/search

Abdul Latif Jameel Poverty Action Lab (J-PAL):
http://www.povertyactionlab.org/search/apachesolr_search?filters=type:evaluation

University of California Center for Effective Global Action (CEGA): Research Projects:
http://cega.berkeley.edu/research/

Development Impact Evaluation Initiative (DIME):
http://web.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTDEVIMPEVAINI/0,,contentMDK:21553788~pagePK:64168445~piPK:64168309~theSitePK:3998212,00.html

Rural Education Action Programme: http://reap.stanford.edu/

IDEAS/RePEc: http://ideas.repec.org/

NBER (National Bureau of Economic Research): http://www.nber.org/

Social Science Research Network (SSRN):
http://papers.ssrn.com/sol3/DisplayAbstractSearch.cfm

Bureau for Research and Economic Analysis of Development (BREAD) working papers:
http://ipl.econ.duke.edu/bread/papers.htm

Proceedings for past American Economic Association (AEA) and the Northeast Universities Development Consortium (NEUDC) conferences 2008 to 2013.

We searched the following websites of bilateral and multilateral organisations relevant to this review:

AUSAID

CIDA

DANIDA

DFID (including Research for Development (R4D): http://r4d.dfid.gov.uk/

SIDA

UNDP

USAID (Including USAID Development Experience Clearing House:
https://dec.usaid.gov/dec/content/search.aspx)

World Bank (Including Sector pages, World Bank's Impact Evaluations in Education (IE2):
http://datatopics.worldbank.org/EdStatsApps/Edu%20Evaluation/evaluationHome.aspx;

World Bank Impact Evaluation Working Paper Series:
http://go.worldbank.org/0OY9ERG1E0)

Inter-American Development Bank (Including sector pages and:
http://www.iadb.org/en/office-of-evaluation-and-oversight/evaluations,1578.html)

Asian Development Bank (Including sector pages and evaluation resources:
http://www.adb.org/site/evaluation/resources/1232?keyword=)

African Development Bank (Including sector pages and evaluation reports:
http://www.afdb.org/en/documents/evaluation-reports/)

Millennium Challenge Corporation (Independent Evaluations:
http://www.mcc.gov/pages/results/evaluations)

2.2.3 Other searches

We screened the bibliographies of included studies and existing reviews for additional eligible studies and conducted forward citation-tracking of included studies in Web of Science. We also identified and contacted key researchers and organisations working in the education field, including the UK's Department for International Development (DfID), The Abdul Latif Jameel Poverty Action Lab (J-PAL), the Rural Education Action Programme (REAP), UNICEF, UNESCO, World Bank and key bilateral donors.

Titles and abstracts were screened against the inclusion criteria and relevant records were downloaded into the review management software EPPI reviewer. The initial screening of records was conducted by several reviewers screening the records from different databases. At this stage, we were over-inclusive to ensure relevant studies were not omitted because sufficient information was not reported in title or abstract. Two reviewers then independently reviewed abstracts that were judged to be potentially relevant at the first stage in more detail to determine which papers should be retrieved and reviewed at full text. Two reviewers then independently assessed full text studies for inclusion, with any disagreements determined by a third reviewer.

Targeted search for addressing review questions 2a and 2b

When we had determined which studies were included in the review of effectiveness, we undertook targeted searching for qualitative studies, as well as process, implementation and cost information for those interventions evaluated in the included studies. We conducted citation tracking of included studies to identify any relevant sister papers and conducted internet and database searches using the names of programmes from included studies. To identify project documents and process evaluations, we conducted targeted searches of databases of project documents and websites of implementing agencies. Finally, we contacted authors and implementing agencies to request available project documentation. A more detailed description of the targeted search is provided in Appendix A.

2.3 Data extraction and coding procedures

Using a standardised data extraction form, we extracted three main categories of data: (1) descriptive data on study design, intervention and context for purposes of descriptive analysis of the body of research; (2) data on the population, context, study design, intervention design and process and implementation for purposes of moderator analysis and narrative synthesis addressing question 2; and (3) data on outcomes and sample for purposes of effect size calculation.

For 1 and 3 we extracted data primarily from the included impact evaluations, although we supplemented 1 with any relevant information from studies and documents identified through the targeted search. Both impact evaluations and associated documents were coded for data on population, context, study design, intervention design and process and implementation. The information necessary to calculate effect sizes for all outcomes included in the review was collected from each study selected for inclusion where possible, as detailed below. The code book is provided in Appendix C. Due to resource constraints we implemented a more reduced code book for studies assessing cash transfers, focusing on intervention design, study design and effect size data.

Before proceeding with extracting data for effect size calculation we mapped all the included studies to identify which outcomes they assessed, for which follow up periods, sub-groups, treatment arms and specifications. For the majority of the papers we discussed each paper in a group involving substantive reviewers and a statistician. Based on an assessment of how the outcomes met our inclusion criteria and comparing studies with other studies in the same intervention categories we made decisions about which outcomes to extract data from.

One researcher then extracted data used to calculate effect sizes and their variance, and to conduct the critical appraisal of all included studies. A second researcher then independently reviewed the data and made changes where necessary. The detailed coding of contextual, implementation and cost information was conducted by one person, with a second person checking and adding any additional information. This coding of context, implementation and cost data was done in EPPI reviewer.

2.4 Critical appraisal

Review question (1): assessment of risk of bias in included studies of effects

We assessed risk of bias using the following categories, based on categories of bias recommended by the Cochrane Non-Randomised Studies Group and the risk of bias tool developed by Hombrados and Waddington (2012). The tool has been developed to allow consistent assessment of internal validity of social experiments and quasi-experiments including randomised controlled trials (RCTs), regression discontinuity designs (RDDs), non-randomised studies based on participant self-selection (panel data models, propensity score and covariate matching, and cross-sectional regression), and studies using instrumental variables estimation for causal identification.

The risk of bias tool includes evaluation criteria to assess risk of bias across the following domains:

Baseline confounding and selection bias: was the allocation or identification mechanism able to control for baseline confounding and sample selection bias (censored data)?

Time-varying confounding: was the method of analysis executed adequately to ensure comparability of groups throughout the study?

Bias due to missing data: is the estimation method sensitive to non-random attrition?

Biases in outcome data collection: was the process of being observed causing motivation bias (Hawthorne and John Henry effects, courtesy bias, and recall bias)?

Departures from intended interventions: was the study adequately protected against performance bias and survey effects?

Outcome & analysis reporting biases: was the study free from outcome reporting bias and analysis reporting bias?

We assessed the risk of bias within included studies across the domains outlined above, coding papers as ‘Yes’, ‘No’ and ‘Unclear’ according to how well they address each domain. Because of the large number of studies only one reviewer conducted the risk of bias assessment. We followed a similar approach to de Vibe *et al.* (2012) and report a summary of the risk of bias across all studies for each risk of bias domain.

Critical appraisal of qualitative studies, process evaluations and project documents

Including a broader range of evidence can complicate critical appraisal, particularly as there is a lack of existing tools and criteria for quality (Noyes *et al.*, 2011). We anticipated that additional sources included to address questions 2a and 2b would fall into four main categories as outlined in the inclusion criteria. We adopted different approaches to appraise these three types of studies/ documents, as outlined below.

We assessed the quality of included qualitative studies and descriptive quantitative studies using an adapted version of the Critical Appraisal Skills Programme checklist (CASP, 2006), making judgments on the adequacy of reporting, data collection, presentation, analysis and conclusions drawn. The checklist is included in Appendix D. We filtered out studies of particularly low quality at this stage (Noyes *et al.*, 2011) and studies where questions 1-5 are assessed as "No" were excluded at this stage.

There are no commonly used critical appraisal tools for process evaluations. Such analysis needs reliable data from a representative sample, so assessment of sampling and methods of data collection are obvious issues to consider. We drew on existing guidelines for process evaluations (Scriven, 2007) and again adapted the CASP checklist to better suit such evaluations. The checklist is included in Appendix D.

Project documents provided information about planned, on-going or completed programmes, providing information about the design or resources available for a project for instance. As such these documents do not typically include much analysis of primary evidence, but they provide factual information about interventions. The purpose of including them in our review was to ensure we had sufficient information about the context and interventions included in our review. Thus, we did not formally appraise the quality of such documents, but rather focused our appraisal on assessing the relevance of the documents against the interventions assessed in our review. Before extracting any data we ensured that the name of the intervention, the implementing agency, context and timeline of the intervention described in the project document corresponds to the intervention assessed in the impact evaluation included in our review. Finally, collecting data from a range of sources, especially if used for triangulation, can enhance confidence in the trustworthiness of the information included (Montgomery *et al.,* forthcoming). If several sources were available we extracted data from all sources for purposes of triangulation. If we were in doubt regarding the relevance of a particular document, we contacted the authors.

2.5 Effect size calculation

We extracted data available in included studies to calculate standardised effect sizes, as detailed in our data extraction tool. We had planned to calculate a partial *d* using formula provided by Keef and Roberts (2004). However, while the majority of studies reported regression coefficients, few studies provided the standard deviations required for using this formula. Therefore the decision on how to standardise effect sizes was made taking into account what had been reported in the majority of the studies sharing common outcomes.

Based on the availability of data we decided to calculate standardised mean difference (d) using formulae the following formula, where *n* denotes the sample size of treatment group (t) and control (c):

$$d = \frac{2t}{\sqrt{n_t + n_c}} \qquad Var_d = \frac{2}{n_t + n_c} + \frac{d^2}{4(n_t + n_c)}$$

We calculated the t-statistic (*t)* by dividing the coefficient by the standard error. If the authors only reported confidence intervals and no standard error we calculated the standard error from the confidence intervals. If the study did not report the standard error, but reported *t* we extracted and used this as reported by the authors.

For studies reporting other data than coefficients and standard errors we used different formula to calculate d, as reported below:

Studies reporting mean differences ($\Delta\bar{X}$) between treatment (T) and control (C) and standard deviation (SD) at follow up (p+1) :

$$d = \frac{\Delta\bar{X}_{p+1}}{SD_{p+1}} = \frac{\bar{X}_{Tp+1} - \bar{X}_{Cp+1}}{SD_{p+1}}$$

Studies reporting mean differences between treatment and control, standard error (SE) and sample size (n):

$$d = \frac{\Delta\bar{X}_{p+1}}{\mathrm{SE}\sqrt{n}}$$

Studies reporting means and standard deviations for treatment and control groups at baseline (p) and follow up:

$$d = \frac{\Delta\bar{X}_{p} - \Delta\bar{X}_{p+1}}{SD_{p+1}}$$, where

$$SD_{p+1} = \sqrt{\frac{(n_{Tp+1} - 1)SD^2_{Tp+1} + (n_{Cp+1} - 1)SD^2_{Cp+1}}{n_{Tp+1} + n_{Cp+1} - 2}}$$

$$Var_d = \frac{n_{\mathrm{T}} + n_{\mathrm{C}}}{n_{\mathrm{T}} n_{\mathrm{C}}} + \frac{d^2}{2(n_{\mathrm{T}} + n_{\mathrm{C}})}$$

Studies reporting proportions (r) in treatment group and control:

$$d = \ln\left[\frac{r_{\mathrm{T}}(1-r_{\mathrm{T}})}{r_{\mathrm{C}}(1-r_{\mathrm{C}})}\right]\frac{\sqrt{3}}{\pi}$$

Finally, a few studies only provided exact p-values and sample sizes. For these studies we calculated *d* using the t-test p-value, unequal sample size formula provided in the Practical Meta-Analysis Effect Size Calculator (Wilson, n.d).

Our data set included both extremely small and extremely large sample sizes. Therefore we Winzorised the sample sizes (Lipsey and Wilson, 2001) so that if the sample size of any single effect size was less than or equal to the 1st quintile of the overall sample size for the data set for any single meta-analysis the sample size for that effect size was made equal to the first quintile value. Similarly, at the top end, if the sample size was larger than or equal to the 3rd quintile value the sample size for that effect size was made equal to the 3rd quintile value. This value was then used for calculating the variance for the effect size.

2.5.1 Dependent effect sizes

A range of issues may cause there to be dependencies between effect sizes. For instance, there are at times several publications from one study, or several studies based on the same data set. Some studies report outcome measures from several time points, or use multiple outcome measures to assess related outcome constructs. Finally, many studies include multiple treatment arms with only one control group. We do not treat all outcome estimates as independent of each other (Borenstein *et al.*, 2009) and only include one effect estimate per independent sample in any single meta-analysis. The exception to this was if we had more than 10 studies for any one intervention-outcome combination. In such cases we included dependent effect sizes and applied robust variance estimation in the meta-analysis which allows for the inclusion of dependent effect sizes (Hedges *et al.,* 2010; Tanner-Smith, E. & Tipton, E. (in press)).

When selecting effect sizes we used the following rules: If we had several publications reporting on the same study we assigned the most recent publication to be the 'main impact evaluation' and extracted effect size data from this publication (only including estimates from earlier publications that were not included in the most recent version). For studies with outcome measures at different time points we selected the follow up period that was most similar to the other measures included in any single meta-analysis. If studies reported on the same outcome measure for independent samples (e.g. different grades, genders, locations) without providing an estimate for the full sample we included all independent samples in the meta- analysis. If there were estimates from overlapping samples we used the data for the most comprehensive sample.

When studies included multiple outcome measures to assess related outcome constructs, we followed Macdonald *et al.* (2012) and selected the outcome that appeared to most accurately reflect the outcome construct of interest without reference to the results. The exception to this rule was maths and language arts test scores. For these outcomes authors often reported a number of disaggregated measures of the broader outcome construct. If we had less than ten studies for any single meta-analysis we calculated an average or 'synthetic' effect size (Borenstein *et al.,* 2009, Chapter 25). If studies only reported disaggregated outcomes (e.g.: different components of literacy) without reporting an aggregate effect size we calculated an average or 'synthetic' effect size which we then included in the meta-analysis (Borenstein *et al.*, 2009, Chapter 25). Alternatively, if we had sufficient studies to use robust variance meta-analysis we included all disaggregated outcomes measures in the meta-analysis.

When several specifications are reported in one study we included the specification preferred by the author. When this was not explicit we included the specification with most controls. If studies included multiple treatment arms with only one control group and the treatments represented separate treatment constructs, we calculated the effect size for treatment A versus control and treatment B versus control and included in separate meta-analyses according to the treatment construct. If the treatments A and B represented variations of the same treatment construct we selected the treatment arm that was most similar to the other treatments included in any single meta-analysis. If we had more than ten included studies and multiple treatment arms represented variations of the same treatment construct we included more than one treatment arm and used robust variance meta-analysis as above.

For the cash transfer intervention area we identified several large scale programmes that have been rigorously evaluated. These studies produced large data sets which have been made available to many study teams, resulting in a large number of studies analysing the same data sets in different ways. For Progresa/Oportunidades alone we identified over 50 documents which met our inclusion criteria. We could have coded all these studies and used robust variance estimation as outlined above. However, for practical reasons we decided it was better use of our resources to select one study per programme. Therefore we developed additional criteria to select which study to include in our analysis. The general decision rule we adopted was to choose a primary impact evaluation when more than one study evaluated the same cash transfer programme as follows:

1) Prioritise RCT with DID estimation, followed by RCTs without DID estimation. This decision was based on the most commonly occurring study design within the cash transfer category, allowing us to have the largest number of comparable study designs included in the meta-analyses.

2) If this did not help identify the primary impact evaluation we chose the study that used the most comprehensive data set in terms of sample size and scope (broadest age / grade / geographic coverage / data sources).

For Progresa/Oportunidades, we chose the original published study as the primary impact evaluation (this also used an RCT+DID design) and then followed rules 1) and 2) above to select studies for data extraction for other relevant outcomes not reported in the primary impact evaluation. We extracted all includable outcomes from the primary impact evaluation and data was only extracted from the remaining studies evaluating the same programme if these were additional to the ones reported in the primary study (following the same rules if a choice had to be made between several additional studies).

2.5.2 Unit of analysis

We assessed studies for unit of analysis errors, where the unit of the treatment is different to the unit of analysis, without taking account of clustering in the analysis (The Campbell Collaboration, 2014). Most regression based studies already correct their standard errors for clustering. We did not correct the standard errors of the studies that did not account for differences in the unit of treatment allocation and unit of analysis.

2.5.3 Missing Data

Where included studies did not provide the data required to calculate effect sizes, we contacted the authors of the primary studies to get access to the missing information. In cases where we were not able to obtain sufficient data from authors to calculate effects sizes we excluded the study from the meta- analysis. However, we still included of the study in our descriptive findings and qualitative synthesis.

2.6 Methods of synthesis

2.6.1 Review questions 1a, 1b: Statistical analysis

We synthesised evidence on the effects of education interventions to address review questions 1a and 1b. If meta-analysis was feasible, we synthesised studies using an inverse-variance, random effects model due to the anticipated heterogeneity in our included studies. By accounting for the possibility of different effect sizes across studies, random effects meta-analysis produces a pooled effect size with greater uncertainty attached to it, in terms of wider confidence intervals than a fixed effect model (Higgins & Green, 2011). We present the syntheses ordered by where the outcome falls on the causal chain, from intermediate to final outcomes.

We only conducted meta-analysis of studies which we assessed to be sufficiently similar. The studies included in our review evaluate the effect of interventions falling under the broad category of being an education intervention, but we included studies that assess different treatment constructs. Therefore we did not pool the results of all studies in a single meta-analysis, and we only synthesised findings from studies with comparable intervention constructs. We followed the approach adopted by Wilson *et al.* (2011) and conducted meta-analysis for interventions where we identified two or more studies with comparable effect-sizes for a common outcome construct and where the condition in the comparison group was judged to be similar. All statistical analyses were conducted using R software (R Development Core Team, 2008).

Assessment of heterogeneity

We assessed heterogeneity of effect sizes graphically, and tested for heterogeneity formally by calculating the Q-statistic, the I^2, and Tau^2 to provide an overall estimate of the amount of variability in the distribution of the true effect sizes (Borenstein *et al.,* 2009).

Moderator analyses

We conducted a moderator analysis at the review level for primary outcomes to explore possible sources of heterogeneity in combined effects. We used meta-regression for primary outcomes to assess the potential impact of the following study level variables: (1) Methodological variables: study design and risk of bias; (2) Substantive variables: region, country income classification, type of implementing agency (government pilot, government, NGO, research team) and length of follow- up.

We also conducted sensitivity analysis to assess whether our results were sensitive to the removal of any one single study. [3]

We were also able to conduct moderator analysis for cash transfer interventions due to the relatively large number of studies identified for this intervention. We assessed the potential role of the following characteristics specific to cash transfer interventions: intensity of condition (using a scale from 1-6, as developed by Baird *et al.,* 2013); recipient of transfer (mother, father, child, grandparent); targeting of cash transfer (based on poverty/ income, specific age groups, gender) and size of the transfer.

To allow us to compare cash transfer amounts across interventions, the amounts given were standardised to an annualised value with denomination per person and in international dollars. Specifically, conversions from local currency rates were applied using the available purchasing power parity (PPP) data from the World Bank Data Catalogue. One-off annual amounts were also added to the annualised value of regular cash transfers to create a total annual transfer value. For example this includes transfer items such as an annual school materials grant which complimented a regular monthly cash transfer. Where the amount of a cash transfer was reported per household and the transfer is a “general” cash transfer to the household, the transferred amount was divided by the estimated average household size of the country of origin using data from the Demographic and Health Survey Programme Database. If the amount of the transfer was reported per household and the transfer was an “education specific” cash transfer, the amount was divided by the estimated average number

[3] Results of the sensitivity analysis are reported in the text for each intervention levels and in full in appendix H. The results for the sensitivity analysis in appendix H where we remove each study one by one from the meta-analysis do not follow the order of the forest plots – please contact the authors if you are interested in more information regarding the ordering of studies in the sensitivity analysis in Appendix H. We discuss any interesting findings from the sensitivity analysis in text.

of children aged 5-19 per household, again using estimates from the Demographic and Health Survey Programme Database. To calculate the estimated average number of children aged 5-19 per household, estimates from the average household size were multiplied by the estimated average percentage composition of a household that lie in this age group. Where conversion and survey estimates were not directly available for the exact years of intervention activity, we used estimates from surveys closest to the time of intervention delivery and available PPP data.

Publication bias

We attempted to reduce publication bias by searching for and including unpublished studies in the review, but we also assessed possible publication bias and under-reporting of small sample studies using funnel plots and Egger *et al.*'s (1998) test.

2.6.2 Review questions 2a and 2b: Qualitative synthesis

After having completed the detailed coding of all of the included studies as described above, we re-reviewed the coding of data on context, intervention design and implementation to identify descriptive findings which remain close to the findings in the primary studies (following Thomas and Harden, 2008). Due to the quality and quantity of additional data identified we were unable to conduct the interpretive synthesis outlined in our protocol. Instead we reported the descriptive findings from the qualitative and process evidence in summary tables and integrate the findings in our discussion when relevant. We provide the full text of the qualitative synthesis in Appendix J.

2.7 Summary of findings

We produced summary of findings tables for each intervention, reporting details of sample size, average effects and confidence intervals, and heterogeneity statistics for all primary outcomes. To present readers with a more interpretable estimate of effects we converted the SMDs into more practical and intuitive values using an improvement index method which is based on Cohen's Index (see Cohen, 1977; 1988), as described by Durlak (2009) and Sussman (2001). This measure shows the estimated change in percentile rank for an average student in the control group if they had received the intervention (Brewer and Picus, 2014). The approach uses the z-score corresponding to the estimated effect size, which in turn indicates the proportion of area under the symmetrically normal curve, and then interprets this area in terms of percentiles in order to compare the treatment and control group distributions (Bickman and Rog, 2009; Durlack, 2009). Therefore, if an intervention has no effect (d=0.00) then one assumes only 50 per cent of the treatment group would have scored higher than the control group mean. Applying this method to an estimated Cohen's d of 0.1, one would find ~54 per cent of treated observations would have achieved better than the control group mean. The treatment mean is 4 percentiles higher than the control group mean.

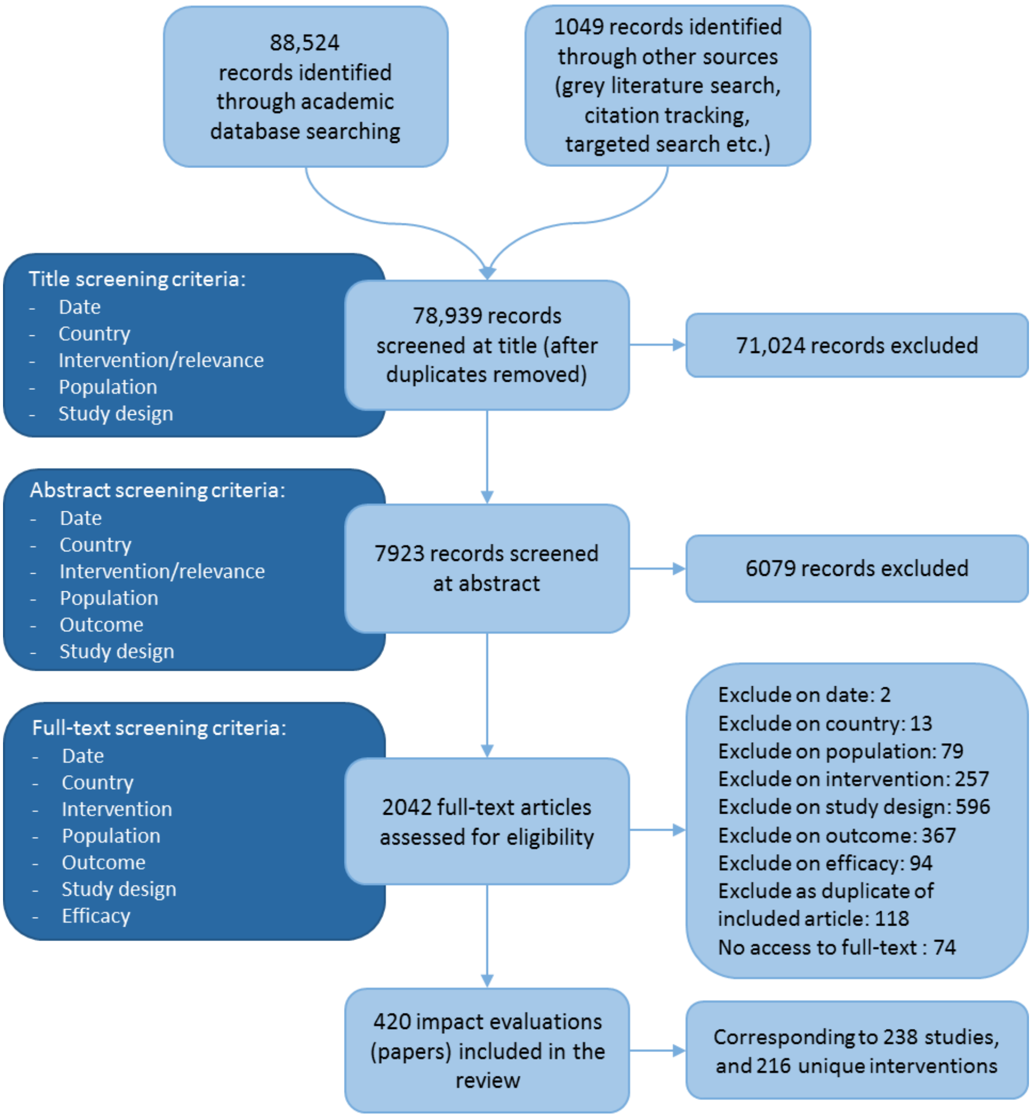
3. Results: search and characteristics of the evidence base

In this chapter we report the descriptive results for the review to provide an overview of the characteristics and distribution of the evidence base across all interventions covered in the review. We start by providing the results of the search and screening of the literature, followed by a section providing a summary of the characteristics of included studies, including the risk of bias and quality appraisal of included impact evaluations and qualitative documents.

3.1 Search results

Figure 3a below provides an outline of the search and screening process to identify included studies.

Figure 3 a: PRISMA diagram:



After removing duplicate records we identified 78,939 citations for screening at title stage. Seven thousand nine hundred and twenty three records were retained for detailed abstract screening by two people. Two thousand and forty two papers were screened at full text by two or more reviewers. Finally, we included 420 papers reporting on impact evaluations meeting our PICOs criteria. These report on 238 studies and 216 unique interventions across all intervention areas. This number of unique interventions is lower as there were cases where we included several studies that use the same dataset but had different author teams undertaking different analysis or reporting additional outcomes.

The number of included papers is much larger than the number of included studies. This is because it is common for studies to be reported in more than one paper, typically one or more working papers and a journal version. As noted in Chapter 2, our approach was to make the most recent version the main paper, and then include any other version(s) of the paper which contained additional information. Typically the working papers and other 'un-published' reports would include more detail than journal versions, including effects on additional outcomes, sub-groups or more detail about the programme and would therefore be included.

3.1.2 Reasons for exclusion

The main reason for exclusion at title stage was a failure to meet the intervention criteria/ relevance of the paper (n= 53,055). That is, the paper was on a topic irrelevant to the review or did not have improving educational outcomes as a primary focus. Six thousand three hundred and twenty records were excluded on population criteria, followed by 6200 on country, 5280 on study design and 69 on date. At the detailed abstract screening stage, where papers were screened by two reviewers, the main reason for exclusion was again intervention/relevance (n = 3754), followed by 1093 on study design, 698 on outcome, 349 on population, 173 on country, and 12 on date.

Figure 3.1a provides a detailed breakdown of the reasons for exclusion at full text. The most common reason for excluding studies at full text was that they did not meet our study design inclusion criteria. Of the 596 studies being excluded for this reason, 103 studies did not have a comparison group, 220 studies did not have baseline data on the outcome variable, 272 studies had other design issues such as no control for confounding and one was excluded as an interrupted time series design. The second most common reason for exclusion at full text was that studies did not report on any of the outcomes of interest in our review (n= 367). Two hundred and fifty seven studies were excluded because they evaluated an intervention beyond the scope of our review and 79 studies did not meet our population inclusion criteria. In addition, 94 studies were excluded as they were classified as an efficacy study as per the criteria outlined in Chapter 2. Hundred and eighteen papers were excluded as they were an effective duplicate of another included paper (for example a working paper version of a journal article that did not contain additional information). We were not able to access the full- text of 74 studies. The list of all studies excluded at full text with reasons for exclusion is available on request and will be published as a web appendix.

3.2 Characteristics of included studies

With over 230[4] included studies our review has identified a large number of impact evaluations covering a broad range of education interventions. The individual chapters in the report provide detailed descriptions of the included studies as they apply to specific interventions. Here we describe the characteristics of the evidence base as a whole.

[4] We identified a large number of includable impact evaluations of cash transfer programmes, 156 papers in total. As per the criteria outlined in Chapter 2 we did not use data from all of these papers in the review. These includable but unused studies are not included in our total number of studies or in the write up below. We list the references to all of these studies in a separate section in the reference list below.

3.2.1 Population

The reach of our included studies is global, covering programmes across 52 L&MICs. This included 59 studies from Sub-Saharan Africa, 38 studies from East Asia & the Pacific, 87 from Latin America & the Caribbean; 51 from South Asia, two from Middle Eastern & North Africa and one from Europe & CIS.

Figure 3 b: Coverage of included studies[5]

3.2.2 Interventions

Table 3a provides a list of all included interventions and the corresponding number of studies identified for inclusion. It highlights there is divergence in the extent to which different interventions have been evaluated. The intervention level with the lowest number of included studies is teacher level interventions. Only 19 studies of teacher specific interventions were identified. These evaluated interventions providing teacher incentives, teacher training and different approaches to hiring of teachers.

Table 3 a: Number of studies by intervention type

Intervention Level	Intervention Category	No. of studies
Child level	School-Based Health	16
	School Feeding	16
	Merit Based Scholarships	11
	Providing Information	4
Household level	Reducing/Eliminating Fees	9
	Cash Transfer	49
School level	Computer Assisted Learning	18
	Pedagogy	22
	Extra Time	3
	New Schools and Infrastructure	7

	Providing Materials	4
	Remedial Education	4
	Grade Retention	1
	Tracking	2
Teacher level	Teacher Hiring	8
	Teacher Incentives	10
	Teacher Training	1
	Diagnostick Feedback	2
System level	School Based Management	14
	Community Based Monitoring	11
	Public Private Partnerships	13

3.2.3 Outcomes

The most frequently reported outcomes included for meta-analysis is language arts test scores (n=114), followed by maths test scores (n=126). Reporting of cognitive outcomes was scarce (n=13) in the included studies and was reported mainly for school-feeding and school based health studies.[5] The most frequently reported access outcomes were enrolment (n=74), attendance (n=68) and drop-out (n=46).

Throughout our included studies we consistently found a lack of clear reporting of outcome definitions and varying interpretations of measures for outcomes. This presented a barrier to evidence synthesis. Standardised outcome measures would have made this easier. Few studies reported sub-group analysis and reporting of effects at longer follow up was scarce.

3.2.4 Study designs

Across the entire review, 122, or 52 per cent, of the included studies were cluster-randomised controlled trials (Cluster-RCTs) and eight per cent were randomised controlled trials, where random assignment was done at the individual level. Seven per cent were natural experiments. Eleven per cent of the included studies were regression discontinuity designs and the remaining 23 per cent used a controlled before-after study design, with estimation strategies such as difference-in-difference estimation or propensity score matching to control for potential selection bias. See Table 3b for an overview.

Table 3 b: Overview of included study designs

Study design	**Number**	**% of total**
Randomised Controlled Trial (RCT)	19	8
Cluster RCT	122	51.5
Natural experiment	17	7.2
Regression Discontinuity Design (RDD)	25	10.5
Controlled Before and After (CBA)	54	22.8

[5] These numbers correspond to the number of included studies by outcome for those studies included in the meta-analysis only. Several other studies reported on these outcomes but were not suitable for synthesis.

3.3 Critical appraisal of included studies

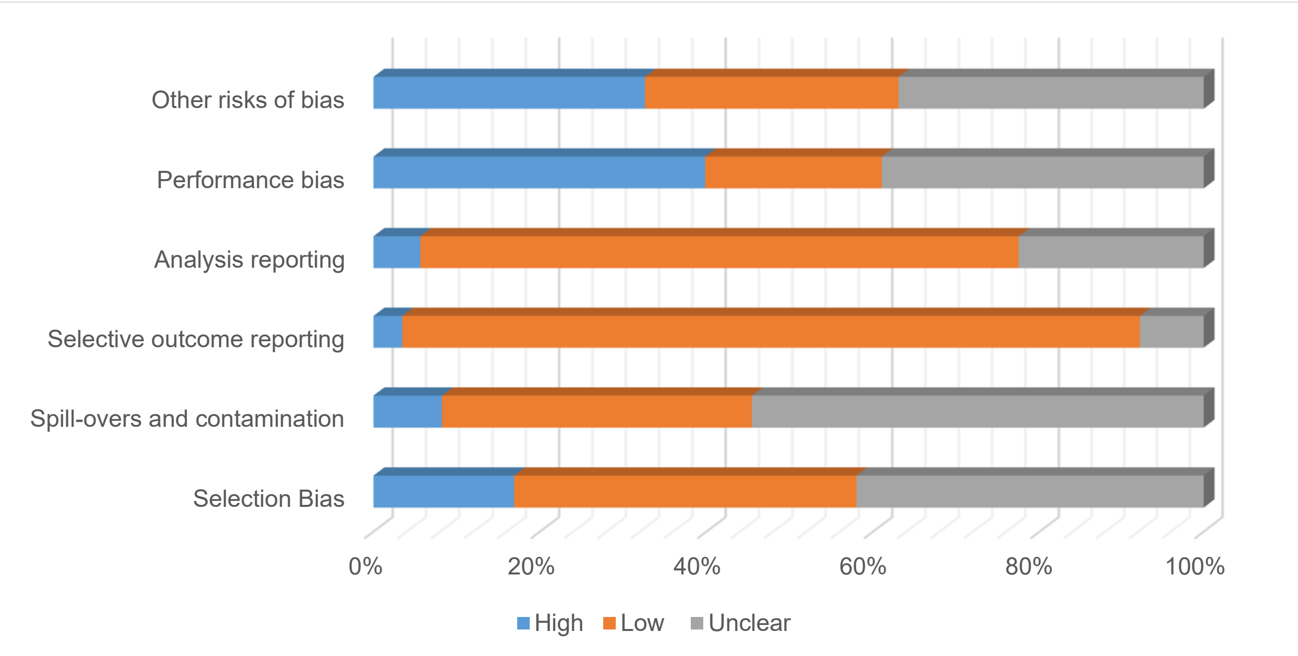
3.3.1 Risk of bias

The risk of bias tool aims to assess the risk of bias to estimates of effects in the included studies based on an assessment of key features of study design and analysis. Figure 3 c provides a summary of the assessment of risk of bias in the included studies, with a breakdown for each domain. Appendix G gives a full summary of our risk of bias assessment at the study level. As we can see, the quality of the included impact evaluations varied.

About half of the included studies (41%) were judged to be of low risk of selection bias and confounding, while for a large percentage of the studies it was unclear (42%). Most studies were judged to have low risk of outcome reporting and analysis reporting bias (89% and 72% respectively), but less than a third (37%) were judged to have a low risk of bias of spill-overs, cross-overs and contamination, and less than a quarter were judged to have low risk of performance bias (21%) and low risk of other biases (30%). In 55 per cent of the studies, it was unclear if the authors had adequately addressed spill-overs, cross-overs and contamination, and so it is difficult to assess if estimated effects over- or under-estimate the true effects of the interventions. Similarly, just under half of the included studies (40%) had a high risk of performance bias. Seventeen per cent were judged to have a high risk of selection bias or suffer from confounding.

In several cases, the reporting in the included studies was not sufficiently clear to make a judgement about particular aspects risk of bias. For example, in 42 per cent of the studies it was unclear whether selection bias and confounding had been adequately addressed, over a third of the included studies (39%) did not report sufficient information to make a judgement about performance bias, and almost half (37%) of the included studies did not report enough information to make a judgement about the risk of other types of biases.

Table 3 c: Overview of risk of bias of included impact evaluations



3.3.2 Critical appraisal of process evaluations and qualitative studies

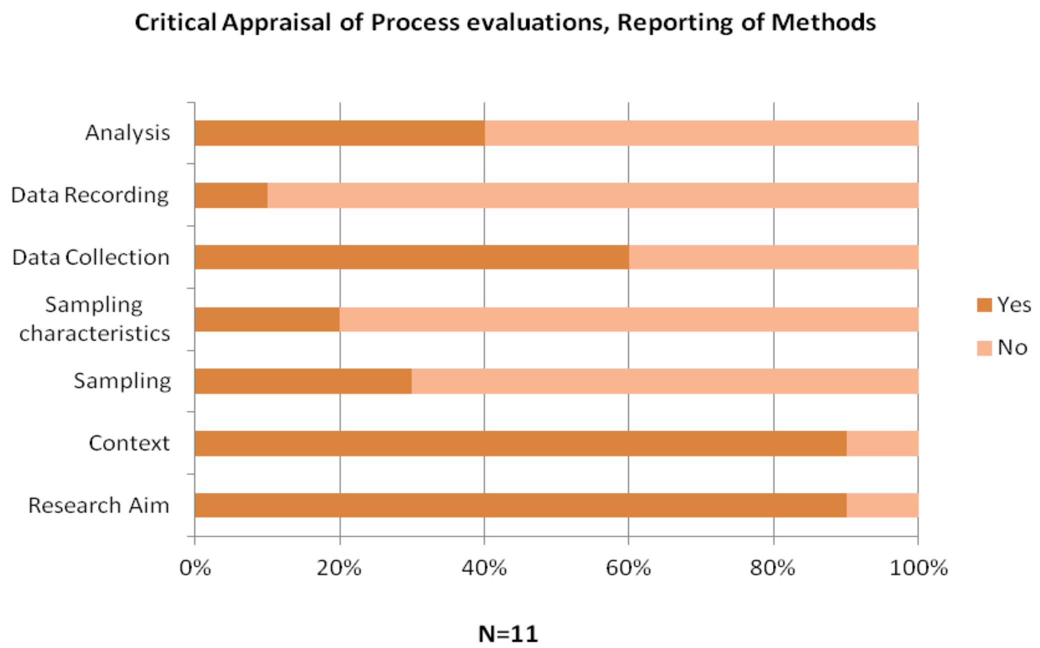
The critical appraisal has been conducted for two types of documents: qualitative studies (including qualitative, mixed-methods and descriptive quantitative studies) and process evaluations. The types of document belonging to the first category are quite diverse in their nature. They can be published articles, PhD thesis or reports. The process evaluations are more similar as they are generally produced by independent monitoring and evaluation agencies following a fairly similar structure.

The appraisal has been conducted on two areas, namely, the appropriateness of the reporting and the appropriateness of the methods. While the first area tells us whether the study provides basic information on sampling design, sampling characteristics, data collection, the second area tells us whether the methodological choices related to sampling, data collection and analysis have been appropriate in relation to the objective of the study. A brief description will follow for the critical appraisal of process evaluations and qualitative documents separately.

Critical appraisal of process evaluations

The figure below gives us information on the reporting methods of the process evaluations included in the review. As we can see, 90% of the studies provided information on the context and on the research aims. Information on the instruments used for data collection is reported only in 60% of cases. Information on all the remaining characteristics is not provided in most of the cases. For instance, information on how the data was recorded once collected is given by 10 % of the documents. Details on sampling and characteristics of the samples are provided for respectively 30% and 20% of the documents. Information on how the analysis was conducted was also provided for 40% of the documents.

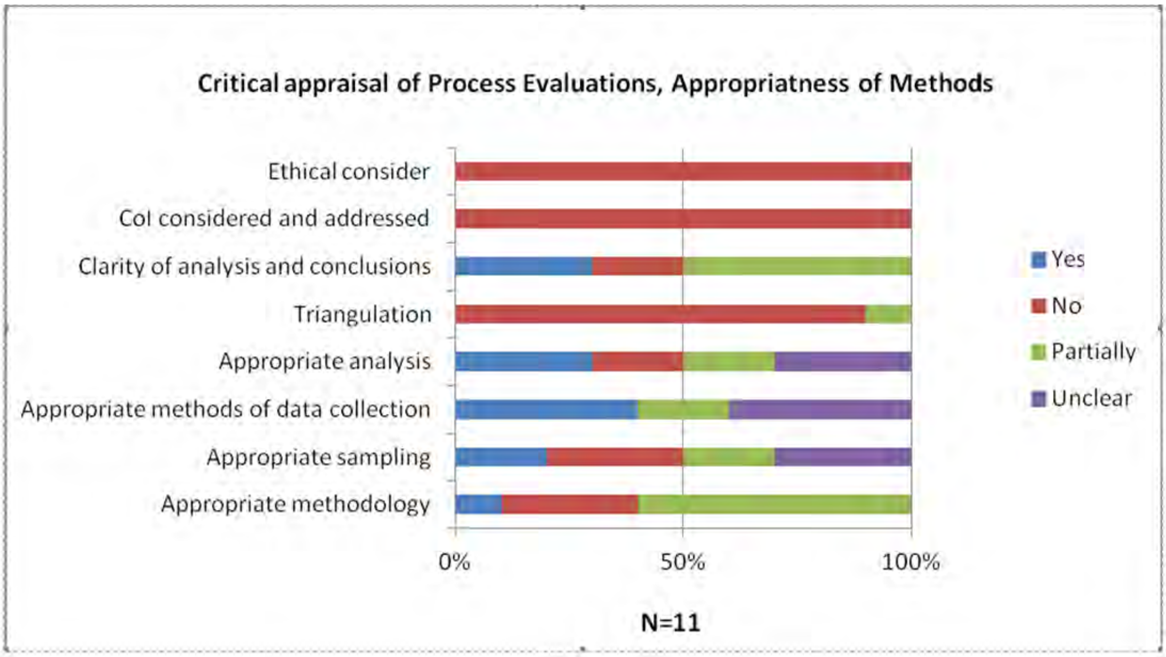
Table 3 d: Overview of critical appraisal of process evaluations (reporting methods)



The results do not change much for the appropriateness of the methods chosen. As we can see in the figure below, most of the studies fall between ‘No’, ‘Partially’ and ‘Unclear’. The only case where we at least 30 per cent of the documents have been appraised to choose an appropriate methodology is for data collection. By this, we mean that the method/s chosen for collecting the data, were appropriate to answer the research questions. For example, if the study wanted to investigate whether the teachers had changed their teaching methods as a result of a particular intervention, classroom observation appears to be a more appropriate method than interviews. In 20 per cent of the cases the documents justify information on the sampling. That is, most of the studies did not explain how they sampled the participants or why those participants were the best ones to provide relevant information on their research questions. Similarly, only 10 per cent of the evaluations justified their methodology over other possible choices. Information on analysis is also weak.

Very little information was provided on clarity of analysis and conclusions. While the results or findings were normally discussed, only in 30 per cent of cases was there an explicit link with the research questions the study started with. Additionally, studies rarely discuss limitations or potential biases in the interpretation of their findings. Finally, few of the studies provided details on triangulation, ethical consideration or potential conflict of interest.

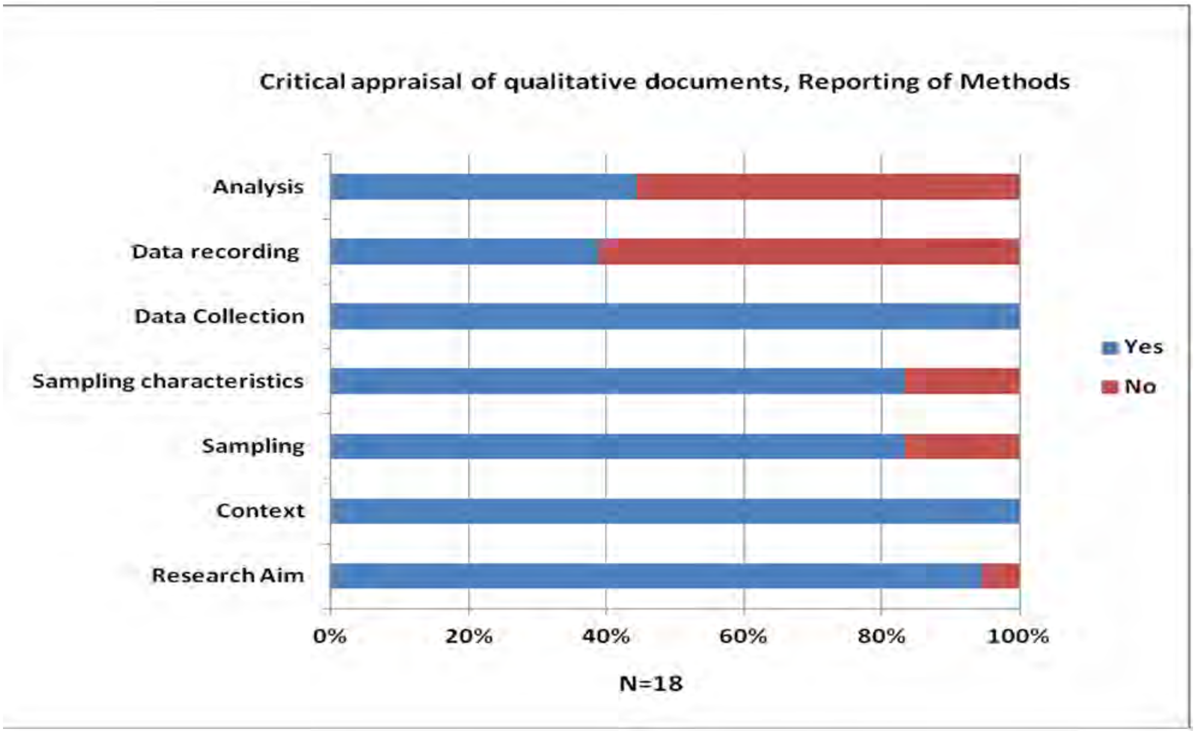
Table 3 e: Overview of critical appraisal of process evaluations (appropriateness of methods)



Critical appraisal of qualitative studies

Most of the qualitative documents provide information on basic information such as research aims, context sampling, data collection and sampling characteristics (see Figure 3 f). However, information about how data was recorded is provided for less than 40 per cent of the documents. Information on how the data was analysed is available for less than 40 per cent of the studies.

Table 3 f: Overview of critical appraisal of qualitative documents (reporting methods)

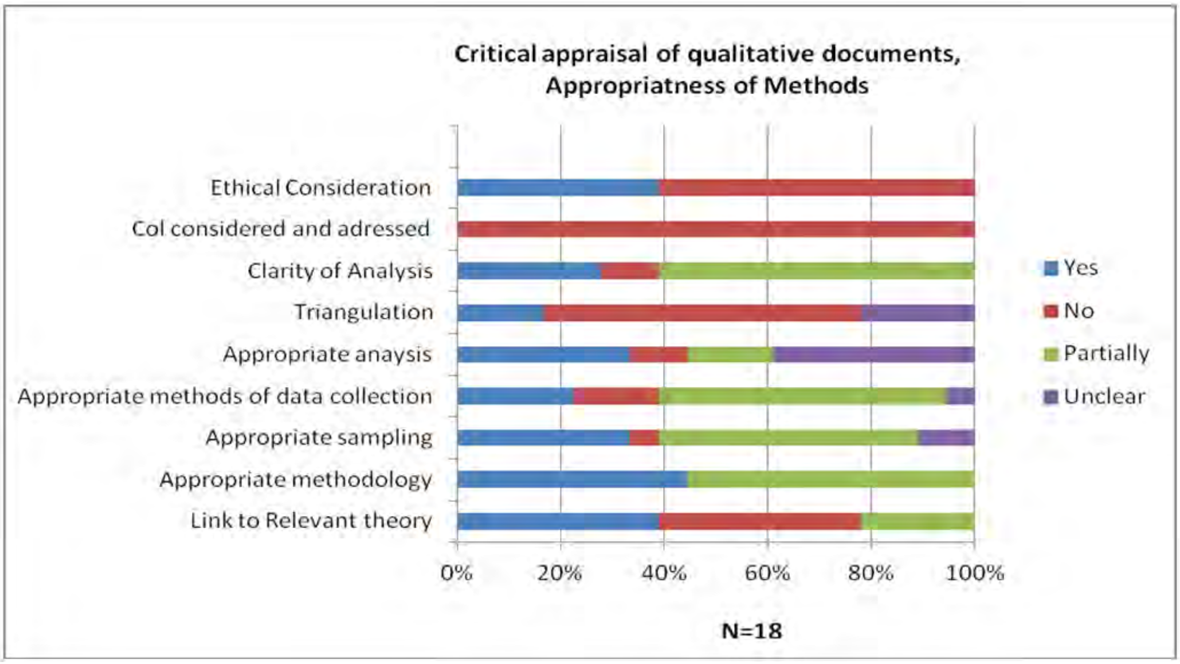


While qualitative studies score relatively well on the critical appraisal of reporting of methods, the situation is different for appropriateness of the methods. As we can see in Figure 3 g, the area which was most discussed was appropriateness of methodology. However, this is the case for only around more than 40 per cent of the studies. In this case, most of the studies provided partial information. A link to the relevant theory used for data analysis was provided by almost 40 per cent of the studies. However, only 30 per cent of the studies provided information on how the data was analysed. This means that it is not clear how the authors interpreted the data to support the findings.

Similarly, only 30 per cent of the studies mentioned how the participants were sampled and why these had been identified as the most relevant persons to provide the information sought by the authors. For this category, most of the studies reported partial information. This means the sampling strategy may have been appropriate, but the study did not explain why some people rather than others were selected for inclusion in the study.

Barely 20 per cent of the studies mentioned triangulation. In this case, it is important to specify that if the authors did not explicitly mentioned triangulation but collected data in a more comprehensive way (that is, through classroom observations, interviews and/or document review) the study was still considered as having used triangulation. Ethical considerations, such as issues of anonymity or potential consequences arising from data collection were mentioned by 40 per cent of the studies. Similarly to the process evaluations, no studies addressed or mentioned whether there may have been any conflict of interests.

Table 3 g: Overview of critical appraisal of qualitative documents (appropriateness of methods)



3.3.3 Quality of reporting in included studies

An issue with many of the included studies, and a persistent challenge for making the best use of the evidence base, is the poor quality of reporting in study reports. This is reflected in the relatively large share of studies where one or more domains of risk of bias for instance are rated as 'unclear'. Details about the randomisation procedure are often lacking and it is not clear which covariates have been controlled for in regressions. Similarly, many studies do not report any details on attrition, including level of attrition and characteristics of attritors in treatment and comparison groups.

Another issue is the lack of reporting of sample characteristics and statistical outputs. For many studies we were not able to identify details about sample characteristics, including key features such as gender, age and grades of students. Studies often fail to report, or do not report very clearly, the sample size associated with different effect estimates. Additionally, many studies do not report the information required to transform estimates into a common metric for the purposes of meta-analysis and cross study comparison. Studies typically report p-values and/or statistical significance, but many do not include standard deviations, standard errors or confidence intervals. We have tried to address this issue my writing to authors to request this information, but this is labour intensive, in particular for a review of this scale. Several studies have been excluded from meta-analysis because of a lack of usable data (Newman *et al.*, 2002; Bellei, 2009; Simwaka *et al.,* 2009; Grigorenko *et al.*, 2007).

A lack of clear definitions of key study constructs, such as outcomes is another limitation. For instance, studies may report that they measure 'language' without specifying how this was defined and indeed which language this refers to. Finally, reporting of intervention design and implementation is often limited, so it can be difficult to assess what was delivered, by whom and at what cost. Issues with reporting quality is not confined to impact evaluations. In particular reporting of methods is even worse for qualitative studies, as we have seen in section 3.3.2.

Conclusions

These issues with the quality of reporting limit both the usefulness of study reports and our confidence in the results of studies. It also adds to the time needed to identify, appraise and extract data from studies. To address issues related to the lack of reporting of details of intervention design and implementation we conducted a targeted search to identify studies and documents with this information. To address issues with the lack of reporting of complete statistical data we have written and requested this information from authors. While many authors were helpful in providing this information, four studies will not be included in the statistical analysis because of a lack of necessary data to calculate effect sizes (Newman *et al.,* 2002; Bellei, 2009; Simwaka *et al.,* 2009; Grigorenko *et al*., 2007).

4. Child- Level Interventions

We define child-level interventions as those interventions targeting children directly, focusing on improving their ability to benefit from schooling or their incentives and motivation for investing time and resources in their own education. We have included school- based health interventions, school feeding interventions, providing information to children and parents and merit based scholarships in this category.

This chapter provides the findings of our synthesis of the 47 included studies evaluating the effect of these interventions on access to schooling and learning outcomes. The chapter is organised by intervention group. Each sub-section starts with a description of the intervention type and its theory of change, followed by descriptive results and the findings addressing our research questions.

4.1 School- Based Health interventions

Evidence shows that ill health and malnutrition have negative effects on educational outcomes (Glewwe *et al.,* 2008; Jukes *et al.,* 2008; Bundy, 2005). Micronutrient deficiencies and diseases such as malaria and worm infections that affect school- age children in particular pose challenges to gains in education. They may result in fewer years being enrolled in school, lower attendance rates and reduced ability to learn when in school. Recognising the significant impact of ill health on educational outcomes, school- based health (SBH) programmes have gained increasing attention over the past decades. This is reflected in the initiation of a number of international partnerships and programmes, including the launch of the framework for focusing resources on effective school health (FRESH) developed jointly by UNESCO, WHO, UNICEF, The World Bank and Education International (Jukes *et al.*, 2008). This consensus framework describes health as 'an input and condition necessary for learning, as an outcome of effective quality education and as a sector that must collaborate with education to achieve the goal of Education for All.' (Bundy *et al.*, 2006: 1097). The focus in this review is on a range of SBH interventions to prevent or treat illness that are delivered to children at school with the aim of improving education outcomes. Treatment-based SBH interventions include deworming for intestinal worms and schistosomiasis (Simeons *et al.,* 1995; Ebenezer *et al.*, 2013) and the treatment of malaria (Simwaka *et al.,* 2009; Fernando *et al.,* 2006). Prevention based interventions may include provision of micronutrient supplements such as Vitamin A, Iodine or iron (Mahawithanage *et al.,* 2007; Wong *et al.*, 2015), or eye tests, followed by provision of eye-glasses (Glewwe *et al.*, 2011).

4.1.1 How may SBH affect education outcomes?

Figure 4.1a below provides an ideal type programme theory, mapping out the causal chain of how SBH interventions may improve education outcomes. The theory of change suggests two main pathways through which provision of SBH may contribute to improved education outcomes:

Receiving the correct treatment regime for the targeted health problems of the programme leads to a reduction in new infections, disease prevalence, malnutrition or anemia and thus increases the physical well-being of students. As a result of this, students are less likely to be absent or dropout due to ill health and better able to enrol and participate in school. SBH programmes may affect a child's cognitive ability, as observed by improvements in skills such as concentration, short-term and long-term memory, auditory and visual processing (Ebenezer *et al.*, 2013; Taylor- Robinson *et al.*, 2012). These abilities may in turn affect a child's efficiency of learning and thus school performance.

Thus, SBH interventions have the potential to improve school performance by increasing the total amount of time a student spends in school, as measured by improved enrollment and attendance, and improve learning abilities while the student is in school, which in turn may lead to improved academic achievement. Miguel *et al*. (2004) argue that two assumptions must hold for a deworming treatment and other SBH programmes to have positive effects on students' test scores: 1) Increased school attendance of previously ill students should not lead to classroom congestion; 2) the presence of previously ill students in the classroom does not impose negative learning externalities on other pupils. Furthermore, as pointed out by Jukes *et al*. (2008), programme success is dependent on the support and participation of a number of stakeholders including parents, teachers association and community leaders who should be consulted and informed throughout the process of the programme.

Figure 4.1 a: School-Based Health Interventions Programme Theory

Inputs

- Financial resources to pay for drugs, material, training and human resources
- School staff/ health staff

Assumptions:
- Long- term funding is available from donors, ministries to guarantee sustainabilit y of program

Activities

1. Deworming for intestinal worms and schistosomiasis
2. Treatment of malaria & provision of insecticide-treated nets
3. Programmes to target micronutrient deficiencies (for example supplemention provision (Vitamin A, Iodine, iron)
4. Referral to youth- friendly clinics
5. Provision of eyeglasses to students with compromised vision

Assumptions:
- Children/ parents/ guardians/ community members are informed about the intervention and are supportive
- Students and parents have incentive for students to go to school to receive treatment
- School committees are trained and informed about intervention to gain support and guarantee sustainability of the project
- School/ health staff are adequately trained on how to diagnose health problems, administer drugs/ micronutrients and identify possible adverse events and have the capacity to take on additional school health activity

Outputs

1. Reduction in prevalence of worm infection
2. Reduction in incidence of malaria
3. Reduction in anaemia and malnutrition
4. Access to specific treatment
5. Improved vision

Assumptions:
- Effective treatment regimen/ micronutrient supplement is used in treating infection/ micronutrient deficiencies
- Students adhere to treatment regimen
- Personnel/school staff administering drugs to students adhere to the right doses and timeframes of treatment

Immediate outcomes

- Increased physical wellbeing (improved nutrition/ health)
- Increased cognitive abilities (concentration, memory, auditory and visual processing)

Intermediate Outcomes

- Improved school attendance and school participation
- Improved cognitive performance (efficiency of learning per unit of time spent in school)

Final Outcomes

- Improved academic achievement (test scores)

Assumptions:
- Increased school attendance of previously ill students does not lead to classroom congestion
- Presence of previously ill students in the classroom does not impose negative learning externalities on other pupils

Characteristics of intervention that may affect project outcomes :
The type of health services provided by the intervention (simple to administer, safe, familiar to students/ parents/ teachers)

Contextual factors that may affect project outcomes
Policy: Multisectoral policies that ensure collaboration between the health and education sector; health and nutrition related school policies
School environment: sanitation facilities, access to safe water, facilities for girls
Household Education Environment:
Other alternative educational inputs: reducing class size, improving school facilities etc.
Disease Profile of the country: Disease prevalence in programme, population, intensity of infections

4.1.2 Description of included studies

We included 16 studies reported in 20 different papers that evaluated the effect of 16 different school based health programmes in a L&MICs. We included more than one paper that evaluated the same programme if the author team undertook different analysis or reported different education outcomes over several papers. For example, we included two papers that evaluated the Treatment of Trichuris trichiura Infections in Jamaica (Simeon *et al.*, 1995a, 1995b); two papers that evaluated the Primary School Deworming Project (PSDP) in Kenya (Miguel *et al.,* 2001, 2004) and three papers that evaluated the Health and Literacy Intervention (HALI) project in Kenya (Brooker *et al.*, 2013, 2015; Halliday *et al.*, 2014).

Population

All interventions were targeted at the primary school level and included students at different ages and grades. For the majority of studies the authors do not specify the school type of the study sample (n= 10). The remaining programmes were targeted at public (government- run) schools only.

Setting

The included studies evaluated 16 different programmes and trials covering countries in Latin America and Caribbean, East Asia, South Asia and Sub- Saharan Africa. Five studies took place in China (Sylvia *et al.*, 2013; Kleiman-Weiner *et al.*, 2013; Luo *et al.*, 2012; Glewwe *et al.*, 2014; Wong *et al.,* 2014), three in Kenya (Miguel *et al.*, 2004; Brooker *et al.*, 2015; Clarke *et al.,* 2008), three in Sri Lanka (Mahawithanage *et al.,* 2007; Fernando *et al.,* 2006; Ebenezer *et al.,* 2013), and one in Malawi (Simwaka *et al.*, 2009), Guatemala (Watkins *et al.*, 1996), Zambia (Grigorenko *et al.*, 2007), Philippines (Jukes *et al.,* 2014) and Jamaica (Simeon *et al.,* 1995).

Most studies included samples from rural areas only (N= 11), two studies included samples from both rural and urban areas within a country (Grigorenko *et al.*, 2007; Simeon *et al.,* 1995). The location is not clear in the remaining two studies (Fernando *et al.*, 2006; Ebenezer *et al.,* 2013; Simwaka *et al.*, 2009).

Intervention

The included studies evaluated a range of different school- based health interventions including de-worming (n= 6), malaria prevention and control (n= 4), micronutrient supplementation (n= 7), the provision of eye glasses (n= 1) and provision of incentives for anaemia reduction (n= 1). The School Health and Nutrition intervention in Zambia, the Deworming/ Iron Supplementation Trial in Sri Lanka and the SHN programme in the Phillippines all provided a combination of de-worming and micronutrient supplementation (Ebenezer *et al.,* 2013; Grigorenko *et al.,* 2007; Jukes *et al.* 2014). Table 4.1b provides details of the intervention design components of included studies as detailed below.

The included SBH interventions differed in terms of treatment duration and frequency. We describe these below by relevant intervention sub-groups.

Multivitamin/ Micronutrient supplementation

The programmes studied by Luo *et al.*'s (2012) and Kleiman-Weiner *et al.*'s (2013) in China provided students with a daily chewable vitamin for five and six months respectively. Similarly, Wong *et al.*'s (2014) study in China assessed the effect of daily iron supplement tablets (containing five milligrams of iron and 20 other vitamins and minerals per tablet) for a period of six months). Students in Mahawathanage *et al*.'s (2007) study in Sri Lanka received three doses of 200,000 IU of Vitamin A every four months. As part of the School Health and Nutrition Study in Zambia, Vitamin A supplementation and treatment for helminth infection was provided annually, while iron supplementation was given on a weekly basis for one, two or three years depending on the treatment arm (Grigorenko *et al.*, 2007). In Ebenezer *et al*.'s (2013) study in Sri Lanka the treatment group received a weekly dose of iron supplementation (tablets containing 200mg of ferrous sulphate equivalent to 60mg of elemental iron) for six months. The Save the Children School Health and Nutrition (SHN) programme in the Philippines administered 60mg iron tablets weekly (Jukes *et al*. 2014).

Malaria Control

The included malaria prevention and treatment interventions used various treatment regimes. In Fernando *et al*.'s study in Sri Lanka, one chloroquine tablet (150 mg of chloroquine phosphate base) was given to each child after a meal under the supervision of a research assistant or teacher at weekly school visits over a nine month period. The Intermittent Preventative Treatment (IPT) Malaria Trial in Kenya involved the periodic mass administration of a full therapeutic course of an antimalaria drug, irrespective of infection status. Children received one dose of sulphadoxine-pyrimeth-amine (SP) and three daily doses of amodiaquine (AQ) on three occasions within a 12-month period. Brooker *et al.* (2015) trialled an alternative school- based malaria control strategy based on intermittent screening and treatment (IST) for malaria as part of the Health and Literacy Intervention (HALI) project. All children were screened for malaria using a rapid diagnostic test (RDT) once a school term. Children found to be RDT- positive were treated with a six dose regime of artemether- lumefantrine (AL) over three days during five rounds of screening (Brooker *et al*., 2015). Simwaka *et al.,* (2009) studied a programme where teachers detected suspected malaria cases and provided age- appropriate doses of sulphadoxine-pyrimethamine tablets and paracetamol to the students at school and to take home.

De-worming

In Ebenezer *et al.'*s study in Sri Lanka the treatment group received one 500- mg oral dose of mebandazole. As part of the School Health and Nutrition Study in Zambia children received annual treatment for intestinal worms with albendazole and schistomiasis with prziquantel (Grigorenko *et al.,* 2007). In the case of the Watkins *et al.* study in Guatemala, students received albendazole (2x200 mg tablets) twice, at baseline and 12 weeks later. In Miguel *et al*.'s (2004) study in Kenya, schools with geohelminth prevalence over 50 per cent were mass treated with albendazole every six months (600mg in 1998; 400mg in 1999), and schools with schistosomiasis prevalence over 30 per cent were mass treated with praziquantel annually (40mg/kg). The de-worming study in Jamaica treated children with T. trichiura infections with 800mg of albendazole (400mg on each of two days) given at baseline measurements, 12 weeks later, and 24 weeks later (Simeon *et al.*, 1995). As part of the Save the Children School Health and Nutrition (SHN) programme in the Philippines children received one 400mg oral dose of albendazole at baseline and three months later after the first follow up survey (Jukes *et al.* 2014).

Incentives for Health Promotion

The Anaemia Reduction Programme in China differs to the other included programmes in that it tested whether providing school principals with pay- for performance contracts tied directly to health outcomes of children in their school could increase the effectiveness of a school based anaemia reduction programme (Sylvia *et al.,* 2013). Schools in the trial were allocated to either (a) a 'subsidy' group in which school principals were given information about anaemia and a subsidy of 1.5 yuan per students per day earmarked for anaemia-related expenses; (b) a 'health incentive group' were school principals were provided with a pay- for- performance contract based on school- level anaemia prevalence in addition to the information and subsidy. The intervention allowed educators control over which anaemia reduction strategy to implement and it is not explicitly stated what the subsidy was used for (Sylvia *et al.* 2013).

Vision Correction

The Gansu Vision Intervention Project (GVIP) in Western China provided eyeglasses to primary school students with poor vision in two counties (Glewwe *et al.* 2014)

Additional Features

Several included studies provided training and incentives for school stakeholders to undertake SBH activities alongside the specific health intervention Luo *et al*.'s, Wong *et al*.'s and Kleiman- Weimar *et al*.'s trials in China all provided a small honorarium to the principal and teachers for their services in distributing the multivitamins to the students (100 yuan- about the equivalent of one to two days' salary). Simwaka *et al*.'s (2009) study in Malawi included training for teachers and school health committees to ensure successful implementation of the Malaria treatment programme by the school staff. Two to three teachers from each school participated in a five- day training on how to diagnose and treat malaria. The training sessions also focused on developing teachers' skills to persuade students to take their medicine and come to school if sick.

The one- day training for school health committees intended to help them understand, support, and sustain the project from the community perspective. In addition, students' received pupil's treatment kits, which also included tetracycline (for eye infection) and iodine and dressings for minor cuts and wounds. Similarly, as part of the Primary School Deworming Project in Kenya, treatment schools received worm prevention education through regular public health lectures, wall charts and training of teachers. A series of community and parent meetings were held at which the project was described and parents who did not want their child to participate in the project were asked to inform the school headmaster (Miguel *et al.,* 2004). In the case of the iron supplementation trial in China, teachers received training and training material (Wong *et al.,* 2014). The authors also report that the teachers delivered health education in addition to the treatment (although content and duration of the treatment is not clearly defined).

Table 4.1 a: Intervention Design Features of Included Studies

	Luo (2012)	Kleiman - Weiner (2013)	Mahawithanage (2007)	Wong (2014)	Clarke (2008)	Simwaka (2009)	Brooker (2015)	Fernando (2006)	Ebenezer (2013)	Grigorenko (2007)	Jukes (2014)	Watkins (1996)	Miguel (2004)	Simeon (1995)	Glewwe (2014)	Sylvia (2013)
Intervention design features																
MicronutrientSupplementation	✓	✓	✓	✓					✓	✓	✓					
Malaria Prevention and Control					✓	✓	✓	✓								
De-worming									✓	✓	✓	✓	✓	✓		
Vision Correction															✓	
Incentive for Health Promotion																✓
Additional Features	✓	✓		✓		✓				✓			✓			
Implementation features																
Delivered by Teachers (1), Health staff (2), Researcher (3)	1	1	1 and 3	1	Not clear	1	1 and 2	1 and 3	1	1	1 and 2	Not clear	2	Not clear	2and 3	Not clear
Length of treatment	5 months	6 months	13 months	8 months	12 months	Approx. 4 years	9/24 months	9 months	6 months	36 months	3 months (FU1)	3 months	12- 24 months	24 weeks	10 months	7 months
Intervention Frequency/ Intensity	Daily (one multivitamin tablet containing iron and 20 other vitamins and minerals)	Daily (one multivitamin tablet containing iron)	Every 4 months (3 doses, of Vitamin A capsule (200 000 IU of Vit A)	Daily (one multivitamin tablet containing iron and 20 other vitamins and minerals)	Three rounds of IPT (one dose of sulfadoxine-pyremethamine and three daily doses of amodiaquine)	Not clear; doses of sulphadoxine-pyrimethamine tablets etamol	Five rounds of screening and treatment (Artemether-lumefantrine)	Weekly (150 mg of chloroquine phos-phate base)	Deworming(500 mg of mebandazole) once; Iron Supplementation (200 mg of ferrous sulphate):Weekly for 6 months	Deworming Albendazole and praziquantel, annually); Vitamin A: Annually; Iron Supplementation: weekly	Deworming400mg Albendazole: 400mg; 60mg Iron Supplementation: weekly	Twice (400mg albendazole tablets)	Every 6 months: (Albandazole: 1998: 600mg; 1999: 400mg Praziquantel: annually: 40mg/kg	Three rounds of treatment (400mg albendazole on each of 2 days)	Once	Not clear, subsidy and training seems to be have given once

Comparisons

Eight of the included studies compared the effect of an intervention to a comparison group with no intervention. For two studies, the comparison group received one component of the SBH intervention: In Luo *et al.*'s study in China, principals and teachers in the control schools did not receive the health intervention but still received the same honorarium as the teachers in the intervention group for distributing the vitamins and eggs. The School Health and Nutrition Study in Zambia provided health education sessions to the children in the control group (Grigorenko *et al*., 2007). The remaining six studies were placebo- controlled studies (Mahawithanage *et al.,* 2007, Clarke *et al*., 2008, Ebenezer *et al.*, 2013, Watkins *et al.*, 1996, Simeon *et al.,* 1995, Fernando *et al*., 2006). Simeon *et al.*'s study included an additional control group that received no intervention.

Several of the studies included multiple treatment groups. Miguel *et al*.'s (2004) study in Kenya had two treatment groups receiving the same treatment in 1998 and 1999 respectively, compared against a group receiving no treatment. Similarly, Grigorenko *et al*.'s (2007) study in Zambia has three treatment groups receiving the same treatment in year one, two and three of the intervention. As mentioned above, the anaemia reduction programme in China had two treatment arms testing two variations of the intervention. Wong *et al*.'s study included an additional treatment arm to the iron supplementation programme that trained parents about nutrition and anaemia through power point, video and colourful booklets.

Outcomes

The included studies reported on a number of primary education outcomes, including student attendance (n= 6), completion (n=1), dropout (n=1) and different measures of student learning (n= 16). Student attendance was either measured as the number of days absent from school based on school records (Simwaka *et al.*; Fernando *et al.*), number of days present on a given day (Watkins *et al.;* Simeon *et al*.; Miguel *et al.*) or defined as causes of absence based on letters of excuses sent by the students (Mahawathinage *et al.*). Simwaka *et al.* is the only study reporting on completion and dropout.

The majority of studies (n= 8) measured student learning through written test scores of individual subjects (maths, local or official language). These tests were either standardized national achievement tests or developed by the study team. Two studies reported a composite test score (Miguel *et al.*; Glewwe *et al.*). Six studies assessed cognitive abilities of students such as a child's sustained attention and verbal reasoning (Clarke *et al.*, 2008; Ebenezer *et al.*, 2013; Brooker *et al.*, 2015; Simeon *et al.,* 1995; Grigorenko *et al*., 2007; Jukes *et al.*, 2014). A number of studies (n= 9) also reported intermediate outcomes representing different pathways through which SBH may influence final education outcomes. These included health status (n=3) and nutrition outcomes such as anaemia and haemoglobin status (n= 11).

Study Design

All but one of the included studies were randomised controlled studies where the intervention was randomised at either the student or school level. The exception is Simwaka *et al.*'s (2009) who used a controlled before and after design with difference in difference and propensity score matching.

Qualitative studies, process evaluations and project documents

We only identified two additional documents (two project documents) related to one of the programmes covered by the included studies, The School Health and Nutrition Study in Zambia (Grigorenko *et al.*, 2007). The lack of qualitative information for this intervention type may be due to the fact that the majority of included studies are studies evaluating a one- off trial to inform future roll out of a government-run programme.

Table 4.1 b: Characteristics of Included Studies SBH

Included study	Population	Intervention summary	Included outcomes	Follow- Up	Study design	Sample Size
Luo *et al.* (2012) **Micronutrient supplementation**	China (rural), Primary school (Grade 4)	As part of the Multivitamin treatment arm of the study, children were given one tablet of iron supplements per day for five months. Children were also given vitamins to take home at the weekend. Teachers were trained on how to dispense the vitamins and provided with posters. In all schools, principals and teachers received a small honorarium,	Learning; Nutrition	7 months	Cluster- RCT	3,661 students
Kleiman-Weiner *et al*. (2013) **Micronutrient supplementation**	China (rural), Primary school (Grade 4; Age: 10 yrs)	The chewable vitamin treatment arm provided vitamins (containing micronutrients including iron) to students. Children were also given vitamins to take home at the weekend. In all schools principals and teachers received a small honorarium.	Learning; Health	7 months	RCT	1304 students
Wong *et al.* (2014) **Iron Supplementation trial**	China (rural) Primary School (Grade 4; Age: 9-12	Iron Supplementation trial- (1) Children were provided with iron supplement tablets every day for a period of six months. (2) All fourth- grade homeroom teachers were given a small honorarium to encourage high level of compliance (3) Equipment to prepare drinking water and disposable cups (4) Training + Training material provided to teachers of each supplement school	Learning Nutrition	6 months	Cluster-RCT	45 schools (Maths sample: 1815 students; Hb sample 1215 students)

Included study	Population	Intervention summary	Included outcomes	Follow- Up	Study design	Sample Size
Mahawithanage *et al.* (2007) **Micronutrient supplementation**	Sri Lanka (rural), Primary schools (Grade 1- 5; Age: 5- 13 yrs)	As part of the Vitamin A supplementation trial of Children were assigned to either 200,000 IU of Vitamin A (n = 297) or placebo (n=316) once every four months over a period of 13 months.	Attendance; Nutrition	13 months	RCT	659 students
Clarke *et al.* (2008) **Malaria control**	Kenya (rural), Primary schools (Grade: all; Age: 5- 18 years)	The Intermittent Preventative Treatment (IPT) Malaria Trial provided children in intervention schools one dose of sulfadoxine-pyrimethamine and three daily doses of amodiaquine on three occasions within a 12-month period (once each school term). IPT was first given in May, 2005, coinciding with a seasonal peak in transmission, and then repeated in September, 2005, and January, 2006.	Learning; Nutrition	12 months	Cluster- RCT	679 students

Included study	Population	Intervention summary	Included outcomes	Follow- Up	Study design	Sample Size
Simwaka *et al.* (2009) **Malaria control**	Malawi (not clear whether urban or rural), Primary school (Mean age: 10 yrs)	Teachers detected suspected malaria cases and provided age- appropriate doses of sulphadoxine- pyremethamine tablets and paracetamol to the student at the school free of charge. Students also received pupil's treatment kits (PTKs). The intervention involved a teacher training on how to utilise malaria treatment kits and on skills to persuade students to come to school if they are sick.	Dropout; Completion; Attendance	Approx. 60 months	CBA (DID, PSM)	651 students
Brooker *et al.* (2015) **Malaria control**	Kenya (rural), Primary School (Grade 1 and 5; Mean Age 10 years)	Health and Literacy Intervention (HALI) project- The health interventions involved intermittent screening and treatment (IST) of malaria in schools by public health workers using rapid diagnostic tests (RDTs) once a school term. There are 5 rounds of screening and treatment.	Learning; Nutrition	Learning: 9 months and 24 months Nutrition: 12 and 24 months	Cluster- RCT	101 schools, classes 1 and 5); 5177 students
Fernando *et al.* (2006) **Malaria control**	Sri Lanka (not clear whether urban or rural), Primary school (Grades 1-5; Age: 6- 12)	Malaria Control trial- at weekly school visits, one chloroquine tablet (150 mg of chloroquine phosphate base) or placebo (5 mg of nicotinamide) was given to each child after a meal under the direct supervision of a research assistant or the teacher. All children having malarial parasites on blood smear examination were treated with 300 mg of chloroquine phosphate base on days one and two, and 150 mg of chloroquine phosphate base on day three.	Attendance; Learning; Nutrition	9 months	RCT	587 students

Included study	Population	Intervention summary	Included outcomes	Follow- Up	Study design	Sample Size
		Children with a P. falciparum infection were administered 22.5 mg of primaquine on day one. Children with a p. vivax infection were administered 7.5 mg of primaquine for five days.				
Ebenezer *et al.* (2013) **De-worming / Micronutrient supplementation**	Sri Lanka (unclear if urban or rural), Primary school (Grade 4; Age: 9.5 yrs)	Deworming/ Iron Supplementation Trial- the trial provided students with deworming treatment and weekly iron supplementation for six months. Teachers were trained on how to administer the intervention.	Learning; Nutrition; Health	6 months	Cluster- RCT	1090 students
Grigorenko *et al.* (2007) **De-worming/ Micronutrient supplementation**	Zambia (urban and rural) Primary School (Grade 3- 7)	The School Health and Nutrition Study- Interventions included 1) Annual treatment for intestinal worms with albendazole 2) Annual treatment for schistosomiasis with praziquantel 3) Annual vitamin A Supplementation 4) Weekly iron supplementation 5) Health education delivered by teachers	Learning;	5- 6 months	Cluster-RCT	Sample size in Year 1: 1,963. Year 2: 2,567. Year 3: 2,159.

Included study	Population	Intervention summary	Included outcomes	Follow- Up	Study design	Sample Size
Watkins *et al.* (1996) **De-worming**	Guatemala (rural), Primary school (Grade: 1-4; Mean Age: 10 yrs)	Students in this deworming trial were randomly assigned to receive either albendazole or a placebo at 0 and 12 weeks. All children received albendazole at the end of the study.	Learning; Attendance;	6 months	RCT	256 students
Miguel *et al.* (2004) **De-worming**	Kenya (rural). Primary School (Grade 1- 8)	Primary School Deworming Project (PSDP)- (1) Schools with geohelminth prevalence over 50 per cent were mass treated with albendazole every six months, and schools with schistosomiasis prevalence over 30 per cent were mass treated with praziquantel annually. (2) Treatment schools received worm prevention education through regular public health lectures, wall charts, and the training of teachers in each treatment school on worm prevention. (p.169) (3) A series of community and parent meetings were held in treatment schools	Attendance; Learning; Health	12- 24 months (depending on treatment group	Cluster- RCT	24958 students
Simeon *et al.* (1995) **De-worming**	Jamaica (rural and urban) Primary School (Grade 2- 5)	As part of the trial, children with T. trichiura infections were randomly assigned to receive anthelmintic treatment (800mg of albendazole), given at baseline measurements, 12 weeks later, and 24 weeks later.	Attendance; Learning; Nutrition	6 months	RCT	407 students

Included study	Population	Intervention summary	Included outcomes	Follow- Up	Study design	Sample Size
Glewwe *et al.* (2014) **Vision correction**	China (rural) Primary School (Grade 4- 6, Age: 8- 15 yrs)	Gansu Vision Intervention Project (GVIP)- Children with previously identified low vision in intervention schools are given glasses to correct impaired eyesight.	Learning	10 months	RCT	28271 students
Sylvia *et al.* (2013) **Incentives for anaemia reduction**	China (rural) Primary School (Grade 4- 5; Mean Age: 10 years)	Incentives for Health Promotion trial- School principals in both treatment arms were provided with information and training about anaemia reduction. In addition to information, schools in the subsidy group treatment arm received a subsidy of 1.5 yuan (about US$0.22) per student per day earmarked for anemia-related expenses. Schools in the health incentive group received the same information and subsidy. In addition, the school principal was given an incentive (pay- for- performance) contract rewarding reductions in the number of anemic students between the baseline and endline surveys.	Learning Nutrition	7 months	Cluster- RCT	2957students

Included study	Population	Intervention summary	Included outcomes	Follow- Up	Study design	Sample Size
Jukes *et al.* (2014) **Deworming + Iron Supplementation**	Philippines (rural) Primary School (Grade 3)	Save The Children School Health and Nutrition (SHN) programme School teachers, together with staff from the Rural Health Unit (RHU) and the Department of Education administered: Deworming: Two 400 mg oral dose of albendazole during one school year Iron Treatment: 60 mg iron tablets distributed weekly, a regimen found to be as effective as daily supplementation in some countries	Learning Nutrition	3 months	Cluster- RCT	1460 students

4.1.3 Synthesis of findings

We present the results in two sections. First, we present the results of the meta-analysis on the effects of SBH programmes on primary and secondary education outcomes. Second, we present a summary of findings and discussion of the findings incorporating any relevant evidence from the descriptive qualitative synthesis.

Effects of SBH interventions on attendance and learning outcomes

We only combined studies evaluating similar treatment constructs and present the results organised by SBH type: Malaria prevention and control, Multivitamin/ Micronutrient interventions and Deworming. We have structured the presentation of results of each SBH type according to the 'ideal type' theory of change (Figure 4.1 a), starting with intermediate outcomes (nutrition) followed by school participation (attendance) and final outcomes (learning outcomes: cognitive test scores, language arts test scores and maths test scores).

Only fourteen of the sixteen included studies provided data for the analysis. Two studies did not contain the necessary data for us to calculate effect sizes (Simwaka *et al.*, 2009; Grigorenko *et al.*, 2007).[6] The interventions in the incentives for health promotion study (Sylvia *et al.* 2013) and vision correction study (Glewwe *et al.,* 2014) were considered to be too different to the other included SBH studies and therefore were not included in any of the meta- analyses. Mahawithanage *et al*.'s Vitamin A supplementation trial in Sri Lanka was also not included in the meta-analyses as the intervention design was considered too different to the other multivitamin/ micronutrient studies. We present standardised effect sizes for these studies separately.

The studies included a range of different follow up periods, with the majority of studies including a data point within the 0- 12 month period. We selected these for the meta-analysis when available. If studies only provided data for a longer follow up period we included these in the meta-analysis, using sensitivity analysis to assess whether this makes a difference to the results. As part of the Save the Children School Health and Nutrition (SHN) programme in the Philippines, the control group received deworming and iron supplementation after the first follow- up survey. We therefore only included the first follow up period in our analysis. In cases where numerous measures of language arts, mathematics or cognitive scores were presented (for example reading and spelling for language arts), we combined the measures to create one synthetic effect size. Studies reported on a range of different measures of nutrition. To allow us to combine studies in meta-analyses we decided to include both measures of anaemia and haemoglobin outcomes in the same meta-analyses. As the desired effect is positive for haemoglobin levels and negative for anaemia we reversed the sign for anaemia before including the effect size in the meta-analysis.

Several of the studies included multiple treatment arms. Miguel *et al.*'s (2004) study in Kenya had two treatment groups receiving the same treatment in 1998 and 1999 respectively, compared against a group receiving no treatment. For the meta- analysis, we used the 1999 treatment group where group 1 (schools receiving treatment in both 1998 and 1999) and group 2 (schools receiving treatment in 1999 only) are combined and compared against the control (group 3). Wong *et al*.'s study included an iron supplementation programme treatment arm and a treatment arm where parents were given training about nutrition and anaemia through power point, video and colourful booklets, and we only included the iron supplementation treatment arm in the analysis.

[6] Both authors were contacted but no response was received.

With regards to the labelling of the studies in the forest plot, numbers refer to different programmes implemented in the same country whereas letters refer to different samples from the same programme, for example different year groups. Kenya1 refers to the Primary School Deworming Project (Miguel *et al.,* 2004). Kenya2 is the Health and Literacy Intervention (HALI) project with Kenya2a referring to the grade 1 sample and Kenya2b to the grade 5 sample (Brooker *et al.,* 2015). Kenya3 is the Intermittent Preventative Treatment (IPT) Malaria Trial (Clarke *et al.*, 2008). China1 is the Gansu Vision Intervention Project (GVIP) that provided glasses to students with low vision (Glewwe *et al.*, 2014). China2 refers to the chewable vitamin intervention in the Gansu province (Kleiman- Weiner *et al.*, 2013) and China3 refers to Luo et al.'s multivitamin study. China4 is the Iron Supplementation trial (Wong *et al.,* 2014).With regards to the included studies in Sri Lanka, Sri Lanka1 is the Malaria Control trial (Fernando *et al.,* 2006) with Sri Lanka1a being the Female subgroup sample and Sri Lanka1b being the male subgroup sample. Sri Lanka2 is the vitamin A supplementation trial (Mahawithanage *et al.,* 2007) and Sri Lanka3 refers to the deworming and iron supplementation trial (Ebenezer *et al.*, 2013).

All effect sizes are expressed as standardised mean difference (SMD), interpreted as the magnitude of the number of standard deviation changes in the outcome for the intervention group as compared to students in non-SBM schools. SMD scores are interpreted as the number of standard deviation changes in the outcome.

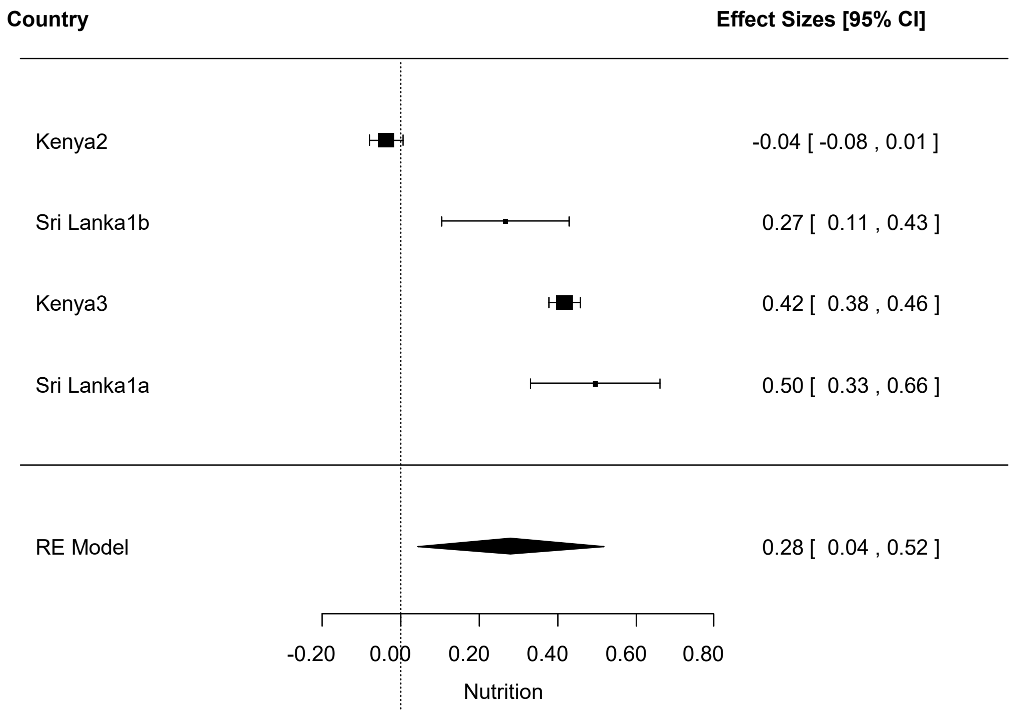
Malaria Prevention and Control

Nutrition

We conducted a meta-analysis of nutrition outcomes for the three Malaria studies that presented data on this outcome (Fernando *et al.,* 2006; Clarke *et al.,* 2008; Brooker *et al.,* 2015). The overall weighted average effect of SBH on nutrition is 0.28 95% CI [0.04, 0.52]), calculated under a random effects model. Figure 4.1b suggests the presence of heterogeneity and this is confirmed by the homogeneity tests (I^2 = 97.98%, τ^2 = 0.0549, Q (df = 3) = 244.25, p-val= < .0001). The effect sizes range from -0.04, 95% CI [-0.08, 0.01] in Kenya2 (Brooker *et al.,* 2015) to 0.50, 95% CI [0.33, 0.66] in Sri Lanka1 (Fernando *et al.,* 2006). Two of the effect sizes in this meta-analysis come from the same SBH programme in Sri Lanka (Sri Lanka1a and Sri Lanka 1b) but correspond to different gender groups and are thus independent samples (Ebenezer *et al.*, 2013). The results are sensitive to the removal of any one of the included estimates. In particular, removing Brooker *et al.* (2015) from the analysis increases the magnitude of the overall effect substantially (SMD=0.40, 95% CI [0.30, 0.50]). Removing any of the other estimates changes the magnitude of the overall effect and makes the estimates less precise.

Figure 4.1 b: Nutrition

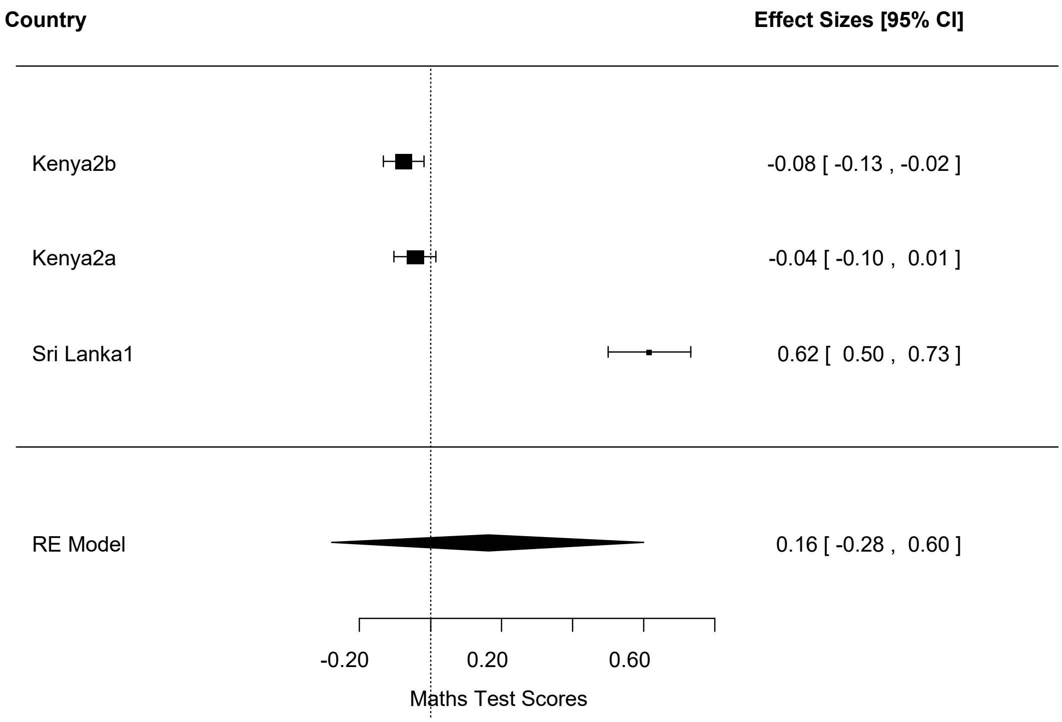
Maths test scores



Two Malaria intervention studies provided effect sizes on maths test scores. The overall average effect is 0.16 (95% CI [-0.08, 0.25]), calculated under a random effects model. Figure 4.1c suggests the presence of heterogeneity and this is supported by the homogeneity tests (I^2 = 99.13%, τ^2 = 0.1490, Q (df = 2) = 115.11, p-val < .0001). The effect sizes range from -0.08, 95% CI [-0.13, -0.02] in Kenya (Brooker *et al.*, 2015) to 0.62, 95% CI [0.50, 0.73] in Sri Lanka1 (Fernando *et al.*, 2006).

Both grade samples from the Health and Literacy Intervention study in Kenya (Kenya2a to Kenya2b, grades 1 and 5 respectively), show small negative effects, whereas the Malaria control trial in Sri Lanka shows a positive effect of substantial magnitude, and their confidence do not overlap. The results are sensitive to the removal of any one single estimate. The overall effect is negative when Fernando *et al.* (2006) is removed from the analysis, whereas removing any one of the estimates from Brooker *et al.* (2015) increases the magnitude of the positive effect substantially, although the estimates remain imprecise.

Figure 4.1 c: Maths Test Scores

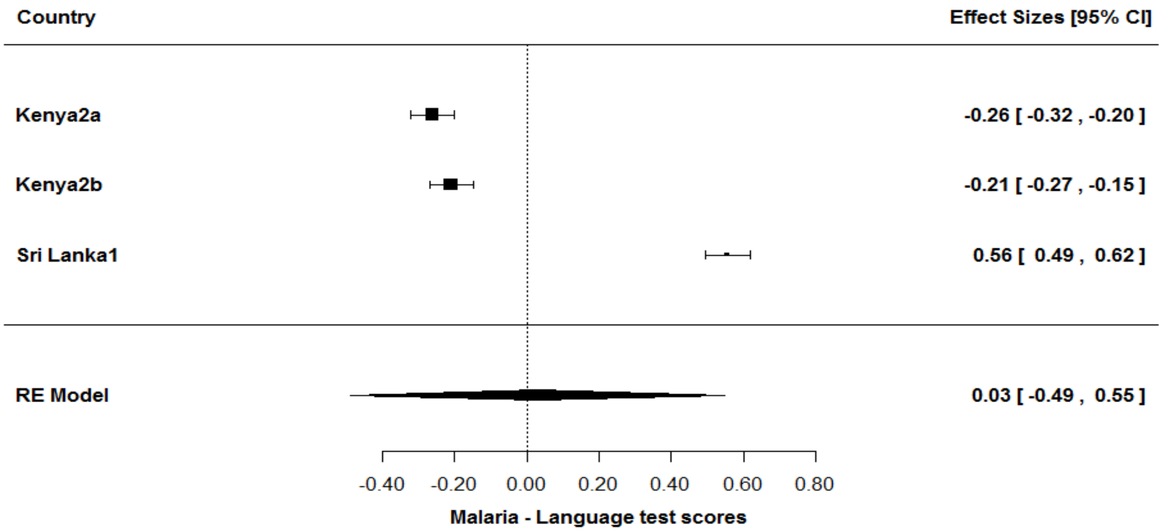


Language arts test scores

The overall average effect of the Malaria interventions on the language arts test scores is 0.03 (95% CI [-0.49, 0.55]), calculated under a random effects model. The homogeneity tests suggests high between-studies variability (I^2 = 99.55%, τ^2 = 0.2086, Q (df = 2) = 423.7535, p-val=< .0001). Figure 4.1d supports the presence of between study heterogeneity.

The effect sizes range from -0.26, 95% CI [-0.32, -0.20] for the grade 1 sample in the HALI project in Kenya (Brooker *et al.,* 2015) up to 0.56, 95% CI [0.49, 0.62] in the Malaria Control trial in Sri Lanka (Fernando *et al.,* 2006). The results are sensitive to the removal of any one single estimate. The overall effect becomes negative when we remove Fernando *et al.* (2006) from the analysis, whereas removing any one of the estimates from Brooker *et al*. (2015) increases the magnitude of the positive effect substantially, although the estimates remain imprecise.

Figure 4.1 d: Language Arts Test Score

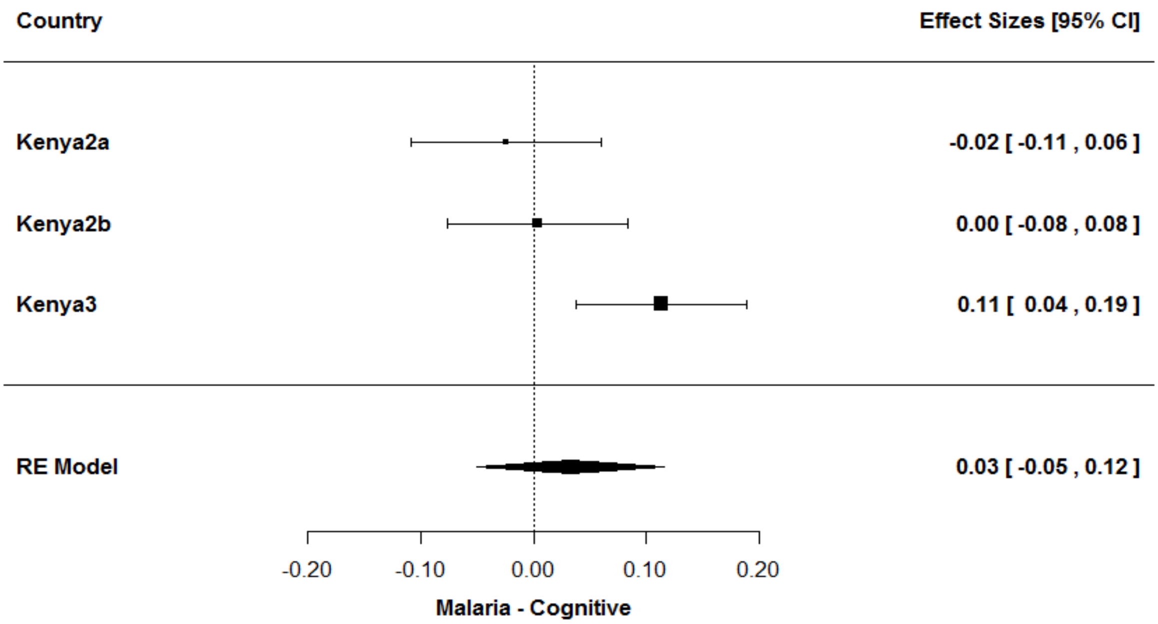


Cognitive Test Scores

The overall average effect of Malaria interventions on cognitive scores is 0.03 (95% CI [-0.05, 0.12]), calculated under a random effects model. The homogeneity tests suggests moderate between-studies variability (I^2 = 69.17%, τ^2 = 0.0037, Q (df = 2) =6.5805, p= 0.04) and Figure 4.1e supports the presence of moderate between study heterogeneity.

The effect sizes range from -0.02, 95% CI [-0.11, 0.06] for the grade 1 sample in the Health and Literacy Intervention (HALI) project in Kenya (Brooker *et al.*, 2015 - Kenya2a, grade 1 sample) up to 0.11, 95% CI [0.04, 0.07] for Intermittent Preventative Treatment (IPT) Malaria Trial in Kenya (Clarke *et al.,* 2008). The results are sensitive to the inclusion of the effect size for the IPT trial. Removing this study results in a negative overall effect estimate of -0.01 (95% CI [-0.07, 0.05]) (See Appendix H for results of all sensitivity analyses).

Figure 4.1 e: Cognitive Test Scores



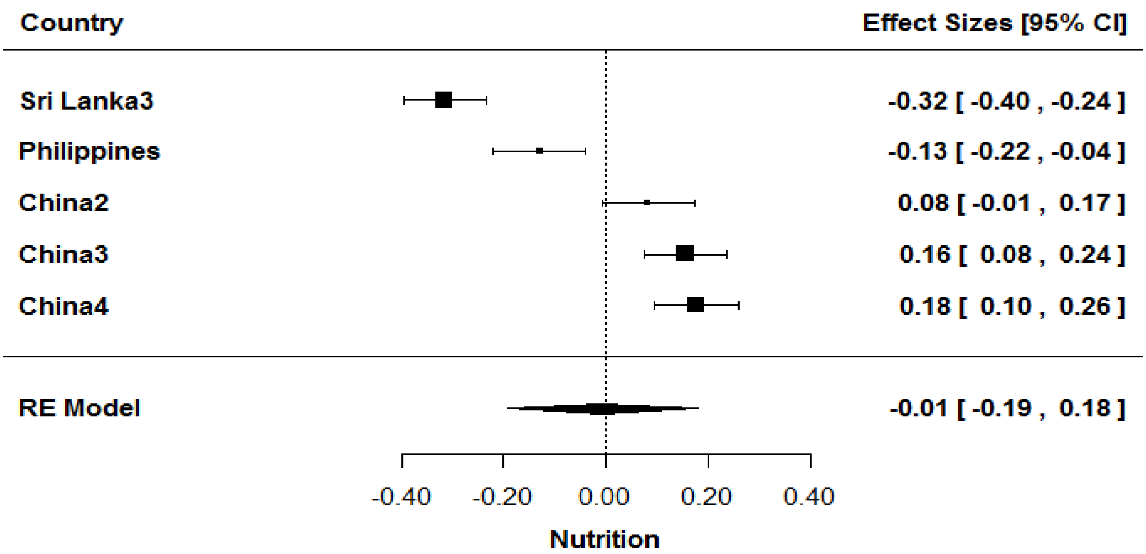
Micronutrient interventions

Nutrition outcomes

We conducted a meta-analysis of nutrition outcomes for the five studies that presented data on this outcome (Ebenezer *et al*. 2013, Luo *et al*. 2012; Kleiman- Weiner *et al*. 2013; Jukes *et al*. 2014, Wong *et al*. 2014) Figure 4.1f presents the forest plot with the results of the individual studies and the overall estimate. The overall average effect of micronutrient interventions on nutrition is -0.01 (95% CI [-0.19, 0.18], calculated under a random-effects model.

The assessment of homogeneity suggests the effects do not arise from a common population (I^2 = 95.86% τ^2 = 0.0130, Q (df = 4) = 102.3580, p = < 0.0001). The effects range from -0.32 (95% CI, [-0.40, -0.24] in Sri Lanka (Ebenezer *et al*. 2013) to 0.18 (95% CI [0.10, 0.26] in China (Wong *et al*. 2014). The confidence intervals of the study from Sri Lanka 3 (Ebenezer *et al*. 2013) and the Philippines (Jukes *et al*. 2014) do not overlap with the three studies from China. As expected the results are sensitive to the removal of any single estimate. Removing the effect sizes for Sri Lanka 3 (Ebenezer *et al*. 2013) and the Philippines (Jukes *et al*. 2014) change the direction of the overall estimate, and when removing any of the studies from China the magnitude of the negative effect is increased. The overall estimate remains imprecise however, with the confidence intervals crossing the line of no effect for all estimates (See Appendix H for results of all sensitivity analyses).

Figure 4.1 f: Nutrition

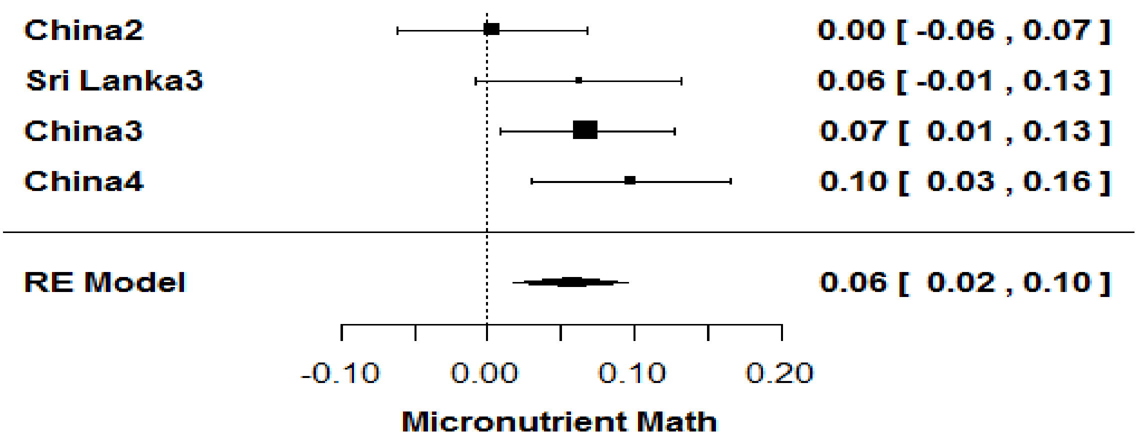


Maths Test Score

The overall average effect of micronutrient interventions on maths test scores is 0.06, 95% CI [0.02, 0.10]. Figure 4.1g suggests some heterogeneity and this is supported by the homogeneity test (I^2 = 29.07%, τ^2 = 0.0470, Q (df = 3) = 4.2227, p- val = 0.2384).

The effect sizes range from 0.00, 95% CI [-0.06, 0.07] in China2 (Kleiman- Weimar *et al.,* 2013) to 0.10, 95% CI [0.03, 0.16] in China4 (Wong *et al.,* 2014). The overall results are relatively robust to the removal of any one study, although removing China 3 (Luo *et al.,* 2012) reduces the precision of the overall estimate, with confidence intervals crossing the line of no effect.

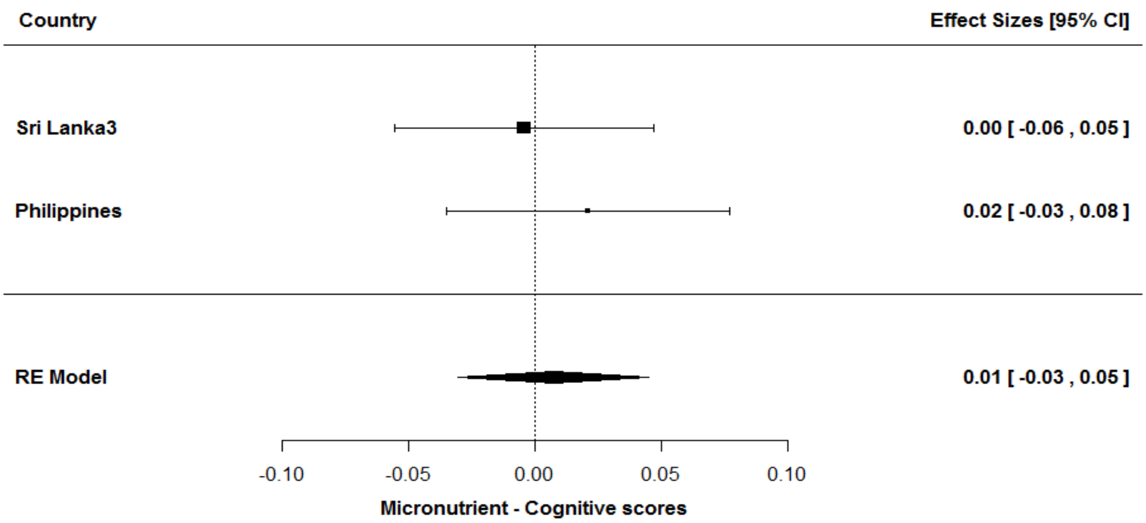
Figure 4.1 g: Maths



Cognitive Test Scores

We conducted a meta-analysis on cognitive outcomes for the two micronutrient studies that presented data on this outcome. The overall average effect of micronutrient interventions on cognitive scores is 0.01 (95% CI [-0.03, 0.05]), calculated under a fixed effects model. The homogeneity test suggests the only source of variation is within-study sampling error (I^2= 0%, τ^2 = 0, Q (df = 1) = 0.4248, p-val = 0.5146).

Figure 4.1 h: Cognitive Outcomes

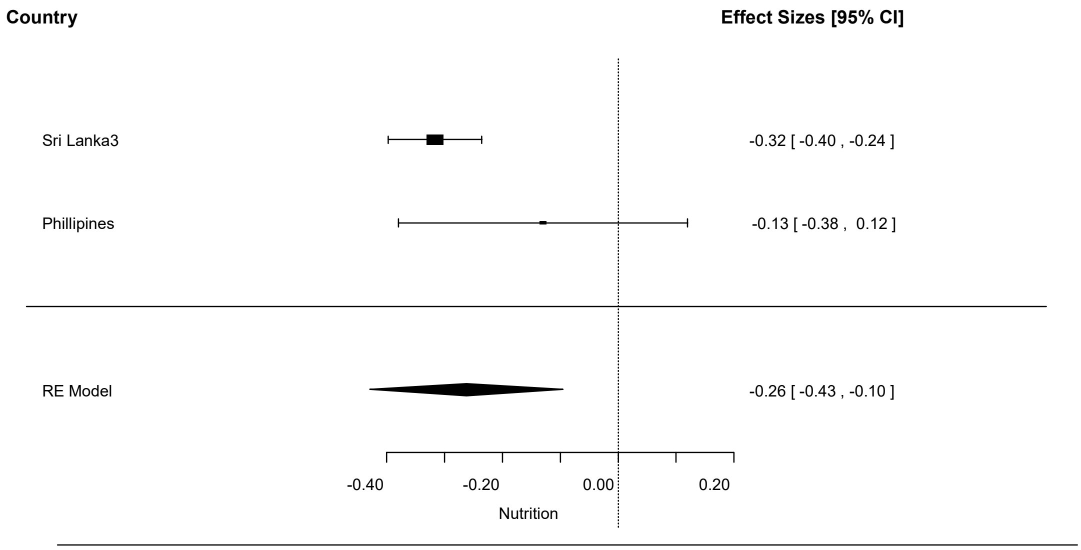


Deworming Interventions

Nutrition

We conducted a meta-analysis of nutrition outcomes for the two studies that presented data on this outcome (Ebenezer *et al*., 2013, Jukes *et al.,* 2014). Figure 4.1i presents the forest plot with the results of the individual studies and the overall estimate. The overall average effect of deworming interventions on nutrition is -0.26, 95% CI [-0.43, -0.10], calculated under a random-effects model. The assessment of homogeneity suggests a moderate amount of between-study variability (I² = 48.38% τ^2 = 0.0084, Q (df = 1) = 1.9373, p= 0.1640). The effects range from -0.32 (95% CI, [-0.40, -0.24] in Sri Lanka (Ebenezer *et al.*, 2013,) to -0.13 (95% CI [-0.38, 0.12] in the Phillipines (Jukes *et al.,* 2014).

Figure 4.1 i: Nutrition

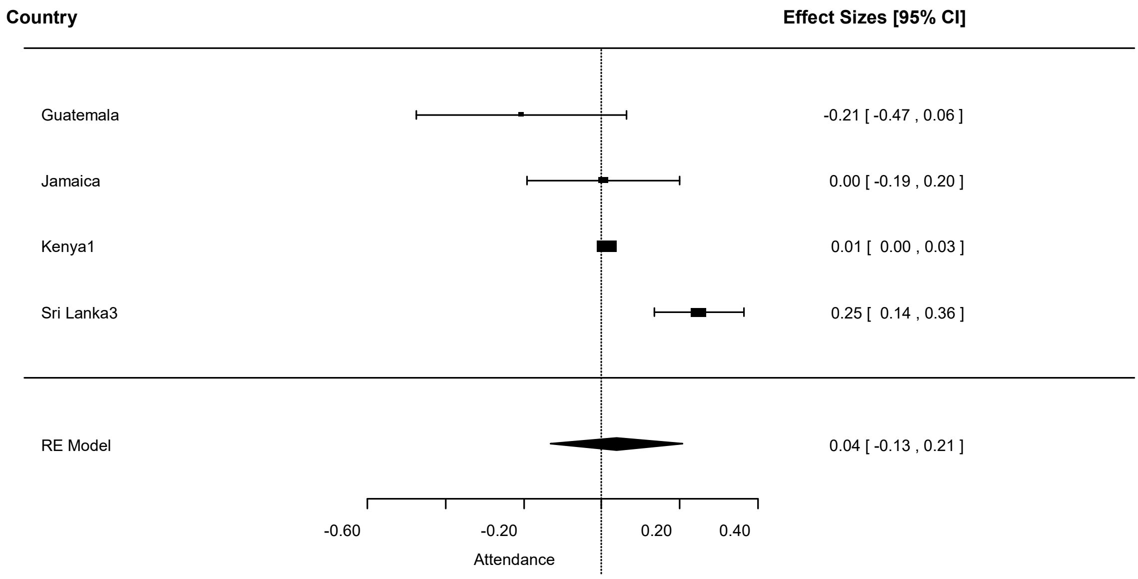


Attendance

The overall average effect of deworming interventions on student attendance is 0.04 (95% CI [-0.13, 0.21]), calculated under a random-effects model. The homogeneity tests (I²= 87.27%, τ^2 = 0.0231, Q (df = 3) = 18.9626, p = 0.0003) indicate a high amount of between study variability. Figure 4.1j supports the presence of heterogeneity.

The effect sizes range from -0.21 (95% CI [-0.47, 0.06]) in Guatemala (Watkins *et al.,* 1996) to 0.25 (95% CI [0.14, 0.36]) in one of the studies from Sri Lanka (Ebenezer *et al*., 2013 – Sri Lanka3). The results are sensitive to the removal of any one of these studies. Removing Watkins *et al.* (1996) increases the magnitude of the overall estimate to 0.9, but it remains imprecise (95% CI [-0.07, 0.25]), whilst removing Ebenezer *et al*. (2013) reduces the magnitude of the overall effect and the estimate is more precise (SMD=0.1, 95% CI [0.00, 0.03])

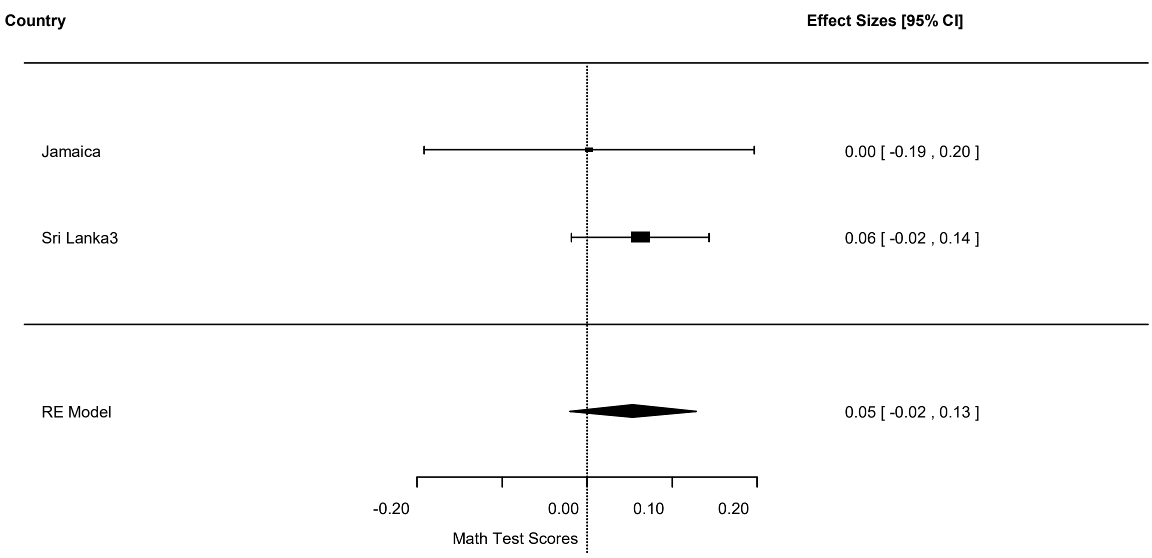
Figure 4.1 j: Attendance



Maths Test Scores

The overall average effect of deworming interventions on maths test scores is 0.05 (95% CI [-0.02, 0.13]). The assessment of homogeneity suggest the only source of variation is within-study sampling error (I^2 = 0%, τ^2 = 0, Q (df = 1) = 0.3157, p=< 0.5742).The effect sizes range from zero, 95% CI [-0.19, 0.2] in Jamaica (Simeon *et al.*, 1995) to 0.06, 95% CI [-0.02, 0.14] in Sri Lanka3 (Ebenezer *et al.*, 2013).

Figure 4.1 k: Maths Test scores

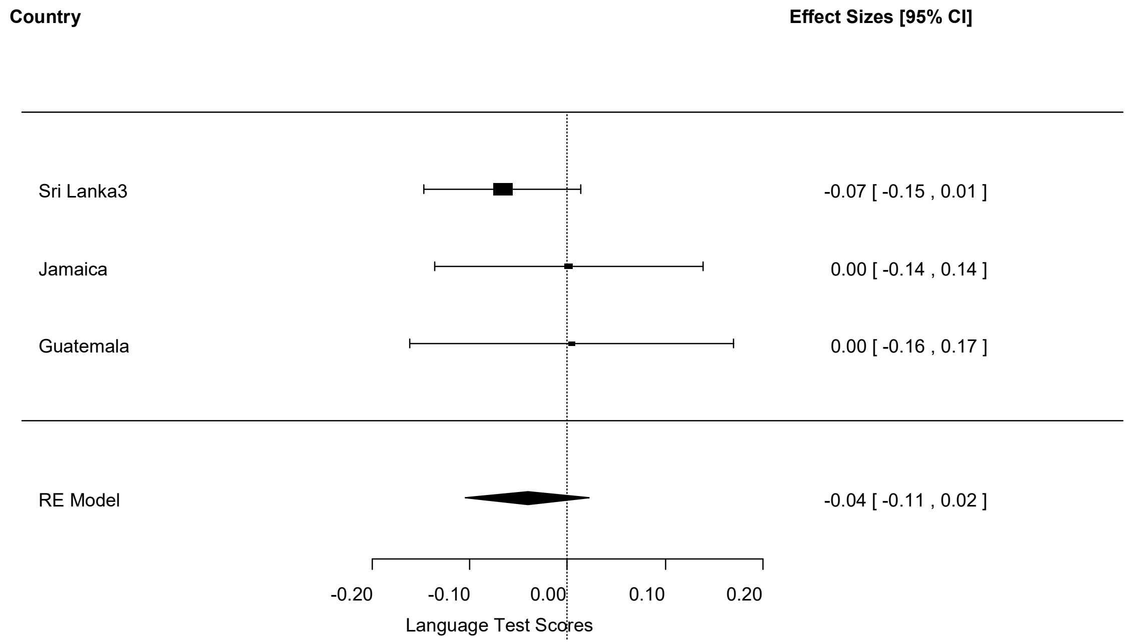


Language Arts

The overall average effect of deworming interventions on learning as measured by language arts is negative (SMD= -0.04, 95% CI [-0.11, 0.02], calculated under a fixed-effects model. The homogeneity test (I^2 = 0%, τ^2 = 0, Q (df = 2) = 1.0281, p= 0.5981) indicates that the effects did arise from the same population. As can be seen from Figure 4.1l the confidence intervals of the included studies are overlapping.

The overall average effect is sensitive to the inclusion of one study (Ebenezer *et al.,* 2013). Removing this study changes the point estimate from a negative effect to zero, although the confidence intervals are still crossing the line of no effect (95% CI [-0.10, 0.11]).

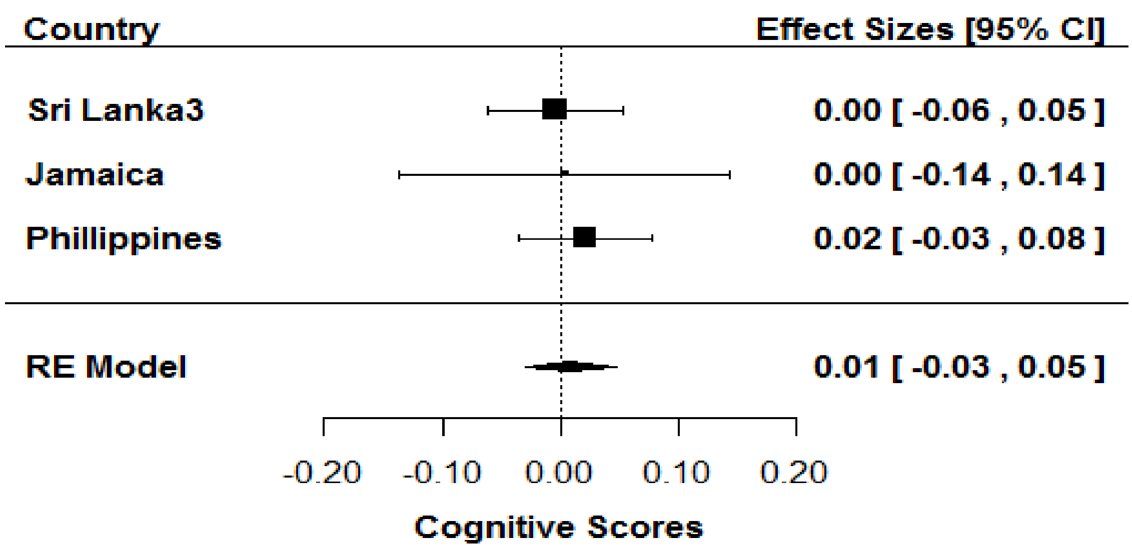
Figure 4.1 l: Language Arts



Cognitive Test Scores

We conducted a meta-analysis of cognitive outcomes for the three deworming studies that presented data on this outcome. The overall average effect of deworming interventions on cognitive scores is 0.01 (95% CI [-0.03, 0.05]). The homogeneity test suggests low between-studies variability (I^2 = 0%, τ^2 = 0, Q (df = 2) = 0.3891, p-val = 0.8232).

Figure 4.1 m: Cognitive Test scores



Other programmes

Both the vision correction study (Glewwe *et al.,* 2014) and the incentives for health promotion study (Sylvia *et al.,* 2013) were considered too different to be included in any of the meta- analyses. Glewwe *et al.'*s study (2014) measured composite, maths and language arts test scores and found zero or small negative effects. (maths: SMD=-0.03, 95% CI [-0.06, -0.01], language: SMD=0.00, 95% CI [-0.03, 0.03], composite; SMD=-0.03, 95% CI [-0.04, -0.10]). The incentives for health promotion study (Sylvia *et al.,* 2013) reported on maths test scores and nutrition outcomes and found zero or small negative effects of both types of anaemia reduction programme (subsidy or health incentive intervention). Appendix H reports all individual effect sizes for the outcomes reported in these two studies.

4.1.4. Summary of findings and discussion

We included 16 studies across seven countries in Latin America and the Caribbean, East Asia, South Asia and Sub- Saharan Africa that evaluated the effect of a school based health programme or trial in L&MICs. The included studies evaluate a range of different school-based health interventions including de-worming (n= 6), malaria prevention and control (n= 4), micronutrient supplementation (n= 7), the provision of glasses (n= 1) and the provision of incentives for anaemia reduction (n= 1). Three studies provided a combination of de-worming and micronutrient supplementation. We were able to examine effects on nutrition outcomes, student attendance, cognitive scores, maths and language arts test scores using meta-analysis.

The results suggest that Malaria Prevention and Control programmes have had beneficial effects on education outcomes for participating children in some contexts. The overall average effects range from 0.03, 95% CI [-0.05, 0.12] for cognitive test scores to 0.28, 95% CI [0.04, 0.52]) for nutrition outcomes. The effects on nutrition are relatively consistent, but there is a large amount of variability for all other overall estimates. For learning outcomes we observed positive effects of a large magnitude for the Malaria control programme in Sri Lanka (Fernando *et al.*, 2006, language arts: SMD=0.56, 95% CI [0.39, 0.72]; maths: SMD=0.62, 95% CI [0.45, 0.78]). Brooker *et al.* (2015) is the only other study assessing learning outcomes and in this case there appears to have been zero effects and, in the case of language arts and maths in particular, negative effects.

The authors provide several explanations for the lack of overall impact of IST, including high rates of re-infection, acquisition of new infections between screening rounds and variability in RDT performance between screening rounds. The authors also reported some evidence of a decline in full supervision (a proxy for compliance) due to logistical difficulties in providing the complex treatment regime. In addition, teachers considered delivering the malaria control programme disruptive and beyond the scope of their work (Brooker *et al.*, 2015 – see appendix J for the full synthesis of qualitative findings).

The meta-analyses of included micronutrient interventions suggest no overall effect on nutrition (SMD= -0.01, 95% CI [-0.19, 0.18]). The results suggest a beneficial effect on maths on average (SMD=0.06, 95% CI [0.02, 0.10]), while the analysis of the two studies that measured cognitive scores suggest no difference (SMD=0.01, 95% CI [-0.03, 0.05]).

The meta-analyses of studies assessing de-worming programmes suggest small, if any observable benefits for children receiving such programmes. There may be small positive effects on attendance in some contexts, but the average effects are small and imprecise for all outcomes, apart from nutrition, where we observed a negative effect (SMD=-0.26 SMD, 95% CI [-0.43, -0.10]).

Overall the results indicate the effects of SBH interventions have mostly been beneficial. However the average effects are relatively small in magnitude and in most cases the confidence intervals cross the line of no effects. For some outcomes, we observed negative or no effects. The exception to this is the relatively larger magnitude of effects on nutrition for malaria programmes. It is not clear why these improvements in nutrition outcomes do not appear to have translated into observable effects on learning outcomes in all cases. The evidence on SBH interventions is limited and combined with the presence of heterogeneity the average effects should be interpreted with caution.

Table 4.1 c: Descriptive findings: Process and implementation

Descriptive findings: Process and implementation	**Context**	**Citation (Info type)**
Teachers in Kenya considered delivering malaria control programme disruptive and beyond the scope of their work	Kenya	Brooker *et al.*, 2015: *Impact Evaluation*
The use of health workers to implement the SBH intervention may be critical in terms of training and supervising teachers and handling referral cases	Kenya	Brooker *et al.*, 2015 *Impact Evaluation*
Treatment may not have been delivered as intended, with divergence from treatment regime and target population	Kenya, Sri Lanka	Miguel *et al.*, 2004 Clarke *et al.*, 2008 Mahawithanage *et al.*, 2007; Fernando *et al.*, 2007: *Impact Evaluation*
High compliance rates for delivery of treatment	Kenya, China	Kleiman- Weiner *et al.*, 2013; Luo *et al.,* 2012: Wong *et al.*, 2014 *Impact Evaluation*
There was an apparent decline in full supervision due to logistical difficulties in providing a complex treatment regime that involved more than one visit	Kenya	Brooker *et al.,* 2015: *Impact Evaluation*
In China, highly decentralised implementation of the programme and poor project management might have resulted in the distribution of eyeglasses in ways inconsistent with project criteria	China	Glewwe *et al.*, 2014: *Impact Evaluation*
In Zambia, funding limitations resulted in the scale down of the intervention in the third year.	Zambia	Grigorenko *et al.*, 2007: *Impact Evaluation*
Community sensitisation activities played a key role in gaining parents support.	Zambia	CAI, 2007: *Project Document*
Parents adequate knowledge of the disease may be important for compliance with the treatment	Kenya, China	Brooker *et al.*, 2015, Glewwe *et al.*, 2014: *Impact Evaluations*

Table 4.1 d: Descriptive findings: Contextual Factors

Descriptive findings: Contextual factors	Context	Citation (info type)
Infection intensity at baseline did not moderate the treatment effect	Kenya, Guatemala	Watkins *et al.,* 1996 Brooker *et al.,* 2015: *Impact Evaluations*
Baseline nutritional status, prevalence and intensity of infection may be important in mediating the effect of SBH programmes	Jamaica, China	Simeons *et al*., 1995, Sylvia *et al*., 2014, Luo *et al*., 2012, Wong *et al*., 2014 *Impact Evaluations*
External events, including natural disasters and disease outbreaks, during the intervention period may have been a barrier to the success of SBH programmes.	China, Kenya, Sri Lanka	Glewwe *et al*., 2014; Miguel *et al*., 2004; Fernando *et al*., 2006; Jukes *et al*., 2014 *Impact Evaluations*

4.2 School feeding programmes

School feeding programmes typically aim to improve the general health of children, provide a safety net for vulnerable and food insecure families, and improve children's ability to learn (Jomaa *et al*., 2011). School feeding programmes use many different modalities to provide food to schoolchildren but can be classified into two main groups of interventions:
In-school feeding programmes, where children are served food in the school. This food can be provided as a breakfast, snack(s), and/or lunch and is often prepared within the school facilities.
Take-home rations, food supplies usually provided on a monthly basis, provided to households conditional on a child's enrolment and a minimum level of attendance (usually 80- 85 per cent of school days) (Adelman *et al*., 2008). With a specific focus on improving the food security of households, families are allowed to redirect the food rations to members of the household or sell it for other goods or cash.
The food provided through these programmes is sometimes fortified with essential minerals or vitamins in order to address multiple nutritional deficiencies. A strong case has been made by organisations such as the World Food Programme that micronutrient fortification should always be an integral part of school feeding if there is a demonstrated need (Bundy *et al.,* 2009). In addition, the complex set of needs of students has led researchers and practitioner to argue for the need for feeding programmes to adapt an integrated approach which combines health, nutritional and educational components in order to increase their effectiveness (Powell: 1998; Jomaa *et al*., 2011).

4.2.1 How may school feeding programmes affect education outcomes?

As outlined in Figure 4.2a, school feeding programmes may affect participants' education outcomes through three main pathways: increasing school enrolment and attendance, alleviating short- term hunger and improving nutrition and health. The food provided under a school feeding programme, either at school or through take- home rations, represents an income transfer to the family by subsidising the costs of sending a child to school (Adelman *et al*., 2007; Lawson *et al.,* 2012). As the food rations are conditional on the attendance of the student on that day or a minimum threshold of attendance, these benefits are thought to act as an incentive for parents to send their children to school more regularly. As a result, school enrolment and attendance will be boosted while dropout rates of participating students reduced (Bergeron *et al.*, 2001; Kristjansson *et al*., 2009; Petrosino *et al.*, 2012; Ahmed, 2004).

Short- term hunger can adversely affect attention span and the interest of students to participate in class and may affect performance on cognitive tasks (Kristiansson *et al.*, 2012). School meals or snacks, especially provided early in the day to alleviate hunger before or during classes, have shown to improve children's attention span and cognitive functioning as well as motivation (Chandler *et al.,* 1995; Mathews *et al.*, 1995). In addition, a child may prefer to attend school if he or she is not hungry.

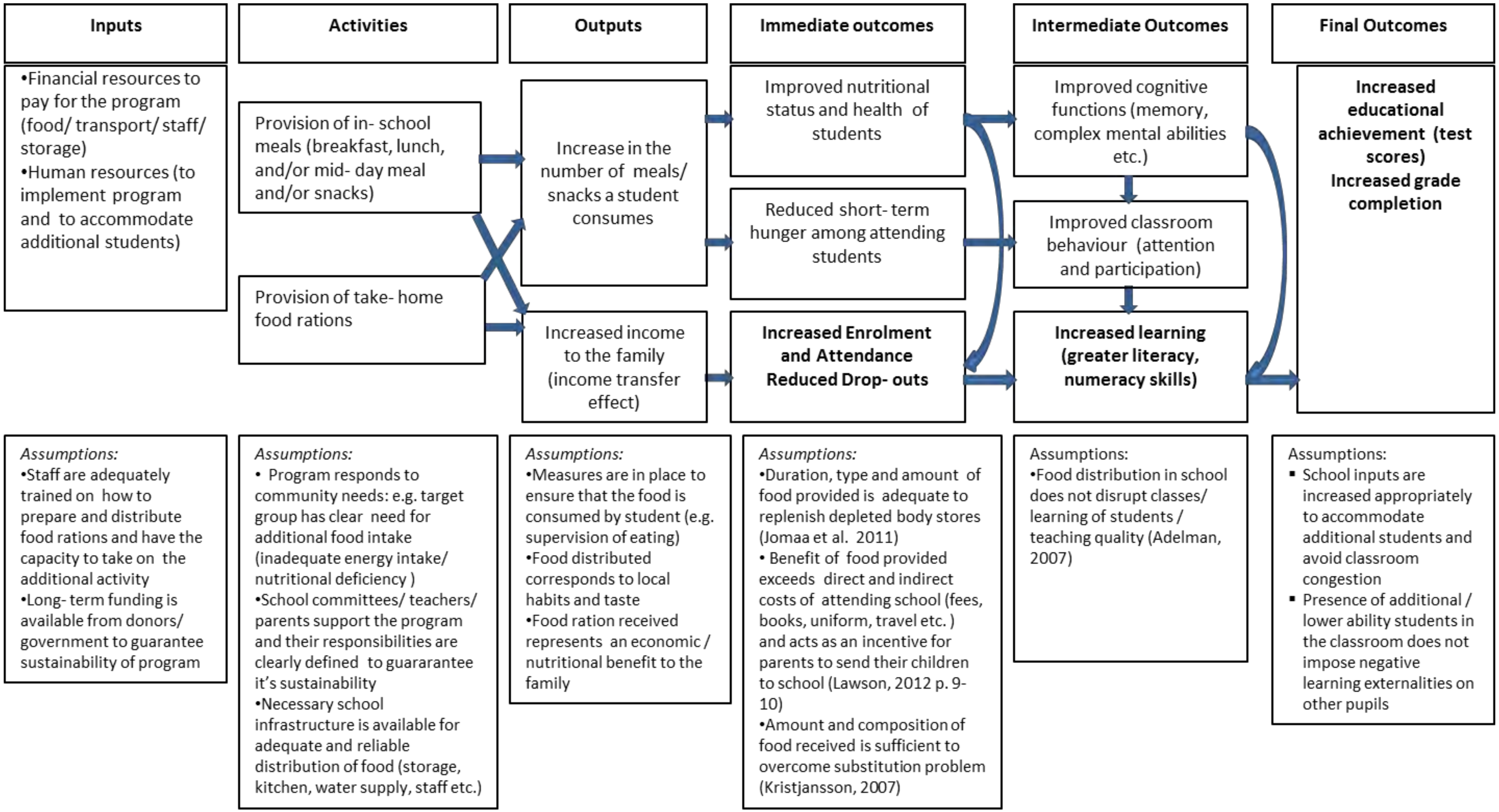
School feeding programmes can enhance diets through increasing total energy intake, addressing nutritional deficiencies (through appropriate selection of food and school menus) and improving micronutrient status. Children who are better nourished will have a better immune system which in turn reduces their susceptibility to infectious diseases and other illnesses (Adelman *et al.*, 2008). Reduced morbidity may decrease the number of school days missed due to illness, thus increasing attendance while potentially reducing dropouts (Jomaa *et al.,* 2011). In addition, school meals may prevent or reduce nutritional deficiencies that affect cognition, such as iron deficiency (Adelman *et al*., 2008). In this way school feeding is thought to affect learning and school performance as children spend more hours in school, allowing them to learn more and improve their school performance. School performance and retention may also be enhanced through improved cognitive functions and classroom behaviour as a result of reduced hunger and through improved nutritional status and health (Adelman *et al.,* 2008).

For school feeding interventions to improve education outcomes a number of assumptions must hold. For a feeding programme to increase school enrolment rates, the net benefit of participating in the programme must exceed the direct and indirect costs of attending school (opportunity cost of a child's time; fees, books, uniform, travel etc.) and act as an incentive for parents to send their children to school (Lawson, 2012; Adelman *et al.,* 2008). Moreover, for the food ration to correct nutritional deficiencies it is assumed that they are rich in energy, protein and vitamins. It also needs to address a nutritional deficiency that is suffered by the target group and provide adequate types of food (Greenhalgh *et al.*, 2009). At the same time, the type and quality of food need to be culturally acceptable to the students and households (Bundy *et al*., 2009).

A repeated concern with school feeding programmes is substitution which occurs when children receive a reduced home diet as a result of receiving food at school as parents, especially if poor households try to spread limited resources (Kristjansson *et al.*, 2007; Jomaa *et al.*, 2011; Greenhalgh *et al*., 2009). Therefore, the model assumes that the amount as well as composition of food provided offsets any diminished intake at home. Important assumption of the model are therefore that measures are in place that the school meal/ snack is consumed (for example, close supervision of eating) and that necessary school infrastructure is in place to ensure that the programme is implemented as intended. For example, the handout of food should not disrupt classes (Adelman, 2007; Greenhalgh *et al.,* 2009).

Finally, the benefits of providing a school meal may vary according to classroom conditions such as teacher/ student ratios, the availability of schooling inputs and teacher quality (Chang, 1996; Powell, 1998; Adelman *et al.,* 2008). It is assumed that school and classroom conditions are conducive for learning and school inputs are provided to improve school quality and to avoid classroom congestion due to additional students attending classes (Bergeron *et al.,* 2001; Adelman *et al.*, 2008).

Figure 4.2 a: School feeding programme theory



4.2.2 Description of included studies

We included 16 studies reported in 21 different papers that evaluated the effect of a school feeding intervention. These studies evaluate 15 unique programmes, as two studies evaluate the impact of the Mid-Day Meal Scheme in India. Table 4.2a provides an overview of the table of characteristics of the included studies and in the following section, we describe the characteristics of these studies in detail.

Population

All of the included studies evaluated programmes targeted at primary school children (n=16). In addition, the two programmes in Sri Lanka (both reported in He, 2010) also provided meals to children in early secondary school. The studies in Cambodia, Senegal, India (n=2), Chile, Argentina, Guyana, Sri Lanka (n=2) and Peru explicitly stated that they were targeted at public schools (Cheung and Berlin, 2014; Diagne *et al.*, 2014; Jayaraman and Simroth, 2015; Afridi, 2014; McEwan, 2013; Adrogue *et al.*, 2011; Ismail *et al.*, 2012; Jacoby *et al.*, 1996; He, 2010a. He, 2010b). In the other studies it was not clear in which type of schools the studies took place, but they are presumed to be public schools. Two programmes targeted the treatment to girls only; these were the take-home rations treatment arm of the WFP school feeding programme in Burkina Faso (Kazianga *et al.*, 2012) and the take-home rations component of the WFP Food for Education Programme in Cambodia (Cheung and Berlin, 2014).

Settings

The included studies covered programmes in a broad range of settings. Latin America and the Caribbean was the most well-represented continent with five studies, one each respectively from Chile (McEwan, 2013), Peru (Jacoby *et al.*, 1996), Argentina (Adrogue *et al.*, 2011), Guyana (Ismail *et al.,* 2012) and Jamaica (Powell *et al.,* 1998). Three studies took place in Sub- Saharan Africa, in Burkina Faso (Kazianga et al., 2013), Kenya (Omwami *et al*., 2011) and Senegal (Diagne *et al.,* 2014). There were four studies that took place in East Asia and Pacific; one each respectively from Laos (Buttenheim *et al.* 2013), the Philippines (Tan *et al.,* 1999), Cambodia (Cheung and Berlin, 2014) and China (Kleiman-Weiner *et al.,* 2013). Finally, there were four studies from South Asia, two of which evaluated the Midday Meal Scheme in India (Jayamaran and Simroth, 2015; Afridi *et al.,* 2014) and two that took place in Sri Lanka, both reported in He (2010).

Many of these studies evaluated programmes that took place primarily in rural areas or evaluated rural districts of a programme only (n=9). One of the studies in India (Jayaraman and Simroth, 2015) and the study in Argentina (*ibid*) evaluated nationwide programmes that presumably covered rural, urban and peri-urban areas. The two programmes in Sri Lanka (He, 2010) both covered large parts of the country and presumably also cover rural, urban and peri-urban locations. The study in Peru (*ibid*) was a small trial that took place in schools on the periphery of the city of Huarez only and Afridi's (2014) study in India evaluated the introduction of the Midday Meal Scheme in Delhi only. It is not clear in Cheung and Berlin's (2014) evaluation of the WFP school feeding programme in Cambodia where it primarily took place.

Intervention

All 16 included studies evaluated feeding programmes or trials that provided an in-school feeding component. In addition, the two studies that evaluated the World Feeding Programme school feeding programmes in Burkina Faso and Lao PDR also included an additional treatment arm that provided take-home rations. The WFP school feeding programme in Cambodia introduced take home rations for girls alongside the in-school feeding programme, and later also provided de-worming and complementary health activities. Table 4.2b provides an overview of key intervention components, as summarised below.

Delivery: Thirteen of the included programmes provided a meal to children at school. Three of the included programmes provided a snack only, while the Chilean school feeding programme (McEwan, 2013) and the Guyanese school feeding programmes (Ismail *et al.,* 2012) were free to provide either meals at breakfast and lunch, or a meal at lunch and a snack, depending on the school. Three of the programmes provided an additional intervention component alongside the feeding programme; in the Philippines this was an intervention to encourage improved teacher-parent partnerships through regular group meetings (Tan, 1999) and in Cambodia and Sri Lanka this was complementary health and sanitation activities for students (Cheung and Berlin, 2014; He, 2010a). The theory of change presents the timing of the delivery of food as another important design consideration, as this may impact a child's concentration and thus their ability to learn. Six of the programmes provided food at breakfast, four at mid-morning, and eight at lunch time. All of the studies stated explicitly that the meal was provided daily when school was in session. The randomised controlled trial in China (Kleiman-Weiner *et al.*, 2013) was the only in-school programme that also provided children with a snack for the weekend.

Nutritional content: An important variation between interventions was the nutritional content of the meal or snack provided. For the food ration to correct nutritional deficiencies it is assumed that it is rich in energy, protein and/or vitamins and addresses a nutritional deficiency that is suffered by the target group (Greenhalgh *et al.,* 2009). Of the studies that presented information on the calorific content of food (n=7), content varied from a 240 kcal snack provided in the school feeding trial in Kenya (Omwami *et al.*, 2011) up to 1000 kcal per day meal in the Chilean feeding programme, Programa de Alimentación Escolar (PAE). The Mid day Meal Scheme in India mandated the provision of at least 300 calories per child and 8-12 grams of protein, but ultimately left it up to the implementing state-level governments to decide on the meal provided. Jacoby *et al.*'s trial of school breakfasts in Peru provided 30 per cent of daily energy requirements (WHO requirements), 60 per cent of RDA for minerals and vitamins and 100 per cent of iron needs. The programme in Cambodia and the standard feeding programme in Sri Lanka (He, 2010a) provided similar food provisions, made up of typical WFP rations such as rice, vitamin A-fortified vegetable oil and iodised salt. In the feeding programmes in Guyana and Sri Lanka, the content of the meals were left up to the discretion of the local communities (Ismail *et al.*, 2012; He, 2010b). Three of the impact evaluations provided no information about the nutritional content of the meals (Tan *et al.,* 1999; Adrogue *et al.,* 2011; Kazianga *et al.*, 2012).

The WFP take-home rations intervention arm in Laos gave children monthly rations to take home, conditional on attendance, as well as a one off provision at the beginning and the end of the year (Buttenheim *et al.,* 2013). The WFP take home rations programme in Burkina Faso gave girls monthly rations of food, conditional on 90 per cent attendance at school. Meals provided in-school are typically are more nutritious than take- home rations, which are primarily cereals and oils (Adelman *et al.*, 2008). This was true of the two included take-home rations program. In the programme in Burkina Faso, children received 10kg bag of cereal flour every month that the attendance conditionality was met. In the programme in Laos, students received 15kg of rice upon enrolment at the beginning of the year and 30kg of rice at the end of the school year if the attendance goal was met, and a can of fish if monthly attendance was also 80 per cent.

School level implementation: In the majority of the programmes, school teachers or members of the local community, for example Parent-Teacher Associations or specific school feeding management committees, were responsible for the day-to-day implementation of the school feeding programme (n=10). The WFP programme in Laos required substantial input from the target community (Buttenheim *et al*., 2013), where villages had to convene a school feeding committee, build food storage facilities, prepare the foods, and to travel to WFP food distribution points to pick up food. The Hinterland Community-Based SFP in Guyana reported in Ismail et a. 2012 was intended to both improve student's nutrition and learning outcomes, as well as build more community participation in schools. The implementation of the programme was therefore left entirely up to the community, who were required to submit proposals to receive funding for the feeding programme, receive training and then able to purchase and prepare food themselves. Similarly, the welfare feeding programme in Sri Lanka reported in He (2010) distributes payments to parents of school students to purchase food and prepare school meals.

The preparation of meals in the Mid-Day Meal Scheme in Delhi was outsourced to external providers (Afridi, 2014); this is compared to the Scheme in the rest of the country that was for the part managed by village governments who often delegate implementation to local charities or Parent-Teacher Associations (Jayaraman and Simroth, 2015). The PAE programme in Chile is unique in this regard as the entire process throughout the country is managed by private companies, from the purchase of ingredients to the actual distribution (Epstein *et al*., 2004).

Comparisons

All but two of the included studies (n=14) compared the effect of an intervention to business as usual (that is, a comparison group with no school feeding intervention). Powell *et al.*'s (1998) study in Jamaica provided children in the comparison group with a placebo school breakfast (a quarter of an orange). Afridi's (2014) evaluation of the Midday Meal Scheme in Delhi compared the introduction of a meal to the existing feeding programme, which provided packaged food rations.

Several of the studies included multiple treatment arms. The programme in the Philippines included four treatment arms, two of which included a school feeding intervention. The first treatment arm simply provided a free school meal while classes were in session, while the second provide the free meal alongside an intervention to encourage parent-teacher partnerships. Buttenheim *et al.*'s (2013) study in Lao PDR evaluated three different treatment arms versus one comparison; one treatment arm providing in school meals, one arm providing students with take home rations and one arm combining the two previous treatments.

Similarly, Kazianga's (2012) study in Burkina Faso had two treatment arms, one providing meals in school canteens and one providing take home rations. Cheung and Berlin (2014) evaluation in Cambodia reports results for the feeding programme at three different stages; when the original WFP school feeding programme was introduced, when it was expanded to include take-home rations for girls and when it was further expanded to be combined with deworming and complementary health activities. In all cases, we chose the treatment for the meta-analysis that was most similar to other treatments in the included evaluations. In all cases this was the in-school feeding intervention. However, we have reported effect sizes for all treatment arms in the narrative synthesis and in full in Appendix H. Omwami's (2011) experimental trial in Kenya tested the impact of three different types of food provided as a mid-morning meal through three different treatment arms. In this case, we combined the effect across all three treatment arms.

Outcomes

The included studies reported on a wide range of education outcomes. The most commonly reported outcome group was achievement outcomes (n=10). Ten of the included studies reported impacts on a maths/numeracy outcome; these were the studies in the Philippines (Tan *et al.*, 1999), Chile (McEwan, 2013), China (Kleiman-Weiner *et al.*, 2011), Peru (Jacoby *et al.*, 1996), Burkina Faso (Kazianga *et al.*, 2013), Jamaica (Powell *et al.,* 1998), Senegal (Diagne *et al.*, 2014), Kenya (Omwami *et al.*, 2011), Guyana (Ismail *et al.*, 2012) and Argentina (Adrogue *et al.*, 2011). Eight studies reported on a literacy and language arts outcomes (Tan *et al.*, 1999; McEwan, 2013; Jacoby *et al.*, 1996; Powell *et al.,* 1998; Diagne *et al.,* 2014; Adrogue *et al.*, 2011, Omwami *et al.*, 2011; Ismail *et al.*, 2012). We included any measures of language artsor literacy in the meta-analysis, however there was a mix of indicators used across the studies. Six used data on test scores from nationally or officially administered language arts examinations (McEwan, *ibid*; Adrogue *et al.*, *ibid*; Tan *et al.*, *ibid*; Omwami *et al.,* 2011; Diagne *et al.*, 2014; Ismail *et al.*, 2012). The remaining two studies used researcher- administered literacy tests to assess children (Powell, *ibid*; Jacoby *et al.*, *ibid*). Powell *et al.* (1998) administered tests of reading and spelling. Jacoby *et al.* (1996) gave students tests to students in grades 4 and 5 to assess reading comprehension and vocabulary. In cases where two or more language arts results were presented, for example English and Filipino, we included the local language spoken by the most number of people in the meta-analysis.

In addition to subject specific achievement, three studies in the Philippines (*ibid*), Kenya (*ibid*) and Senegal (*ibid*) reported on a composite outcome measure of achievement. Three studies reported on an outcome related to cognitive ability. These were the studies in Peru (*ibid*), Senegal (*ibid*) and Burkina Faso (*ibid*). Among these studies, the indicators used were fairly diverse. Kazianga *et al.* (2013) report results for the Raven's progressive matrices test, a common test of non-verbal ability. Diagne *et al.* (2014) report results for four measures of general cognitive ability; memorisation capacities, level of knowledge, level of comprehension and reasoning capacities. Jacoby *et al.* reports results of a coding test for grades 4 and 5 children that tested visual perceptual organisation and visual motor coordination.

In addition to achievement and cognition related outcomes, eight studies reported on enrolment status (Buttenheim *et al.*, 2013; Kazianga *et al.*, 2013; Jayaraman and Simroth., 2015; McEwan, 2013; Ismail *et al.,* 2012; Cheung and Berlin, 2014; He, 2010a; He, 2010b) and seven studies reported on a measure of student attendance (Kazianga *et al.,* 2013; McEwan, 2013; Powell *et al.*, 1996; Omwami *et al.,* 2011; Cheung and Berlin, 2014; Ismail *et al.*, 2012; Afridi, 2014). The three studies from Argentina, the Philippines and Senegal reported on a measure of dropout. Two studies reported on repetition rates (Diagne *et al.*, 2011; McEwan, 2013).

The follow up period for many of these interventions was fairly short, in many cases between 9 and 24 months. The shortest follow up period was 30 days in Jacoby *et al.*'s RCT of a School Breakfast programme in Peru. McEwan's (2013) evaluation of the PAE in Chile used administrative schooling data which allowed him to evaluate the programme over a longer period of time; approximately 36 months for achievement outcomes and 48 months for enrolment.

Study Design

We identified a mix of experimental and quasi experimental studies. Only one of the included studies was a randomised control trial (Powell *et al.,* 1998) while six were cluster-randomised controlled trials (Tan *et al.,* 1999; Jacoby *et al.*, 1996; Kazianga *et al.,* 2012; Kleiman-Weiner, 2013; Omwami *et al.*, 2011; Diagne *et al.*, 2014). In each of these studies, assignment to the programme was done at the school or village level.

Eight of the studies were controlled before and after studies, making use of baseline and endline data collection and a comparison group (Buttenheim *et al.*, 2013; Jayaraman and Simroth, 2015; Afridi, 2014; Ismail *et al.*, 2012; Cheung and Berlin, 2014; Adrogué *et al.*, 2011, He, 2010a, He, 2010b). Adrogue *et al.* (2011), Afridi (2014), Cheung and Berlin, (2014) and Jayaraman and Simoroth (2015) used administrative data to evaluate the impact of the programmes. Buttenheim *et al.*'s study in Laos used data from a longitudinal survey undertaken to evaluate the WFP feeding programme. Ismail *et al.*'s (2012) study in Guyana combined survey data collected as part of the trial with administrative data. Finally, the study in Chile (McEwan, 2013) used a regression discontinuity design. In this paper, McEwan exploits the fact that the Chilean government assigned varying calorific content for their national school meals programme, using a vulnerability index with defined cut-off points for assignment to each calorie level within the programme.

Qualitative studies, process evaluations and project documents

We identified eighteen additional qualitative, mixed methods studies, process evaluations and project documents related to the eleven programmes covered by the included impact evaluations. For most programmes, there was limited additional information identified for the included programmes outside of the impact evaluations. The exception is the Mid-day Meal Scheme in India. For this programme we identified six additional documents (three mixed methods studies and three project documents) assessing the implementation of the programme in one or more Indian states.

Table 4.2 a: Characteristics of Included Evaluations of School Feeding Programmes

Included study	Setting	Description of the intervention	Included outcomes	Length of follow up	Study design	Sample Size
Tan *et al.* (1999)	Philippines (rural) Primary Grades: 1-6 Age: not reported	Dropout Intervention Programme (DIP): The DIP was a pilot programme introduced by the Philippine government, with four intervention arms, two of which included school feeding. Under the first treatment arm, all pupils in beneficiary schools received a free school meal while classes were in session. In the second treatment arm, pupils again received a free school meal, but improved parent-teacher partnerships were also introduced. These partnerships comprised a series of regular (usually monthly) group meetings throughout the school year between school staff and parents. The school feeding programme was often the substantive issue discussed at these parent-teacher meetings.	Dropout; maths test scores; language arts test scores	Approximately 12 months	Cluster RCT – using DID and IV	1609 children (school feeding interventions only)
Powell *et al.* (1998)	Jamaica (rural) Primary Grades: 2-5 Age: average age 8.9 years	Breakfast Feeding Trial (No programme name): The trial tested the effects of providing school breakfast to primary school children. Children in the treatment arm (half of whom were undernourished group and half of whom were nourished) were given cheese sandwich or spiced bun and cheese and flavoured milk for every school day for one school year. Children in the control arm were given one quarter of an orange as a placebo for every school day for one school year. All meals were served before the start of classes on all school days.	Attendance; maths test scores; language arts test scores	Not clear - length of the school year	RCT – using fixed effects regression	791 children

Included study	Setting	Description of the intervention	Included outcomes	Length of follow up	Study design	Sample Size
Jacoby *et al*. (1996)	Peru (peri-urban) Primary Grades: 4-5 Age: average age of 10.5 years in the treatment group, 11.57 years in the control group	School breakfast Programme (BSP) Peru: Students in participating schools received a school breakfast consisting of four cookies and an instant drink. The nutritional content contained in the breakfast was on average 30 per cent of daily energy requirements (WHO requirements), 60 per cent of RDA for minerals and vitamins and 100 per cent of iron needs.	Maths test scores; languages arts test scores; Cognition scores	30 days	Cluster RCT – using other regression techniques	352 children
Buttenheim *et al*. (2013)	Laos PDR (rural) Primary Grades: not reported Age: 6-14 years	WFP School Feeding Programme Laos: The World Food Programme school feeding programme ran on-site feeding, take-home rations, and a combination of the two. The in school programme provided a mid-morning soy-corn snack to school children every morning that they attended school. Schools providing take home rations provided rations conditional on 80 per cent school attendance. These consisted of 15 kilograms of rice upon enrolment at the beginning of the year and 30 kilograms of rice at the end of the school year if the attendance criteria was met, and a can of fish if monthly attendance was at 80 per cent.	Enrolment	Between 12-24 months, due to variation in the start date of the intervention as a result of implementation issues in some schools.	CBA (quasi-experiment with baseline and endline data collection) – using DID and PSM	5667 children across the three treatment groups and control

Included study	Setting	Description of the intervention	Included outcomes	Length of follow up	Study design	Sample Size
Adrogué *et al*. (2011)	Argentina (rural, urban, peri-urban) Primary Grade: 3 Age: not reported	School feeding programmes Argentina: In-school feeding programmes in Argentina are funded by the national government level but are decentralised and vary in characteristics by province. Modalities of delivery of school feeding varies by provinces; including programmes where teachers, cooks and other school members buy the food and decide the daily menu, those where the Provinces or the Municipalities buy the food and the school prepares it, where organisations buy and prepare the meals and then they distribute them to each school and where firms cater in schools	Maths test scores; language arts test scores; dropout	Not clear - up to 36 months	CBA– using DID	3516 public schools
Kazianga *et al*. (2012)	Burkina Faso (rural) Primary Grades: not reported Age: 6-15 years	WFP School Feeding Programme Sahel Region: There were two different treatments provided under this cluster RCT, school meals in canteens and take home rations. In schools providing meals, lunch was served to children on each school day in the school canteen, with the only requirement that the child be present at school. Both boys and girls were eligible for the meals. Take-home ration schools gave girls that attended school for 90 per cent of the time a ration of 10kg of cereal flour to take home. This food could be shared by the households, possibly reaching children who may be in as much or greater need of the food as the student themselves. These were provided on a monthly basis.	Enrolment; maths test scores; cognition scores; attendance	Approximately 9 months (from beginning to the end of the 2006-2007 school year).	Cluster RCT – using DID	4236 children across the two treatment groups and control group

Included study	Setting	Description of the intervention	Included outcomes	Length of follow up	Study design	Sample Size
Kleiman-Weiner (2013)	China (rural) Primary Grade: 4 Age: average age of 10.5 years in the treatment, 10.4 in the control group	No official name 'Egg a day' arm of the study in Gansu province: The treatment consisted of the provision of a cooked egg a day for children in the fourth grade of primary school. The eggs were given to the children by their class teacher to be eaten during the first period of the school day. The researchers provided a small honorarium to the principals in all schools and the homeroom teachers of the fourth graders in the treatment schools; an additional subsidy of 50 yuan a week was provided to the school administration to cover cooking costs.	Maths test scores	7 months	Cluster RCT – using Fixed effects	2686 children
Omwami *et al*. (2011)	Kenya (rural) Primary Grade: 1 Age: average age of 7 years at the beginning of the study	Child Nutrition Project (CNP): This field experiment tested the impact of providing a mid-morning meal during break, either Meat-Githeri, Milk-Githeri, or Energy-Githeri (a Kenyan traditional meal of maize and any type of beans mixed and boiled together). These provided 240 kcal in the first school year and 313 kcal for the remainder of study period.	Attendance; maths test scores; language arts test scores	21 months	Cluster RCT – using comparison of mean differences	554 children across three treatment groups and control

Included study	Setting	Description of the intervention	Included outcomes	Length of follow up	Study design	Sample Size
Jayaraman and Simroth (2015)	India (rural, urban, peri-urban) Primary Grades: 1-5 Age: average age between 8-11 years	Midday Meal Scheme: The Midday meal scheme in India delivers hot midday meals to school children in government run schools. During the time period covered by the impact evaluation, the meal was mandated to provide a minimum of 300 calories and 8-12 grams of protein each day of school for a minimum of 200 days a year. The implementation of the programme was the responsibility of the Indian state-level government, and they are allowed to supplement this minimum food provision as they see fit.	Enrolment	Between 12 - 24 months	CBA – using PSM and DID	Approximately 34,162 schools
Afridi (2014)	India (urban, per-urban) Primary Grades: 2-5 Age: not reported	Midday Meal Scheme (Delhi): The Midday meal scheme in India delivers hot midday meals to school children in government run schools. During the time period covered by the impact evaluation, the meal was mandated to provide a minimum of 300 calories and 8-12 grams of protein each day of school for a minimum of 200 days a year. The implementation of the programme was the responsibility of the Indian state-level government, and they are allowed to supplement this minimum food provision as they see fit. In Delhi, where this study took place, all meal preparation was outsourced to external providers.	Attendance	Approximately three months	CBA – using DID and Fixed effects	19 schools, 1591 children
Diagne *et al*. (2014)	Senegal (rural) Primary grades: 2 and 4	WFP School canteen programme Senegal: The programme provided hot lunches through school canteens set up in some primary schools in rural Senegal. The food basket included the appropriate caloric composition made up of maize, three legumes, vegetable oil and iodized salt, as recommended by the	Completion; dropout; maths test scores; language arts test scores;	13 months	Cluster RCT – using DID	2917 students

Included study	Setting	Description of the intervention	Included outcomes	Length of follow up	Study design	Sample Size
	Age: average age of 10 years in the treatment group 10.36 years in the control group.	WFP, UNESCO and WHO. The WFP provided the food supplies each term. Parents were required to contribute 200 FCFA a month per student for other products needed for the functioning of the school meal programme but which are not included in the WFP food basket	cognition scores			
McEwan (2011)	Chile (rural, unclear where else in the country) Primary Grades: 1-8 Age: not reported	Programa de Alimentación Escolar (PAE): Under the PAE programme, children at eligible publicly funded schools are delivered free meal rations with varying caloric content, depending on their measured “vulnerability".	Enrolment; attendance; maths test scores; language arts test scores; completion	Achievement/ repetition: approximately 36 months Enrolment: approximately 48 months Attendance: approximately 24 months	RDD	School level outcomes (enrolment, attendance): 4469 schools Student Level: (1) completion: 63,336 students (2) Achievement: Language arts test: 34,237; Maths test: 34,162 students

Included study	Setting	Description of the intervention	Included outcomes	Length of follow up	Study design	Sample Size
Cheung *et al*. (2014)	Cambodia (unclear if rural and/or urban) Primary Grades: 1-6 Age: not reported	WFP Food for Education Programme Cambodia: The WFP school feeding programme in Cambodia was phased in, in three stages, gradually introducing more additional interventions to the basic in-school feeding programme. The pilot project started with only in-school feeding, providing children with breakfast before school, which contained rice, canned fish, vitamin A-fortified vegetable oil and iodised salt. In 2001-2002, the programme expanded to include take-home rations for girls in grades 4 to 6 as an incentive to keep these girls in school. The programme expanded further in 2002-2003, introducing a deworming programme to participating schools, as well as complementary health and sanitation activities to improve the education environment; this included identification of safe drinking water, education on making improvements in basic health, hygiene and sanitation practices for students at school and at home.	Enrolment; attendance	Between 12 - 48 months (three follow up periods)	CBA – using DID, Fixed effects	Enrolment: 1053 schools, 1706 schools, 1934 schools, depending on the treatment year
Ismail *et al*. (2012)	Guyana (rural) Primary Grades: 2-6 Age: not reported	Hinterland Community-based School Feeding Programme (SFP): The Hinterland Community-Based SFP intended to both improve children's nutrition and learning outcomes, as well as build more community participation in schools. In order to participate in the programme, schools and communities needed to submit a proposal to receive funding for the programme, take training in financial administration, food hygiene and nutritious meal preparation. Local farmers were required to provide foodstuffs for the	Enrolment; attendance; maths test scores; language arts test scores	Approximately 24 months	CBA – using PSM; fixed effects; two stage method	3877 students

Included study	Setting	Description of the intervention	Included outcomes	Length of follow up	Study design	Sample Size
		meals. The SFP intended to finance meals from a snack or sandwich plus fruit and drink to a full hot lunch.				
He (2010a)	Sri Lanka (presumably rural, urban and peri-urban) Primary and Secondary Grades: 1-9 Age: not reported	WFP School Feeding Programme Sri Lanka (The standard WFP programme): WFP delivers food directly to schools to be disbursed to students in all grades. They provide rice, dhal, and fortified oil and curry/leaf vegetables for meals three days a week, and a fortified corn soya blend mixture two days a week. The corn mixture is fortified with proteins, vitamins A, D, E, B complex, folate, C, and micro and macro minerals and provides 380 Kcal of energy for students. In addition, the programme provides students with de-worming, school health visits and examinations and nutrition education.	Enrolment	Approximately 48 months	CBA – using DID, Fixed effects	Standard programme: 62,323 school grades
He (2010b)	Sri Lanka (presumably rural, urban and peri-urban) Primary and Secondary Grades: 1-9 Age: not reported	WFP School Feeding Programme Sri Lanka (The welfare programme): the programme provides mid-morning meals by distributing funds to Samurdhi (welfare) recipients and local parents to cook and distribute the food in the beginning of the school day. Welfare recipients receive 17 RS (US$0.17) for each student served. Selection and expansion were made by MOE officials to target schools with high malnutrition rates. Those in vulnerable areas were also given priority.	Enrolment	Approximately 48 months	CBA – using DID, Fixed effects	Welfare programme: 84,093 school grades

Table 4.2 b: Intervention Design Features of Included Studies

	Tan Philippines	**Powell** Jamaica	**Jacoby** Peru	**Buttenheim** Laos PDR	**Adrogué** Argentina	**Kazianga** Burkina Faso	**Kleiman-Weiner** China	**Omwami** Kenya	**Afridi** India	**Jayaraman** India	**Diagne** Senegal	**McEwan** Chile	**Ismail** Guyana	**Cheung** Cambodia	**He** Sri Lanka (a)	**He** Sri Lanka (b)
Intervention components																
In-school feeding	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Take home rations				✓		✓								✓		
Other	✓												✓	✓	✓	
In-school feeding characteristics																
Meal	✓	✓	✓		✓	✓			✓	✓	✓	✓	✓	✓	✓	✓
Snack				✓			✓	✓				✓	✓		✓	
Timing of in-school feeding																
Breakfast		✓	✓		✓		✓					✓		✓		
Mid-morning				✓				✓					✓			✓
Lunch	✓				✓	✓			✓	✓	✓	✓	✓			
Unclear															✓	
Intervention objectives																
Improve school participation	✓	✓	✓	✓		✓			✓	✓		✓	✓	✓	✓	✓
Reducing hunger						✓					✓					
Nutritional status		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Achievement	✓	✓	✓	✓		✓	✓		✓	✓	✓	✓			✓	✓
Reduce poverty									✓	✓	✓	✓				
Other									✓	✓			✓		✓	✓

4.2.3 *Synthesis of findings*

The results of our synthesis for school feeding interventions is presented in two sections. First, we present the findings of the meta-analysis on the effects of school feeding programmes on primary and secondary outcomes and results for different population sub-groups where available. We then present a discussion of the findings incorporating evidence from our descriptive qualitative synthesis of intervention design, implementation and context factors that may influence the effectiveness of these programmes (questions 2a and 2b).

Effects of school feeding programmes on enrolment, attendance, dropout rates, completion and learning outcomes

This section reports the results of the meta-analysis of the effects of school feeding interventions, addressing question 1a and 1b of the review. We have structured the presentation of results according to the 'ideal type' theory of change, starting with education access outcomes (enrolment, attendance, drop out) and followed by final outcomes (completion, cognitive scores and learning outcomes: composite test scores, language arts test scores, maths test scores).

All sixteen included studies evaluated programmes or trials that provided an in-house school feeding component and in all cases we included effect sizes for this component in the meta-analysis. We were able to calculate effect sizes for all sixteen studies. Below, we detail some of the decisions on the meta-analyses for school feeding:

As described in the previous section, two studies in Laos and Burkina Faso included a treatment arm that provided take home rations (Buttenheim *et al.*, 2013; Kazianga *et al.,* 2012). In meta-analyses for all relevant outcomes, we included the treatment arm that involved in-school feeding as this was the most common intervention delivery type across the universe of studies but we have also presented effect sizes for other treatment arms and compared where different.

The two evaluations of feeding programmes in Sri Lanka, both reported in He (2010), share a comparison group and so results from both could not be included in the meta-analysis. We decided to include He (2010)a in the meta-analysis, the standard WFP feeding programme, as the intervention in this study was most similar to the other included programmes . We have however reported effect sizes for the other programme, He (2010b) below and in full in Appendix H.

Cheung and Berlin (2014) report on a WFP programme in Cambodia that gradually phased in more components to a school feeding programme, including deworming and complementary health and sanitation activities. We decided to include the results for the basic school feeding programme in our meta-analysis of enrolment outcomes, but once again also report results for the more comprehensive components in Appendix H.

Afridi (2014) reported on enrolment and school attendance but we did not include the results in the meta-analyses as the study did not have a comparison group no receiving school feeding. Tan (1999) reported on a trial in the Philippines that included four treatment arms, two of which included a school feeding intervention. The first treatment arm provided a free school meal while classes were in session, while the second provided a meal alongside provisions for improved parent-teacher partnerships. We included the first treatment arm in the meta-analyses of all relevant outcomes but report results for the other treatment arm separately.

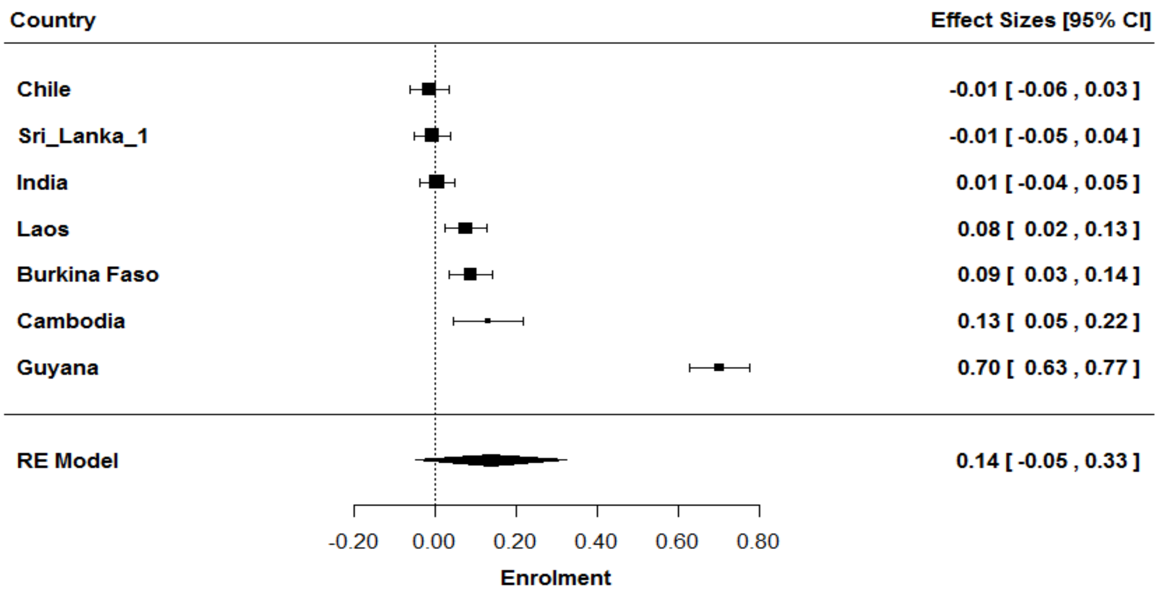
In cases where numerous measures of language arts, mathematics or cognitive scores were presented (for example reading and spelling for language arts), where appropriate we combined the measures to create one synthetic effect size, for example for the measures of spelling and reading in the evaluation of the feeding trial in Jamaica (Powell, 1998). In the case of studies reporting multiple language arts outcomes, for example in the evaluation of the feeding programme in the Philippines (Tan, 1999) which report results for English and Filipino, we chose the local language or language spoken by the most number of children in the sample to include in the meta-analysis but also present effect sizes for the other outcomes.

All effect sizes are expressed as standardised mean difference (SMD), interpreted as the magnitude of the number of standard deviation changes in the outcome for the intervention group as compared to students in non-school feeding schools. SMD scores are interpreted as the number of standard deviation changes in the outcome.

Enrolment

Figure 4.2b presents the forest plot with the results of the individual school feeding studies and the pooled point estimate on student enrolment in school. The overall average effect of school feeding on enrolment is 0.14, 95% CI [-0.05, 0.33], calculated under a random effects model. The homogeneity test suggests a very large amount of between-study heterogeneity (I^2 = 98.83%, τ^2 = 0.0624). The forest plot in Figure 4.2b suggests that a large amount of this heterogeneity is driven by the study from Guyana. Sensitivity analysis revealed that the average point estimate fell to 0.04 SMD and the homogeneity test to 77.26 per cent when this study was removed.

Figure 4.2 b: Enrolment 5



The two treatment arms of the studies in Burkina Faso (Kazianga *et al.*, 2012) and Laos (Buttenheim *et al.,* 2011) were left out of the meta-analysis as they provided take-home rations rather than in-school feeding, found similar effects for take-home rations and in school feeding (Burkina Faso: 0.11, 95% CI [0.06, 0.17]. Laos: 0.10, 95% CI [0.05, 0.16]). The treatment arm with community implementation in Sri Lanka was not included in the meta-analysis as the standard school feeding programme was more similar to the other studies. The effect of the community implemented programme was larger than that for the standard feeding programme (He, 2010b: 0.10, 95% CI [0.06, 0.15]). After the first year of treatment there was a larger effect on enrolment for the treatment schools in Cambodia that also received de-worming and complementary health activities (Cheung and Berlin, 2014, SMD= 0.19, 95% [0.13, 0.25]).

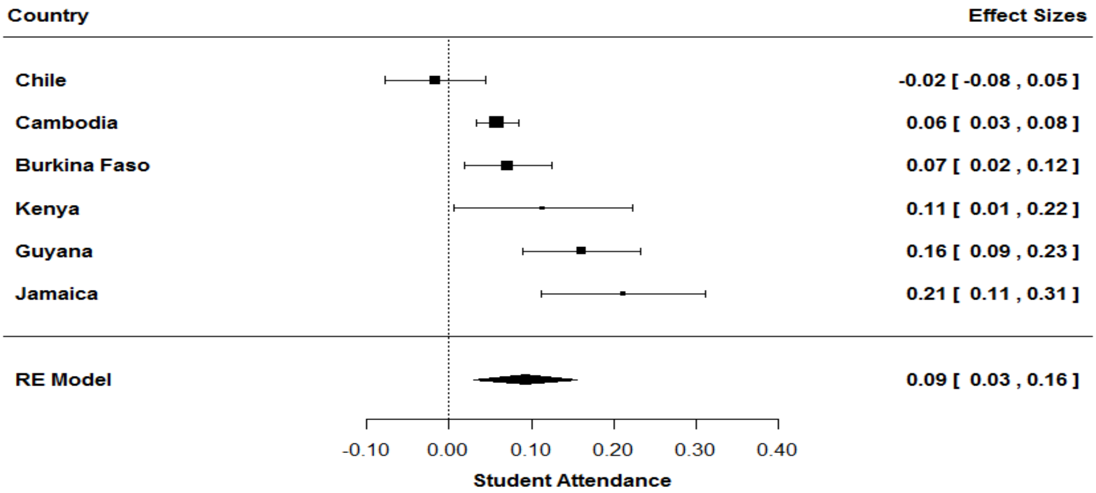
We conducted several sub-group analyses by school grade and gender. For most of the grade level sub-group analyses we only identified two studies, in India and Sri Lanka (Jayaraman and Simroth 2015; He 2010a). The meta-analysis of these two studies resulted in smaller overall estimates than the full sample presented above, ranging between -0.03 95%CI [-0.08, 0.02] for grade 3 students and 0.01 95% [-0.02, 0.03] for grade 1 students. These are in line with the results for the Sri Lankan and Indian studies in the analysis of the full samples. The sub-group analyses for girls and boys found similar effects on enrolment to the analysis of the full sample, however these only included results from two programmes in Laos and Burkina Faso (Buttenheim *et al.,* 2013; Kazianga *et al.*, 2012; Boys: SMD=0.11, 95% CI [0.05, 0.16]; Girls: SMD=0.12, 95% CI [-0.03, 0.28]).

Attendance

The overall weighted average effect of school feeding on student attendance is 0.09, 95% CI [0.03, 0.16)], calculated under a random effects model. The tests of homogeneity suggest a large amount of between studies variability, with an I^2 of 84.58% and τ^2 = 0.0048. This heterogeneity is also apparent from Figure 4.2c. Effect sizes range between -0.02, 95% CI [-0.08, 0.05] in Chile to 0.21, 95% CI [0.11, 0.31] in Jamaica. Results were fairly sensitive to the removal of the study with the negative effect; for example, when we remove the study from Chile, the point estimate increases to 0.11 (SE =0. 0.0284, 95% CI [0.06, 0.17]).

The study from Burkina Faso also included a treatment arm for take home rations treatment that we did not include in the meta-analysis as it was too different from the others. The effect size for this treatment was slightly larger in magnitude than the overall average effect (SMD= 0.10, variance = 0.0007). We also calculated effect sizes for attendance from the study comparing the Mid-Day Meal Scheme in India to packaged food rations. The authors report an effect of 0.11 (95% CI [0.04, 0.18]), which is slightly larger than the overall weighted effect. We did not have enough studies to undertake a meta-analysis of sub-groups of children, for example by gender, age or grade.

Figure 4.2 c: Student Attendance

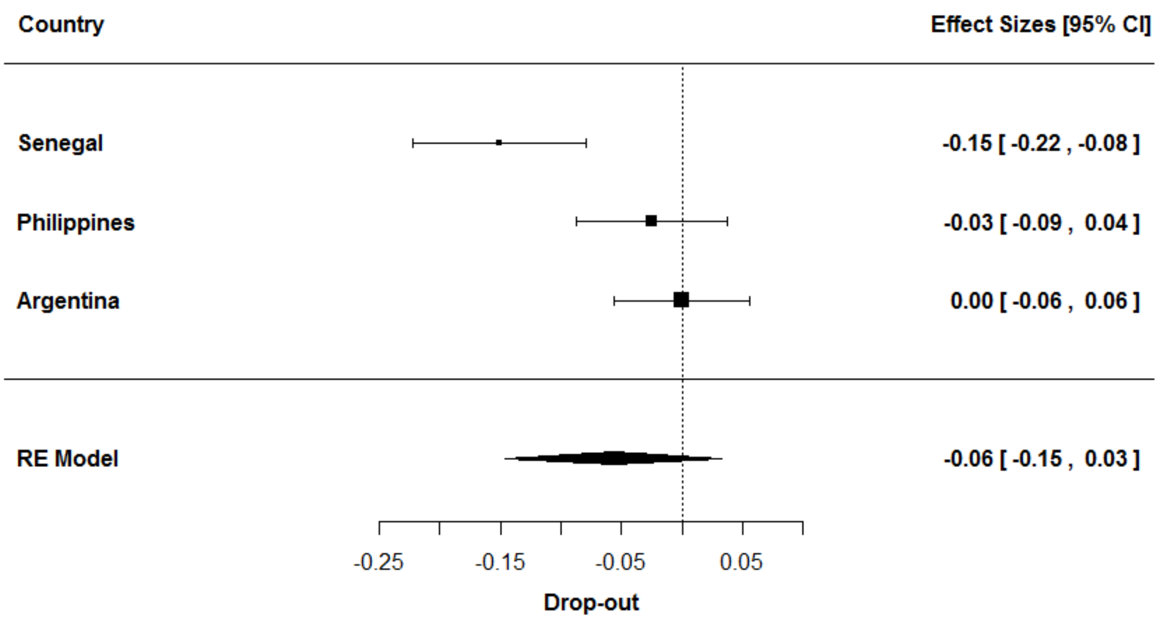


Dropout

We included three studies in our meta-analysis that evaluated the effect of school feeding on student dropout. Figure 4.2d presents the forest plot with the results of the individual studies and the pooled point estimate. The overall weighted average effect of school feeding on student attendance at school was -0.06, 95% CI [-0.15, 0.03], calculated under a random effects model. It should be noted that a negative sign for an effect size for dropout should be interpreted as a reduction in dropout rates.

The tests of homogeneity suggest a large amount of between studies variability (I^2 = 83.52%, τ^2 = 0.0053 and Q(df = 2) = 11.225, p-value = 0.0037). Effects vary between zero (95% CI [-0.06, 0.06]) in Argentina to a reduction in dropout rates of -0.15 (-0.22, -0.09) in Senegal. The results are sensitive to the removal of the study from Senegal; removal of this study causes the point estimate to change to -0.01 (SE = 0.0213). The I^2 also changes to 0.00, suggesting that the only source of variation between the two remaining studies is within-study sampling error.

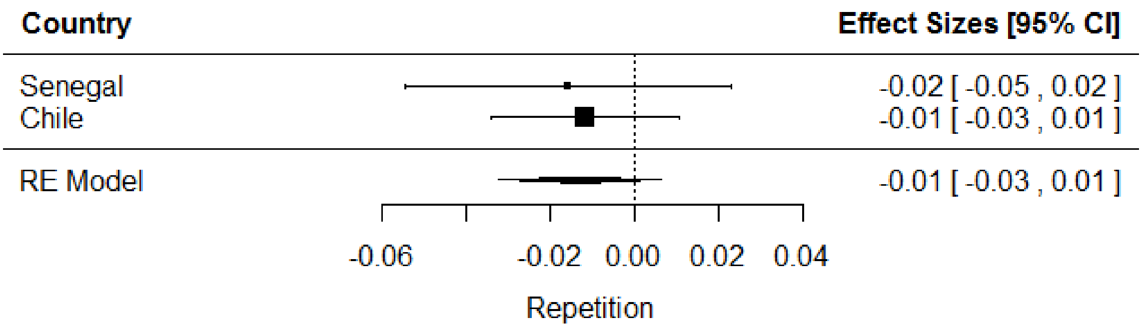
Figure 4.2 d: Dropout



Completion

We only had two included studies that evaluated the effect of school feeding on completion, as measured by grade repetition rates. Figure 4.2e presents the forest plot with the results of the individual studies and the overall estimate. The overall average effect of school feeding on repetition is almost zero (SMD=-0.01, SE= 0.0099, 95% CI [-0.03, 0.01]). As can be seen from the forest plot, there is little variability between the only two included studies and this is also confirmed by the homogeneity test (I^2 = 0.00%, τ^2 = 0).

Figure 4.2 e: Repetition



Cognitive scores

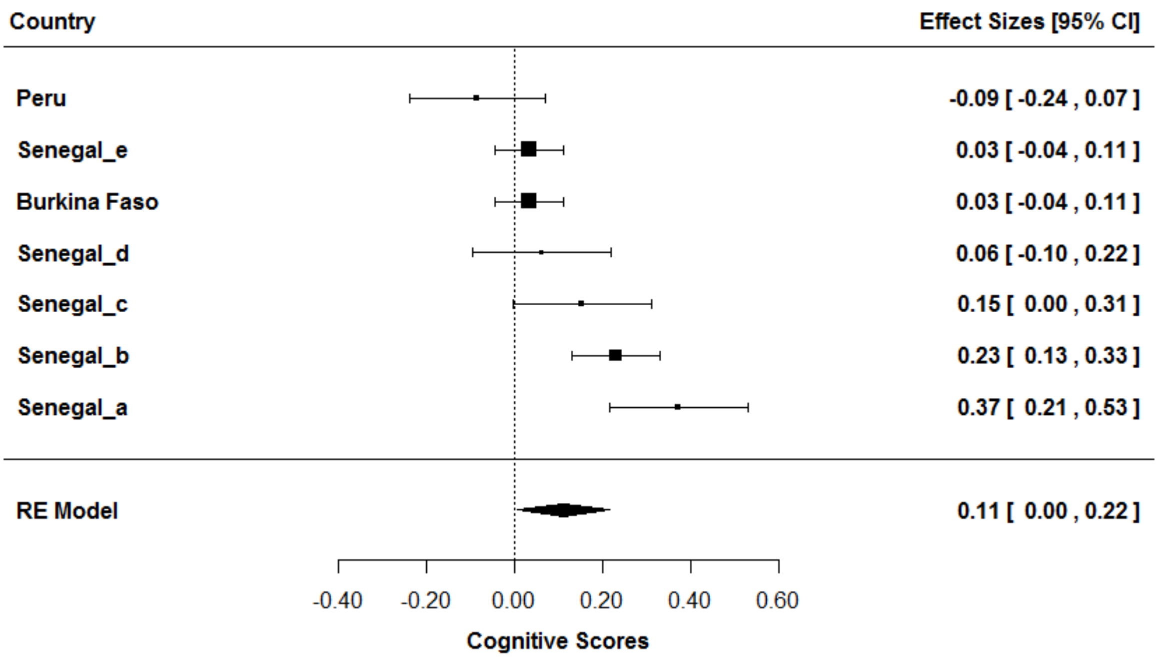
Three studies reported on a measure of cognitive scores. The overall weighted average effect of school feeding on cognitive scores was 0.11, (95% CI [0.00, 0.22]), calculated under a random effects model. The tests of homogeneity suggest a large amount of between studies variability (I^2 = 83.34%, τ^2 = 0.0163). This can also be observed in the forest plot in figure 4.2f. Effect sizes vary from -0.09 (95% CI -0.24, 0.07) from the study in Peru, to 0.37 (95% CI, [0.21, 0.53]) in Senegal.

Five of the effect sizes in this meta-analysis come from the same school canteen programme in Senegal (observations Senegal_a to Senegal_e[7]), but correspond to different grade and age groups and thus form independent samples (Diagne *et al.,* 2014). Even within this programme there is significant variation in the effect of school feeding on cognitive scores. The largest effect is found on children aged 6-7 in grade 2 of primary school (Senegal_a: 0.37, 95% CI 0.21, 0.53) and the smallest effect on children aged 10 and over in grade 4 (Senegal_e: 0.03, 95% CI [-0.04, 0.11]).

The overall results are sensitive to the removal of the two largest observations from Senegal, Senegal_a and Senegal_b, reducing the overall estimate to 0.07 (SE = 0.043) and 0.09 (SE = 0.0593) respectively.

[7] Senegal_a corresponds to children in grade 2 of primary school aged 6-7. Senegal_b corresponds to children in grade 2 aged 8-9. Senegal_c corresponds to children in grade 2 aged 10 and over. Senegal_d corresponds to children in grade 4 aged 8-9. Senegal_e corresponds to children in grade 4 aged 10 and over.

Figure 4.2 f: Cognitive scores[8]

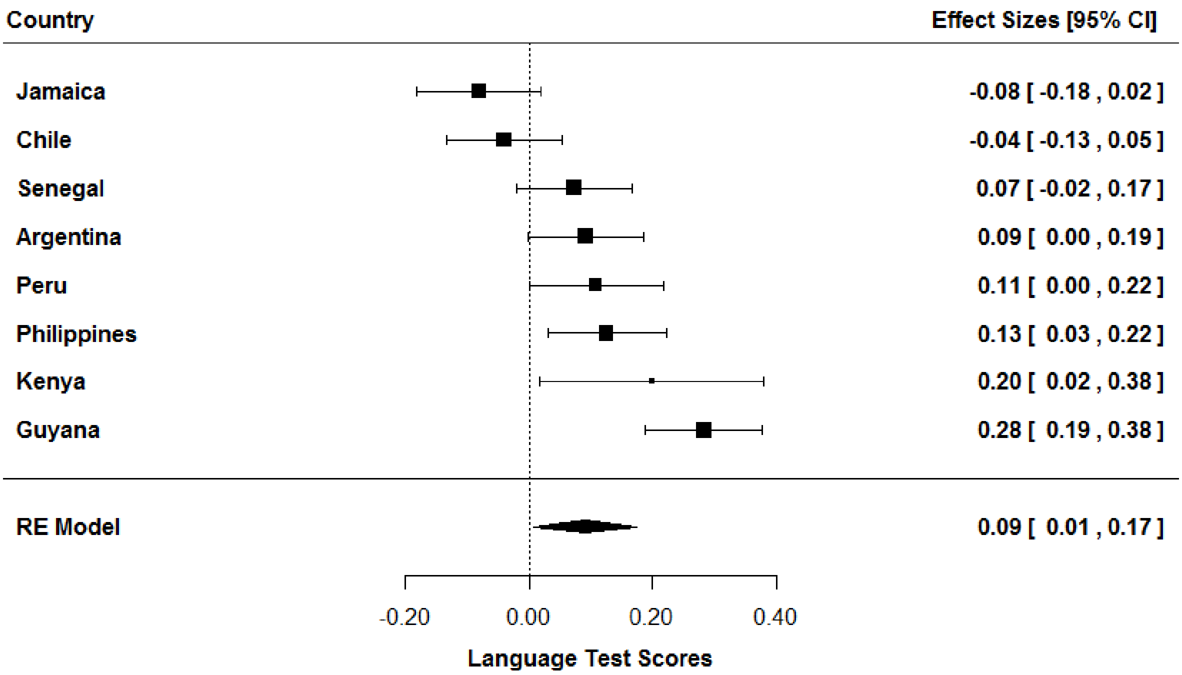


Language arts scores

The overall weighted average effect of school feeding on learning as measured by language arts test scores was 0.09 (95% CI, 0.01, 0.17). The assessments of homogeneity suggest a high amount of between-studies variation (I^2 = 80.49%, τ^2 = 0.0112, Q (df = 7) = = 36.8236, p-value =<0.0001). As can be observed in Figure 4.2g, the studies of programmes in Jamaica (Powell *et al.,* 1998) and Chile (McEwan, 2013) report a negative effect of school feeding on language arts test scores, whereas the remaining six studies found positive effects of different magnitudes. The results are sensitive to the removal of several studies (see Appendix H for results of all sensitivity analyses). For example, the removal of the study from Jamaica (Powell, 1998) causes the point estimate to increase to 0.12 (95% CI [0.04, 0.19]. Removing the study from Guyana (Ismail *et al.*, 2012) reduces the overall estimate to 0.06 (95% CI [-0.01, 0.13]

The study in the Philippines found that for both the treatments arms of the school feeding programme the effect on English test scores was larger in magnitude than for Filipino test scores. In the case of the pure school feeding arm, the effect size was 0.25 (95% CI [0.15, 0.35]), and for the school feeding treatment plus the parent teacher partnership meetings the effect was 0.34 (95% CI [0.24, 0.43]). Similarly, the programme in Kenya (Hulett *et al.*, 2014) found larger effects on English (SMD=0.37 SMD 96% CI [0.03, 0.72]), than on Kiembu (the language included in the meta-analysis for this study).

Figure 4.2 g: Language Arts Test Scores

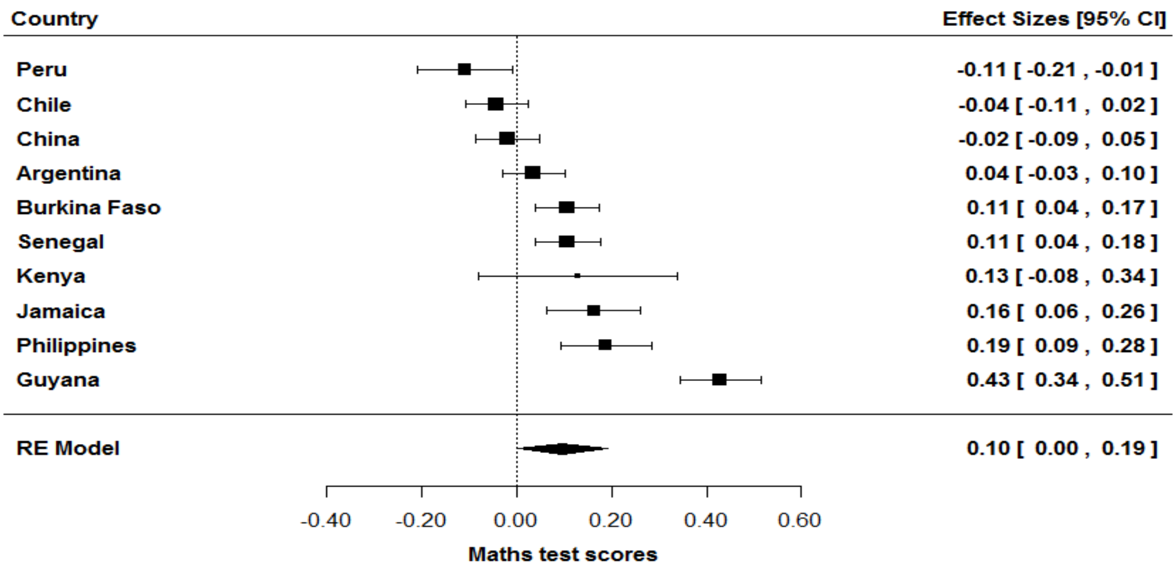


Maths scores

The overall weighted average effect of school feeding on learning as measured by maths test scores was 0.10 (95% CI, 0, 0.19), calculated under a random effects model. This is a similar result to meta-analysis of language arts test scores. The homogeneity tests (I^2 = 92.63%, τ^2 = 0.0210, Q(df = 9) = 110.7828, p-value= < .0001) suggest a large amount of between-studies variability which can also be observed in Figure 4.2h.

The effect sizes range between -0.11 (95% CI, -0.21, -0.01) in Peru (Jacoby *et al.,* 1996) to 0.43 (95% CI [0.34, 0.51]) in Guyana (Ismail *et al.*, 2012). Again, the feeding programme in Guyana report a substantially larger effect on maths test scores than the other studies, and removing this study from the analysis reduces the point estimate to 0.06, 95% CI [-0.01, 0.12].

Figure 4.2 h: Maths test scores

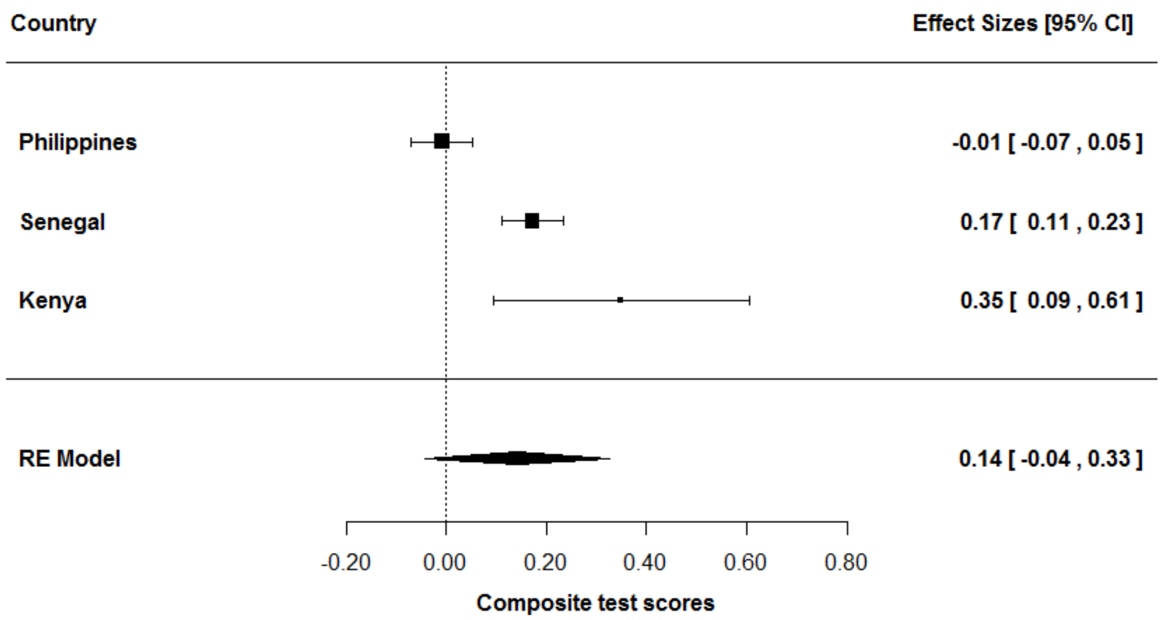


Composite test scores

The overall weighted average effect of school feeding on learning as measured by composite test scores was 0.14 (95% CI, -0.04, 0.33), calculated under a random effects model. We only identified three impact evaluations that reported on a composite test score and the test of homogeneity suggests a large amount of between study heterogeneity (I^2 = 92.15%, τ^2 = 0.0219, Q(df = 2) = 20.228, p-value < .0001).

Figure 4.2i supports the presence of heterogeneity. The effect sizes range from -0.01 95% CI [-0.07, 0.05] for the school feeding programme in the Philippines (Tan, 1999) to 0.35 95% CI [0.09, 0.61] for the feeding programme in Kenya (Hulett *et al.,* 2014). The results are sensitive to the removal of both the study from the Philippines and Senegal, increasing and reducing the magnitude of effect substantially.

Figure 4.2 i: Composite test scores



4.2.4 Summary of findings and discussion

We identified 16 studies that evaluated the effect of a school feeding programme, implemented in Latin America and the Caribbean, Sub- Saharan Africa, East Asia and Pacific and South Asia. The overall average effects range from -0.06 SMD, 95% CI [-0.15, -0.03] for dropout to 0.14 SMD, 95% CI [-0.05, 0.33] for enrolment and 0.14 SMD, 95% CI [-0.04, 0.33] for composite test scores. There is a large amount of variability for most overall estimates and we see the most consistent positive effects across contexts for student attendance.

One of the key assumptions of the theory of change is that the programme responds to community needs, that is, the target group has a clear need for additional food intake (inadequate energy intake or nutritional deficiencies for example) or that that the food ration represents an economic or nutritional benefit to the family. An explanation for some of the observed heterogeneity may therefore be that school feeding programmes are more likely to be effective in contexts with high food insecurity and low existing school participation. Observed effects appear to be slightly larger in contexts with poor baseline school participation and food insecurity. Indeed, many of the included studies came from contexts described as such in the impact evaluations (Tan *et al.*, 1999 – Philippines; Kazianga *et al.*, 2012 – Burkina Faso; Buttenheim *et al.*, 2011 – Laos; Omwami *et al.*, 2011 – Kenya; Jayaraman and Simroth, 2015 – India; Diagne *et al.,* 2014 – Senegal, Cheung and Berlin, 2014 - Cambodia).

For example, the effects of the Hinterland Community-Based SFP in Guyana (Ismail *et al.*, 2012) are consistently positive and relatively large in magnitude for both school participation and learning outcomes. The programme was implemented during the global food price shocks of 2007-08, with a documented increase in food insecurity for poor families. In this context the feeding programme met a clearly defined need and represented an important income transfer to poor families. While school attendance fell among the comparison group, it increased among the children in the programme. Similarly, during the study of the school feeding programme in Kenya the area was hit by drought, which caused food shortages at the household level (Omwami *et al., ibid*). In this case attendance rates declined for students in both the treatment and comparison schools, but the decline was smaller in the treatment schools. On the other hand, in Chile where most extreme childhood malnutrition has been eliminated and primary school enrolment is high, there may have been more limited room for improvements (McEwan, 2013; Altman, 2013). Therefore two of the assumed mechanisms for improved learning in the school feeding programme theory, that is, through changes in nutrition and school participation, appear of limited relevance in this context. Indeed, the results for Chile suggest small and insignificant effects on most outcomes.

Looking at the meta-analysis of maths test scores for instance (the outcome where we have the largest number of studies), we observe consistent positive effects of school feeding in Burkina Faso, Senegal, Kenya, the Philippines and Guyana. The exception to this is the egg-a day trial in China (SMD= -0.02, 95% [-0.09, 0.05]). Many of the school children in Gansu province where the trial took place are described by the study authors as being malnourished, suffering from iron and micronutrient deficiencies. The study authors suggest the food provided was unlikely to meet the nutritional needs of the children in this context (Kleiman-Weiner *et al.*, 2013). The provision of an egg a day is more limited than the food provided in most of the other programmes and may explain the lack of an effect in this context.

Some argue for the need for feeding programmes to adapt an integrated approach that combines health, nutritional and educational components to increase programme effectiveness (Powell, 1998; Jomaa *et al.,* 2011). The evidence from the included studies adopting such an approach is mixed. The effect of the programme in Cambodia that combined school feeding with deworming, complementary health and sanitation activities appears to have been larger in magnitude after the first year of treatment than the standard feeding programme (Cheung and Berlin, 2014, SMD= 0.19, 95% [0.13, 0.25] compared to SMD = 0.13, 95% [0.05, 0.22] for the standard feeding programme). However, the standard school feeding programme in Sri Lanka also adopted this approach there was no effect on enrolment (the only outcome measured in this study, He, 2010a), although this was in a context of already high enrolment levels.

We identified some limited evidence to suggest a lack of local capacity and infrastructure may be a barrier to the effective management of school feeding programmes. For instance, in the case of the Mid-Day Meal Scheme in India poor basic infrastructure and logistical arrangements for preparing meals were reported to obstruct teacher activity in the classroom (Dreze and Goyal, 2003; PEO, 2010; CUTS International, 2006). On the other hand, there is some suggestion that greater local ownership over the running of the feeding programme may facilitate programme implementation and better education outcomes. In the case of the Hinterland Community-Based SFP in Guyana (Ismail *et al*., 2012) implementation of the feeding programme was left entirely up to local communities, providing funding and training for community members to deliver their own programme. The welfare programme in Sri Lanka (He, 2010b) that left the decisions and implementation of the school feeding programme up to the local community also saw larger effects on enrolment than the programme in the country that was centrally implemented by WFP (SMD=0.10, 95% CI [0.06, 0.15], compared to SMD= -0.1, 95% CI [-0.05, 0.04]).

Overall the available evidence suggests school feeding programmes may improve school participation and learning outcome in some contexts. The large effects reported in some contexts suggest school feeding has the potential to improve primary age children's school attendance, the outcome where we saw the most consistent positive effect across different contexts, as well as learning.

Table 4.2 c: Descriptive Findings: Process and Implementation

Descriptive findings: Process and implementation	**Context**	**Citation/ info type**
Uptake of the interventions		
Almost complete child compliance with the snack programme	China	Kleiman-Weiner al., 2013: *Impact evaluation*
Schools unable to set up school feeding programme in time	Senegal	Diagne *et al.,* 2014: *Impact evaluation*
Lack of take-up due to high cost of the feeding programme for target villages	Laos	Buttenheim *et al.,* 2013: *Impact evaluation*
Implementation fidelity and service quality delivery		
Poor basic infrastructure a barrier to successful implementation of feeding programmes	India, Sri Lanka	Dreze and GoyaL, 2003, PEO, 2010, CUTS International, 2006: *Mixed methods*. He, 2010: *Impact evaluation*
Schools meals not provided on a daily basis as intended	Laos	Buttenheim *et al.,* 2013 *Impact evaluation*
Poor basic infrastructure and logistical arrangements for preparing meals obstructing teacher activities in India	India	Dreze and Goyal, 2003, PEO, 2010: *mixed methods study.* Khera, 2006
Food provided to non-enrolled students in Laos	Laos	Buttenheim *et al*., 2013: *Impact evaluation*
School meal schemes are well-liked by children and families	Chile, India	Altman, 2013: *Project document* CUTS International, 2006; PEO, 2010: *Mixed methods*
Programmes targeting school meals at the neediest schools	Chile, Argentina	McEwan, 2013; Adrogue *et al*., 2011 *Impact evaluations*

Table 4.2 d: Descriptive Findings: Contextual Barriers and Facilitators

Descriptive findings: Contextual barriers and facilitators	**Context**	**Citation/ info type**
Existing education and nutritional status/food security	Chile, Kenya	McEwan, 2013; Omwami *et al.,* 2011; He, 2010: *Impact evaluations* Altman, 2013: *Project document*
Presence of school user fees	Kenya	Omwami *et al.*, 2011 *Impact evaluation*
Capacity of local education organisations or the community to manage the feeding programme	Laos, Senegal, India	Buttenheim *et al.*, 2011; Diagne *et al.,* 2014: *Impact evaluations* CUTS International *(2006): Mixed methods*

4.3 Merit-based scholarships

Merit-based scholarships aim to improve learning outcomes by rewarding high performing students with scholarships or one-off cash payments to continue their study (McEwan, 2013; Berry, 2013). For example, an intervention in Kenya provided scholarships to girls who performed well in their 6th grade exams, awarding the top 15 per cent of students in the grade a grant to cover school fees for two years, and also a cash sum to be used for school supplies (Kremer *et al.*, 2009). In another study, conducted in Gurgaon, India, children were told they would receive a one-off cash reward or toy if they achieved a certain pre-determined grade in their reading tests, thereby presenting the child with an additional incentive to perform well in their examinations (Berry, 2013).

4.3.1 How may merit-based scholarships effect education outcomes?

Most incentive programmes focus on rewarding merit as this is thought to generate the highest return for scholarship investment - it is assumed that such scholarships will go to children that can make most use of them as merit suggests proven persistence in education (Chapman and Mushlin, 2008). Merit-based scholarships are thought to work as a price subsidy for academic performance (Sharma, 2011), facilitating increased student involvement through improved attendance and greater effort in school and homework. Incentives are thought to be more effective as they target the child directly, instead of parents or caregivers (Berry 2014: 2). In many cases there is also a considerable opportunity-cost of schooling and so children can be incentivised to stay in school where they may otherwise dropout in order to pursue low-paid work (Liu, 2013; Yi *et al.,* 2015). Financial incentives are also thought to increase parental involvement in education by increasing monitoring of schools and teachers (Kremer *et al*., 2009. When scholarships are provided, continuous testing is considered to incentivise learning in order to maintain such scholarships (Barrera-Osorio and Filmer, 2013). Scholarships based on academic performance at graduation from an education institution (primary or secondary) are provided to facilitate the next stage of education (secondary or tertiary).

4.3.2 Description of included studies

We included eleven studies reported in ten different papers that evaluated the effects of a merit-based scholarship or incentive intervention. These referred to nine unique programmes. We used the term 'study' to refer to a unique output from an author team, which in some cases was reported in several papers or, as was the case with Li *et al.,* (2014), two different experiments reported in the same paper. Li *et al*., 2014 reports results for two different interventions, a peer-incentive experiment and an individual incentive. Similarly, Barrera-Osorio and Filmer (2012) report on two different treatment arms of a scholarship programme in Cambodia. However, in this case the other treatment arm does not fit into the merit-based scholarships category and so is reported in the cash transfers section of the review. Finally, two studies evaluate the impact of a scholarship offer through the Early Commitment of Financial Aid programme, one to grade seven students (Yi *et al*., 2015, 2012) and the other to grade nine students (Yi *et al*., 2015, Liu *et al*., 2013). We have reported results for these separately. In the following section, we described the characteristics of these studies in detail (as also summarised in Table 4a).

Population

Five of the studies identified in this category took place in primary schools while the remaining studies took place in secondary schools. Of the eleven studies, two did not report any information on the school grade of participants, though one of these noted that the average age of students was 16 years (Blimpo *et al.,* 2010). In the Chinese Fall Challenge Programme (for both the peer-incentive and individual-incentive experiments) the authors reported that students were grades 3 through to 6 (Li *et al.,* 2014). Sharma *et al*. (2012) reports that participants in the Nepal programme were from grade 8 and the study by Kremer's *et al.* (2008) on the Kenya Girls Scholarship Programme included students that were grade 6 (roughly aged 13 or 14 years, depending on whether or not they had repeated a grade). In the programme in Cambodia, participants were all in grade 4 (Barrera-Osorio and Filmer, 2012). Both Early commitment of Financial Aid (ECFA) studies used participants from grade 7 (Liu *et al.,* 2013; Yi *et al*., 2015), as did another Chinese study; the Ningshan Tuition relief programme (Chen, 2013). The Gurgaon programme had participants from grades 1 through to 3 (Berry, 2013).

Setting

The included studies covered programmes in a broad range of settings, with six studies from East Asia and Pacific, two studies from South Asia, two studies from Sub-Saharan Africa, and one study from Latin America and the Caribbean. Four of these studies took place in China (Chen *et al*., 2013; Liu *et al*., 2013; two reported in Li *et al.,* 2015), and one each respectively from Benin (Blimpo et al., 2010), India (Berry, 2013), Kenya (Kremer *et al*., 2008), Mexico (Behrman *et al*., 2012), Nepal (Sharma *et al.,* 2012) and Cambodia (Barrera-Osorio and Filmer, 2012). Three of the studies took place primarily in urban and peri-urban areas, one was conducted in a peri-urban area only, while the final five were conducted in rural areas. The final study did not make clear which areas the programme was run in (Barrera-Osorio and Filmer, 2012).

Intervention

Table 8b provides an overview of the different components of the programmes in the included studies. All of the evaluated programmes featured some kind of incentive (usually cash) based on performance. However, the structure of the incentive programmes varied in several ways. For most programmes, rewards were cash grants paid to the individual. One exception to this – the Kenya Girls Scholarship Program (GSP) – offered a scholarship to be paid directly to the child's school (Kremer *et al.*, 2008). The programme also offered a materials grant to children, which was intended for buying school supplies such as stationery and books. This grant was paid directly to the household and was unconditional (*ibid*).

The use of 'pay-for-grades contracts' was common in these programmes. These so-called contracts could be either tournament-style contracts or target-style contracts. Tournament contracts can be described as more competitive since there tend to be a limited number of awards available for top-scoring students. As such, students needed to obtain the best grades in their class or year group in order to receive the incentive. For example, the Fall Challenge Programme in China only awarded the top three scoring students in a class (Li *et al.*, 2014), while the GSP students had to score in the top 15 per cent of their year group (Kremer *et al.,* 2009) in order to be entered into the scholarship programme. Target-style contracts offered rewards to all students who achieve a certain pre-determined grade, either in one particular subject or an average score across all subjects (Behrman *et al.*, 2012; Blimpo *et al.*, 2010; Sharma, 2011). The programme in Cambodia offered scholarships to participants who passed with the highest grades in their national examinations in the previous year (Barrera-Osorio and Filmer, 2012).

Three included studies had slightly different objectives from the others in that they were designed explicitly to incentivise and assist students in attending the next level of education. These programmes were conducted in China, were very similar in design and differentiate only slightly in application. Two of these studies evaluate a scholarship offer through the Early Commitment of Financial Aid programme, one to grade seven students (Yi *et al.*, 2015, 2012) and the other to grade nine students (Yi *et al.*, 2015, Liu *et al.*, 2013). The ECFA was provided as a contract to junior high school students to contribute to the costs of senior high school conditional on the student's matriculation from junior high school (Liu *et al.,* 2015). In the programme implemented for ninth graders (Yi *et al.*, 2015, Liu *et al.,* 2013), the contract was provided half way through the ninth grade, the last year of junior high school, offering a scholarship 1,500 yuan a year for three years of senior high school. Students and parents were invited to the principal's office where targeted students signed the contract and a photograph was taken. This photograph was sent as a reminder of the agreement to the household and another reminder was sent before senior high school examinations (Liu *et al.,* 2013).

The Ningshan County's Tuition relief Programme (Chen *et al*., 2013) offered the promise of the same amount of tuition relief, but rather than being offered in secrecy to selected ninth or seventh grade students, all students were made aware of the programme. The regional government conducted promotional activities. The scholarship offers were given to the 500 best-performing students in high-school examinations. The study evaluates the effects of the scholarship promise on learning outcomes for seventh grade students (i.e., the first year of junior high). The Early Commitment of Financial Aid evaluated by Yi and colleagues (2015, 2012) was implemented in identical fashion to the ECFA for ninth grade students, with the exception that it was provided to seventh grade students. Instead of sending the reminder before the secondary entrance exam it was sent before the evaluation survey (Yi *et al.*, 2012).

All of the included programmes paid rewards to individuals, though performance may not have been measured on an individual basis. Some programmes had treatment arms which offered rewards to students on the basis of the overall performance of their class or group or where parents were offered the incentive instead and were able to choose whether to give the reward to the child or keep it for the household (Berry, 2013). In Benin, Blimpo *et al.*

(2010) evaluated the effects of incentives not only on individuals but on groups as well. In this programme there were two group treatment arms, one on a tournament contract and the other with a target contract. Both team target and tournament target groups were offered incentives based on the average score of students in their group (Blimpo *et al*., 2010).

The Fall Challenge Programme followed a different design by offering high-achieving peers an incentive to support their lower achieving bench-mate. Incentives were paid to both the high-achieving and low-achieving students if the low-achieving students' exam scores were the most improved in the treatment group over the course of the experiment (Li *et al.,* 2014).

Comparisons

Ten of the included studies compared the effect of an intervention to business as usual. Several of the included studies also had multiple treatment arms, allowing for an assessment of comparative effectiveness of different programmes, although we only included one treatment-comparison per study in our analyses. Berry *et al.* (2013) was the only study which did not include a business as usual comparison group, instead comparing the effect of two different incentive programmes.

Outcomes

All eleven studies reported on some measure of achievement or learning. For the most part, achievement was measured through test scores: maths (n=7) (Barrera-Osorio and Filmer, 2012; Berhman *et al*., 2012; Li *et al*., 2014; Sharma 2011, Chen *et al*., 2013; Yi *et al*., 2013) language arts (n=4) (Li *et al*., 2014; Sharma 2011; Berry *et al*., 2013), and composite scores (n=4) (Blimpo 2010, Kremer *et al*., 2008, Liu *et al*., 2013, Sharma *et al*., 2011). Two studies presented composite score results by gender (Li *et al*., 2014). Barrera-Osorio and Filmer (2008) also reported on cognitive scores.

Three studies report on student attendance (Kremer *et al*. 2008; Liu *et al.* 2013; Barrera-Osorio and Filmer, 2012), and one study also reported on teacher attendance (Kremer *et al.* 2008). Two studies reported on student completion (Liu *et al*. 2013; Barrera-Osorio and Filmer, 2012) and two studies assessed effects on dropout (Liu *et al.* 2013; Yi *et al*. 2015). Most of the studies followed up within 12 months of the start of an intervention while two followed up after 24 months (Barrera-Osorio and Filmer, 2012; Berhman *et al*., 2012). Berhman *et al.* (2012) also included a 36 month follow-up.

Study Design

Seven studies were cluster randomised trials with treatment and control groups being assigned at the school level. A further two studies were RCTs and one study used a controlled before and after design with propensity score matching and difference-in-difference analysis.

Qualitative studies, process evaluations and project documents

We identified two additional documents that present qualitative, process and project information for one of our included programmes (ALI) (Behrman *et al.,* 2012). We found a lack of additional documents for all other interventions. Several of the impact evaluations also included qualitative components and therefore a main part of our qualitative synthesis is based on that.

Table 4.3 a: Merit Based Scholarships

Included study	Setting	Intervention summary	Included outcomes	Follow- Up	Study design	Sample Size
Li *et al.*, 2014	China (urban, peri-urban). Primary school. Grade: 3 – 6 Age: Not reported.	Individual incentive: Students were offered a pay-for-grades incentive contract. There was a public ceremony and official certificates for the winners.	Maths test scores; language arts test scores	6 months	Cluster- RCT	994 students
Li *et al.*, 2014	China (urban, peri-urban), Primary school. Grade: 3 – 6 Age: Not reported.	Peer incentive: Students were offered a pay-for-grades incentive contract just as in the individual incentives contract. In addition in the peer-incentive class each of the top ten students in the class were assigned to serve as a benchmate for one of the treated students. To encourage peer interactions the three students in each class with the greatest test score gain and their benchmates were offered an equivalent cash prize.	Maths test scores; language arts test scores	6 months	Cluster- RCT	956 students

Included study	Setting	Intervention summary	Included outcomes	Follow- Up	Study design	Sample Size
Sharma *et al.*, 2011	Nepal (urban/rural setting not reported). Secondary school. Grade: 8 Age: 14.49	The incentive programme provided cash rewards based on students' average aggregate scores in each of the two semester exams and the end-of-the-year district level exam. To encourage students to pass, those who passed all subjects received reward at twice the rate per score compared to those who failed one or more subjects. Furthermore, those who failed one or more subjects in the earlier semesters could earn back the "withheld" amount if they passed all subjects in the later semesters.	Maths test scores; language arts test scores; composite test scores	6 months	Cluster -RCT	4042 students
Kremer *et al.*, 2008	Kenya (rural). Primary schools. Grade: 6 Age: 13.5 (boys) and 13.9 (girls) in Busia. 14.0 (boys) and 14.1 (girls) in Teso.	Merit based scholarship foracademic years of US$6.40 (KSh 500) to cover school fees (paid to school), a grant of US412.80 (KSh1000) for school supplies paid to family and a public recognition awards ceremony.	Attendance; composite test scores	12 months	Cluster-RCT	3343 students

Included study	Setting	Intervention summary	Included outcomes	Follow- Up	Study design	Sample Size
Behrman *et al.*, 2012	Mexico (rural). Secondary schools. Grade: 10 – 12. Age: not reported	The ALI experiment consisted of different types of performance incentives, of which this review discusses the student incentives. The programme incentivizes students to improve their maths grades by giving them financial rewards when they progress from a certain to another grade (bonuses are based on individual performance).	Maths test scores	24 months	Cluster- RCT	2829 students
Blimpo *et al.*, 2010	Benin (nationwide). Secondary school. Grade: not reported Age: 16.22	The intervention consists of three treatment arms: an individual bonus, a team bonus and a tournament bonus: (1) In the ‘’individual target’’ group each participant received a promise of 5000 Francs CFA to be paid if she passed her secondary school certification examination (BEPC); (2) In the ‘’Team Target’’ Group, students were randomly assigned to teams of four. Each team received a promise of 20,000 Francs CFA to be paid to the team if its average score equals or exceeds the passing grade on the BEPC; (3) In the ‘’Team Tournament’’ group, students are randomly assigned teams of four and the three teams with the highest average scores each won a prize of 320,000 Francs CFA.	Composite test scores	12 months	Cluster-RCT	1274 students

Included study	Setting	Intervention summary	Included outcomes	Follow- Up	Study design	Sample Size
Barrera-Osorio and Filmer, 2012	Cambodia (setting not reported). Primary school. Grade: not reported Age: not reported.	Scholarship recipients were applicants who scored well on an assessment test. Once participating schools were selected, scholarship recipients were identified according to their assessment scores. Recipients then had to stay enrolled, attend school, regularly, and maintain passing grades in order to keep the scholarship until they graduate from high school. The scholarship amount was set at US$20 per student, per year.	Maths test scores; cognitive scores	24 months	Cluster-RCT	940 students
Berry 2013	India (urban). Primary school. Grade: 1 - 3 Age: not reported.	Children were assigned a goal determined by a pre-intervention reading test. Two months later they were retested and rewarded if they achieved their personal goal. There were four treatments arms; the first treatment provided a cash prize of 100 rupees to parents, the second treatment arm paid the incentive directly to the children, while the third allowed the child to choose a toy (which had a value of 100 rupees). Children in this treatment were either allowed to choose a toy from a selection presented or given a voucher to redeem in their local toy shop. All participants also received additional after-school classes as well as reminders about their upcoming reading test.	Language arts test scores	2 months	RCT	993 students

Included study	Setting	Intervention summary	Included outcomes	Follow- Up	Study design	Sample Size
Chen *et al.*, 2013	China (rural). Public secondary school. Grade: 7 Age: 12.92 (t), 13.06 (c).	The Ningshan tuition relief programme paid annual tuition (1500 yuan) for 3 years of senior high school for those among the top 500 students in the entrance examination to senior high school.	Maths test scores	11 months	CBA	3121 students
Liu *et al.*, 2013	China (rural). Secondary school. Grade: 9 Age: 15.3	The Early Commitment of Financial Aid programme (ECFA) was provided as a contract to students promising to contribute to the costs of senior high school conditional on the student's matriculation from junior high school. The contract was provided half way through ninth grade, the last year of junior high school, offering 1,500 yuan a year for three years of senior high school. Students and parents were invited to the principal's office where targeted students signed the contract and a photograph was taken. This photograph was sent as a reminder of the agreement to the household and another reminder was sent before senior high school entrance examinations (Liu *et al.,* 2013: pp. 13-4).	Dropout; completion; composite test scores	6 months	RCT	532 students

Included study	Setting	Intervention summary	Included outcomes	Follow- Up	Study design	Sample Size
Yi *et al*., 2012, 2015	China (rural). Secondary school. Grade: 7 Age: 13.5 (pair treatment), 13.6 (pair control), 13.5 (pure control).	The Early Commitment of Financial Aid programme (ECFA) was provided as a contract to students promising to contribute to the costs of senior high school on matriculation from junior high school (Liu *et al.,* 2013). The contract was provided to seventh grade students, offering 1,500 yuan a year for three years of senior high school. Students and parents were invited to the principal's office where targeted students signed the contract and a photograph was taken. This photograph was sent as a reminder of the agreement to the household and another reminder was sent before the end line data collection.	Dropout; maths scores; attendance	6 months	RCT	1892 students

Table 4.3 b: Intervention Design Features of included Merit-based Scholarship studies

Intervention design features	Li *et al.*, 2014a	Li *et al.*, 2014b	Sharma *et al.*, 2011	Kremer *et al.*, 2008	Berhman *et al.*, 2012	Blimpo *et al.*, 2010	Barrera-Osorio and Filmer, 2012	Berry, 2013	Chen *et al.*, 2013	Liu *et al.*, 2013	Yi *et al.*, 2015
Individual Incentive		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Peer Incentive	✓										
Group Incentive					✓	✓					
Target- or threshold- incentive contract			✓		✓	✓	✓	✓		✓	
Tournament-style incentive contract	✓	✓		✓	✓	✓			✓		
Add payment for exam results					✓						
Public recognition		✓		✓							
Intervention delivery mechanisms											
Cash reward	✓	✓	✓		✓	✓		✓	✓	✓	✓
Scholarship				✓			✓				
Material Grant				✓							
Toy reward								✓			

4.3.3 Synthesis of findings

The results of our synthesis are presented in two sections. We first present the findings of the meta-analysis and individual effects of merit-based scholarships on primary and secondary outcomes, followed by a summary of the findings at the end of the section.

Effects of Merit-based scholarships interventions on attendance and learning outcomes

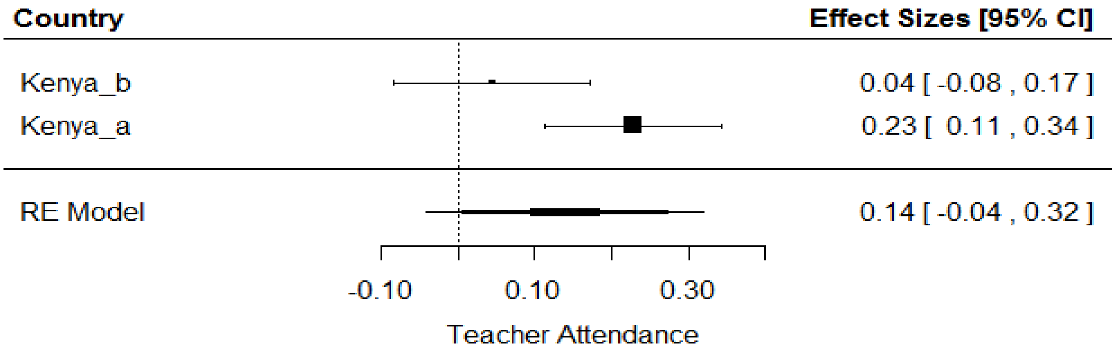
This section reports the results of the meta-analysis of the effects of merit-based scholarships, addressing question 1a. We have structured the presentation of results according to the causal chain, starting with teacher attendance, followed by student attendance, dropout, completion, cognitive scores and finally the learning outcomes (maths, language arts and composite test scores). The studies include a range of different follow up periods, with the majority of studies including a data point within the 12-month period (with the exception of Barrera-Osoria and Filmer, 2012 and Berhman *et al.*, 2012 which also followed up at 24 months). Therefore we selected the 12-month follow up period for the meta-analysis when available.

Ten studies provided data for meta-analysis, but none of the studies reported on all outcomes. The number of comparisons with effect sizes ranges from four for attendance to ten for maths outcomes. We did not include Berry *et al.* (2013) in the meta-analysis as this study did not include a business as usual comparison. As previously described, several studies reported results for multiple treatment arms testing different types of merit-based scholarships. In all of these cases we included treatment arms evaluating individual incentives to children in the analysis (as these were most comparable), but present results for the other treatment arms either in Appendix H or elsewhere in this review. For example, Behrman *et al.* (2012) report on the Aligning Learning Incentives (ALI) experiment which had three treatment arms: incentives for children, incentives for teachers, and incentives for children, teachers and administrators. We have included the incentives for children arm in the meta-analysis for maths test scores but report results for the other arms in chapter six.

In the case of studies reporting multiple language arts outcomes, for example in Sharma *et al.,* (2011) which reported results for both Nepali and English test scores, we chose the local language or language spoken by the most number of children in the sample to include in the meta-analysis (in this case Nepali) but also present effect sizes for the other outcomes. All effect sizes are expressed as standardised mean difference (SMD), interpreted as the magnitude of the number of standard deviation changes in the outcome for the intervention group as compared to students in the comparison group schools. SMD scores are interpreted as the number of standard deviation changes in the outcome. All meta-analyses calculated using a random effects model.

Teacher Attendance
The overall average effect of merit based scholarships on teacher attendance is 0.14, 95% CI [-0.04, 0.32], calculated under a random effects model. The homogeneity test (I^2 = 77%, τ^2 = 0.013, Q(4 df) = 4.1, p = 0.036) indicates that the effects did not arise from the same population. The forest plot in Figure 4.3a supports the presence of heterogeneity.

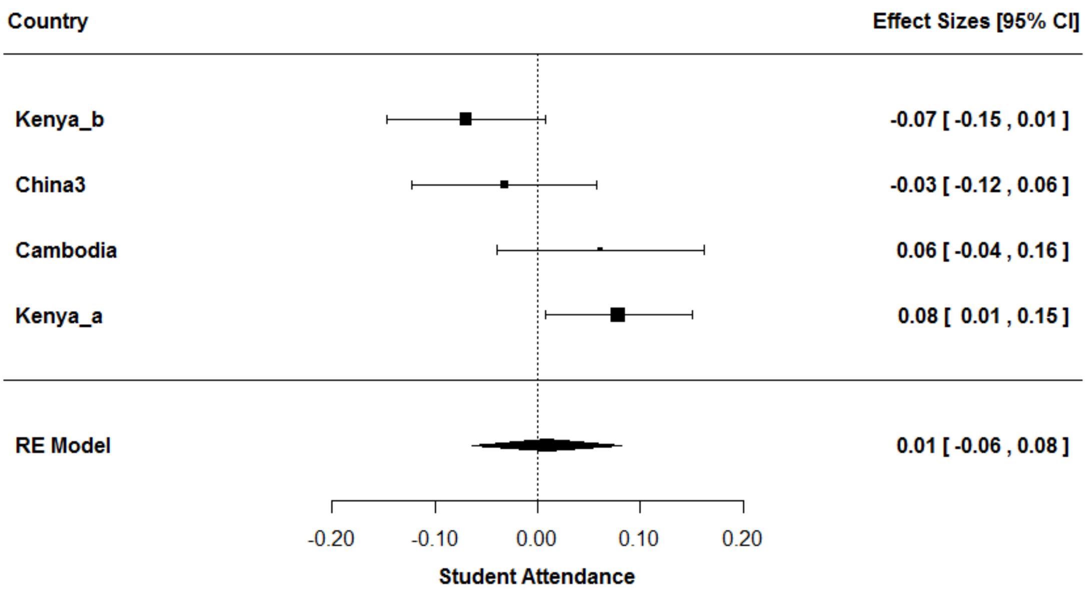
Figure 4.3 a: Teacher Attendance



Student attendance

The overall average effect of merit based scholarships on student attendance is almost zero at 0.01 SMD, 95% CI [-0.06, 0.08[. The results also suggest a moderate amount of heterogeneity (I^2 = 67.02%, τ^2 = 0.0037, Q(df = 3) = 9.5368, p-val = 0.0229). As can be seen from Figure 4.3b, most of the confidence intervals are overlapping. The results are sensitive to the removal of both Kenya_a and Cambodia, which reduces the effect to -0.02 SMD (95% CI [-0.09, 0.05]) and -0.01 SMD (95% CI [-0.10, 0.08]) respectively. Please see Appendix H for details.

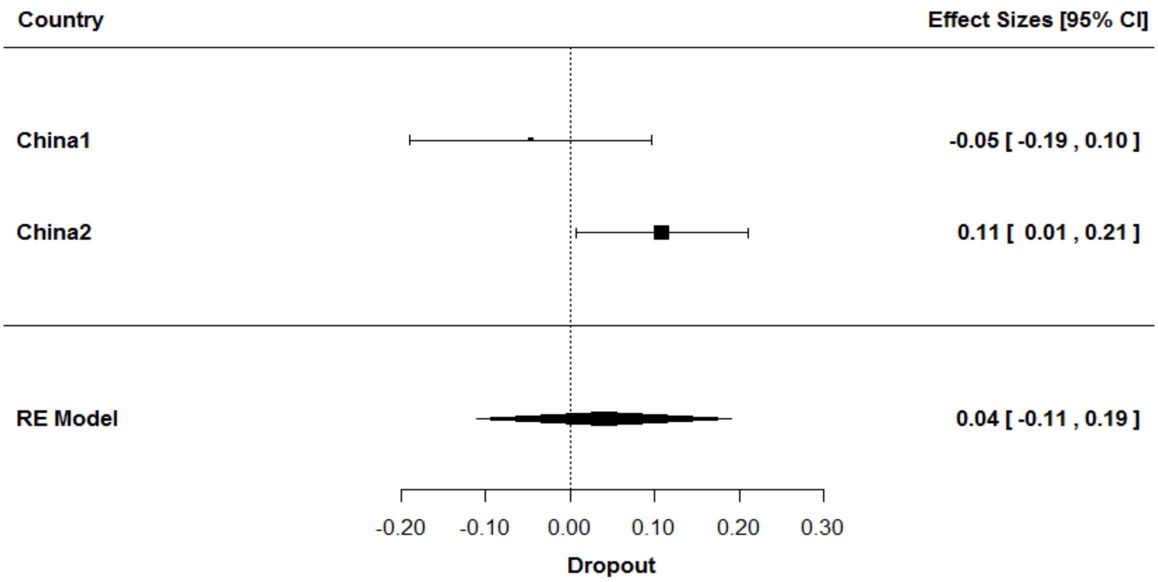
Figure 4.3 b: Student Attendance[8]



Dropout

The overall average effect of merit based scholarships on maths test scores is 0.04 SMD, 95% CI [-0.11, 0.19]. The assessment of homogeneity suggests a moderate amount of between-studies variability (I^2 = 66.84%, τ^2 = 0.0080, Q(df = 1) = 3.0155, p-val = 0.0825). This is also apparent when inspecting the forest plot in Figure 4.3c, where the confidence intervals of both comparisons are overlapping.

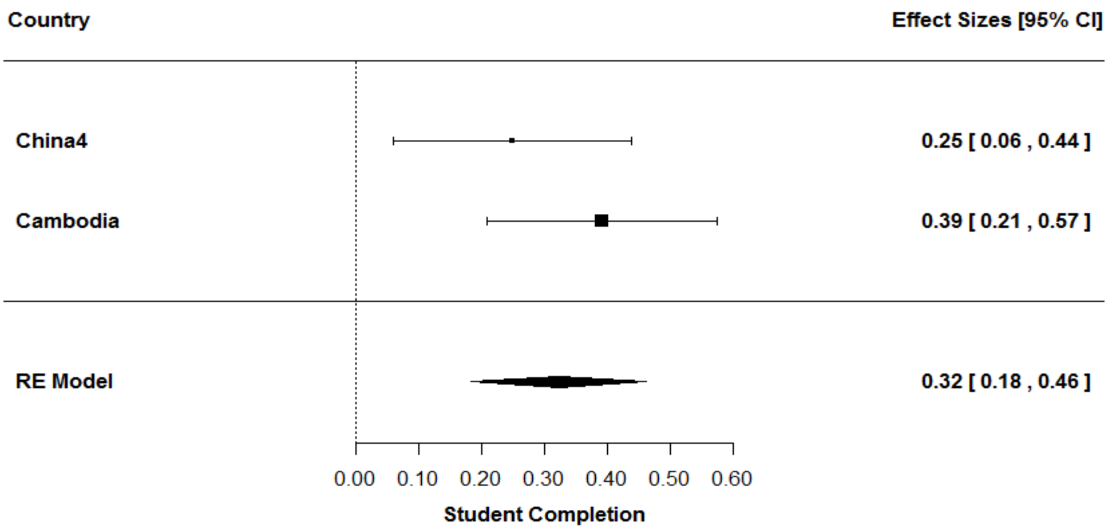
Figure 4.3 c: Dropout



Completion

We only identified two merit-based scholarship studies that reported effects on a measure of completion. The overall average effect of on student completion is 0.32 SMD, 95% CI [-0.18, 0.46]. The homogeneity test (I^2 = 10.87%, τ^2 = 0.0011, Q(df = 1) = 1.1219, p-val = 0.2895) indicates a low level of heterogeneity. The forest plot in Figure 4.3d supports this result.

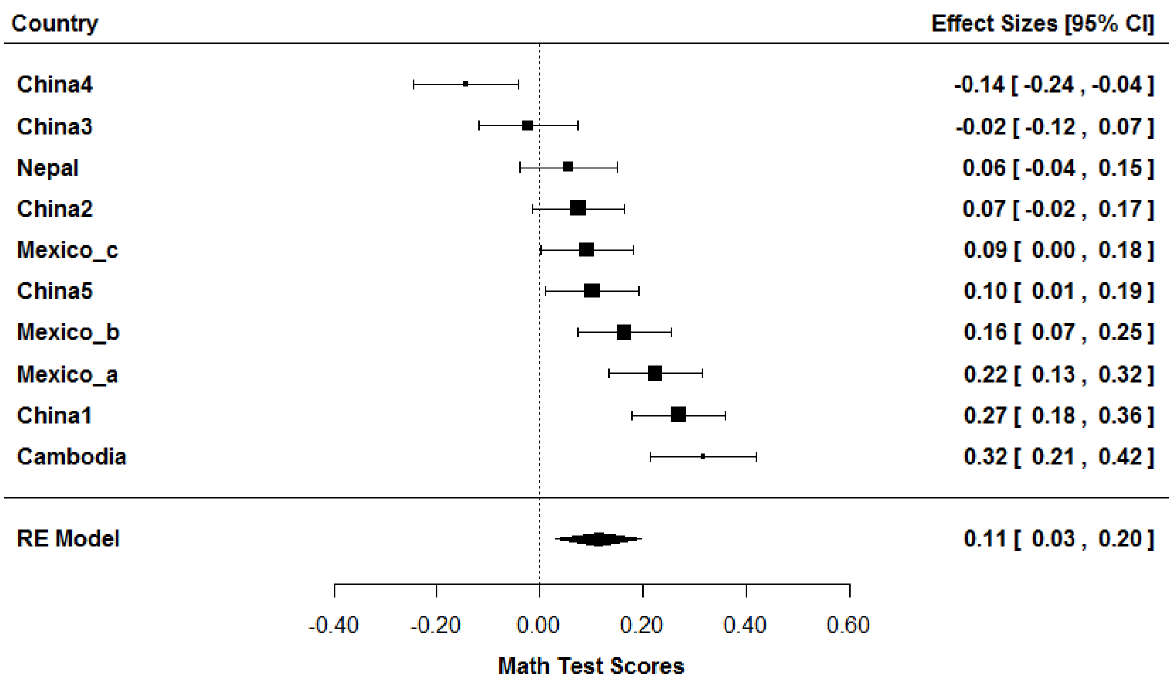
Figure 4.3 d: Completion[8]



Maths Test Scores

The overall average effect of merit based scholarships on maths test scores is 0.11 SMD, 95% CI [0.03, 0.20]. The assessment of homogeneity suggests a high degree of heterogeneity between studies (I^2 = 87.66%, τ^2 = 0.0162, Q(df = 9) = 67.9728, p-val = < .0001). Figure 4.3e presents the forest plot with the results of the individual studies and the pooled point estimate. The effects range from -0.14 (95% CI, [-0.24, -0.04 in the Early Commitment of Financial Aid programme (ECFA) in China, to 0.32 (95% CI, [0.21, 0.32]) for the merit-based scholarship programme in Cambodia. Results appeared to be sensitive to the removal of China4, giving an effect size of 0.14 SMD (95% CI [0.07, 0.21]).

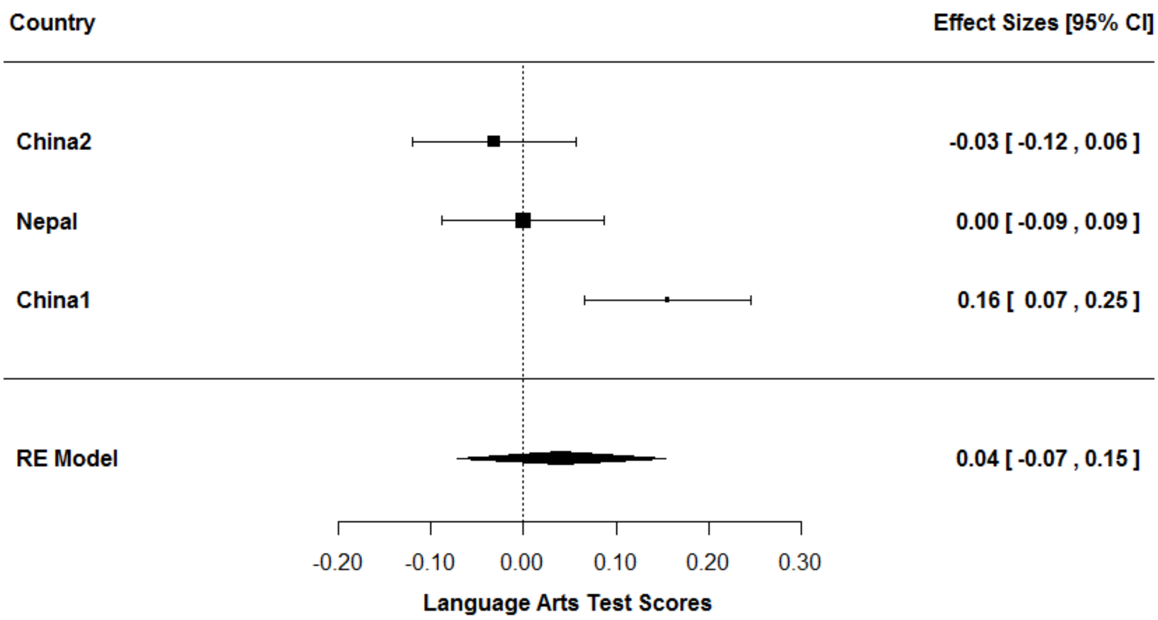
Figure 4.3 e: Maths test scores[8]



Language Arts Test Scores

The overall average effect of merit based scholarships on language arts scores is 0.04 SMD (95% CI [-0.07, 0.15]). The assessment of homogeneity suggests a large amount of between-studies variability ($I^2 = 79.58\%$, $\tau^2 = 0.0079$, $Q(2\ df) = 9.68$, $p = 0.008$). This is also apparent when inspecting the forest plot in Figure 4.3c. We found that these results are sensitive to the removal of the individual incentive experiment arm from Li *et al.* (2014) in China, reducing the effect to -0.02 SMD (95% CI [-0.08, 0.05]) though the results remain statistically insignificant (Appendix H contains further details on sensitivity analysis).

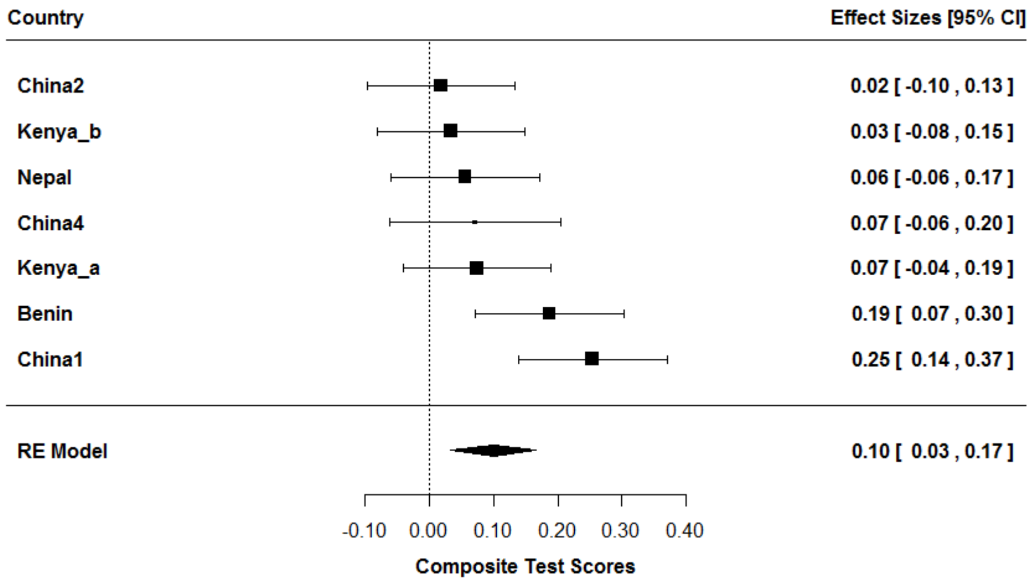
Figure 4.3 f: Language Arts Test Scores[8]



Composite test scores

The overall average effect of merit based scholarships on learning outcomes as measured by composite scores is 0.10 SMD, 95% CI [0.03, 0.17]). The assessment of homogeneity suggests a moderate amount of between-studies variation (I^2 = 54.61%, τ^2 = 0.0043, Q(6 df) = 13.2394, p = 0.0394). As can be seen from the forest plot in Figure 4.3g, most of the confidence intervals are overlapping. Results for composite test scores appear to be sensitive to the removal of two studies. Without China1, the effect reduces to 0.07 SMD (95% CI[0.02, 0.12]), while the removal of the Benin study reduces the effect to 0.08 SMD (95% CI[0.01, 0.15]).

Figure 4.3 g: Composite Scores[8]



Cognitive scores

Only one study reported on cognitive scores. The Pilot Primary School Scholarship Programme in Cambodia use the digit span test and ravens test as measures of cognitive ability. The effect on cognitive test scores as measured by the digit span test were 0.32 (95% CI [0.14, 0.50]) and 0.33 (95% CI, [0.14, 0.51]), as measured by the ravens test.

4.3.4 Summary of findings and discussion

We identified eleven studies assessing the effects of merit-based scholarship interventions on education outcomes. The results suggest no effect of these interventions on attendance (SMD= 0.01, 95% CI [-0.06, 0.08]), although this is based on only a few studies. However, the results for learning outcomes are more promising, with overall positive effects for maths (SMD=0.11, (95% CI [0.03, 0.20]), language arts (SMD=0.04, 95% CI [-0.07, 0.15]), composite scores (SMD=0.10 SMD, 95% CI [0.03, 0.17]) and cognitive (no meta-analysis) outcomes. There is also some evidence to suggest improvements in measures of completion (SMD=0.32 SMD, 95% CI [-0.18, 0.46) and teacher attendance (SMD=0.14 SMD, 95% CI [-0.04, 0.32]).

There is a large amount of heterogeneity for several of the average estimates, suggesting that in some contexts merit-based scholarships led to improvements that were considerably larger in magnitude than what is suggested by the average effect. In particular, for the peer

[8] China_4 (Liu *et al.,* 2013), China2 (Li *et al.,* 2014, individual incentive), China_1 (Li *et al.,* 2014, peer incentive), Kenya_b (Kremer *et al.,* 2008, Teso district), Kenya_a (Kremer *et al.,* Busia district).

incentive programme in China we observe consistently larger effects than the average for all learning outcomes. Similarly, we also observe effects that are relatively large in magnitude for completion and maths for the merit-based scholarship programme in Cambodia. Unfortunately, the evidence does not allow us to identify the reasons for this heterogeneity.

While each meta-analysis is based on a relatively small number of studies, overall the evidence suggests children receiving merit-based scholarships benefit from an improvement in test scores on average, with potential for improvements that are relatively large in magnitude, as was observed in the peer incentives programme in China and the merit-based scholarship programme in Cambodia in particular. More research is needed to identify the programme components that may produce such large effects and whether these effects can be replicated in different contexts.

4.4. Providing information

Providing information to children about the potential future benefits of education in terms of income, employment, and social status is thought to increase school participation, enrolment and continuation where students under-estimate the actual returns to education (Nguyen, 2008). Interventions of this type will typically involve providing information to the students, and in some cases, their parents, about the future potential returns to schooling. The information can be presented in various ways including teachers or external presenters disseminating statistics about average earnings for each level of education. Other interventions make use of role models, who share their experience of education and current achievements with children, with some programmes using a combination of channels (Nguyen, 2008).

4.4.1 Description of included studies

We included four studies that evaluated the effects of providing information to children and/ or their parents. Each study referred to a unique programme. In our conceptual framework, we separated the programmes providing information to children from those providing information to parents. However, as we only identified four studies, we have presented the findings in one section.

Population

Two of the included studies took place in secondary schools (Jensen, n.d. and Loyalka *et al.,* 2013), and both of these assessed grade 8 students (Loyalka *et al.*, 2013) with the other two taking place in primary schools (Dinkleman and Martinez, 2011; Nguyen, 2008), evaluating the effect of the intervention on grade 7 and grade 4 students respectively. One programme took place in rural public schools (Nguyen, 2008), while another of the programmes took place in both private and public schools (Dinkleman and Martinez, 2011). The others did not report any information on school type.

Setting

The included studies covered programmes in East Asia, Latin America and the Caribbean and Sub-Saharan Africa. Two programmes took place in rural areas with one in China (Loyalka *et al.*, 2013) and another in Madagascar (Nguyen, 2008). The others were in urban Chile (Dinkleman and Martinez, 2011) and the Dominican Republic (Jensen, n.d.) respectively.

Intervention

All four programmes shared one key component, namely that participants were informed via written or verbal statements of the returns to education (potential earnings) after leaving school (Table 4.4a). The programme in Chile also provided participants with additional information about available academic scholarships and student loans for further study (Dinkleman and Martinez, 2011) while, in the Dominican Republic students were provided with a written statement and statistics detailing what was provided in the information session (Jensen, n.d.).

The Chinese programme included two treatment arms, either involving provision of information delivered in one session or, delivered as part of a longer series of career counselling sessions to help students identify career interests and highlight the importance of acquiring skills (Loyalka *et al.*, 2013). The final study, which took place in Madagascar (Nguyen, 2008) included three treatment arms. The intervention aimed to deliver information to parents about the returns to education. Both parents and children were invited to a teacher-parent meeting, the content of which differed according to which treatment group they were in. The first treatment, called the “statistics” treatment, provided participants with simple national statistics on the average returns to education, including information on the distribution of jobs and earnings of 25-year-old Malagasy men and women, by level of education. Parents also received an information card displaying mean earnings by gender and education. The second, “role model” treatment in this intervention involved role models delivering a presentation to participants on their background, educational experiences, and current job and standards of living. The third treatment group received a “combined” intervention. In this treatment, the “statistics” intervention was delivered first, followed by the “role model” interventions.

All of the included studies differed somewhat in their delivery method. The programme in China used trained existing class teachers to deliver the intervention (Loyalka *et al.,* 2013). Meanwhile, in Chile the programme implementers produced a DVD where ‘role models’ recounted their experiences of education and their current jobs, salaries. They also told viewers about available loans/scholarships. For the Dominican Republic programme, information was delivered by enumerators trained by the research team (Jensen, n.d.). In Madagascar, role models delivered presentations in person to both children and parents.

Comparisons

Three of the comparison groups in our included studies were given no information intervention beyond the standard provided in the school system. The comparison group in the Madagascar programme attended a meeting in which they discussed typical school matters, but did not receive either statistics or role model presentations on returns to schooling (a form of placebo, Nguyen, 2008).

Outcomes

All four studies reported effects on learning outcomes. Dinkleman and Martinez (2011) and Nguyen report a composite test score, Jensen (n.d) report performance as rated by the class teacher (Jensen, n.d.) and maths test scores (Loyalka *et al.,* 2013). All the studies also reported outcomes for different sub-groups, such as baseline maths score, gender and poverty level (*ibid*), baseline scores and socioeconomic status (Nguyen *et al.,* 2008) and baseline achievement level (Dinkleman nd Martinez, 2011). In addition to learning outcomes, studies also reported results for completion (Jensen, n.d.), dropout (Loyalka *et al.*, 2013), enrolment (Dinkleman and Martinez, 2011) and attendance (Dinkleman and Martinez, 2011; Nguyen, 2008). Follow-up periods were relatively short for three of the studies at around five months (Dinkleman and Martinez, 2011; Loyalka *et al.*, 2013; Nguyen *et al.,* 2008). One study conducted an initial follow-up after five years, with further follow-up surveys being conducted in the subsequent two years (Jensen n.d).

Study Design

Three of the included studies were cluster randomised controlled studies (Dinkleman and Martinez, 2011; Loyalka *et al.*, 2013; Nguyen, 2008) where assignment to the intervention took place at the school level.

Qualitative studies, process evaluations and project documents

We did not identify any qualitative studies, process evaluations or project documents associated with the included studies.

Table 4.4 a: Characteristics of Included Studies Providing Information

Included study	Setting	Intervention summary	Included outcomes	Follow- Up	Study design	Sample size
Jensen (no date)	Dominican Republic. Urban	A Student Survey was conducted to gather information on household and individual characteristics as well as expected earnings by education level. After the student survey respondents were given a written statement with information about estimated earnings and return to education based on the survey.	Completion, Learning	60 months	Cluster-RCT	1125 students
Loyalka *et al.*, 2013	China Rural	The two interventions were an information intervention and a counselling intervention. The information intervention presented statistics on the net returns (wages minus costs) associated with different levels of schooling in simple graphical and tabular forms. In schools that received the counselling intervention, grade 7 homeroom teachers and their principals also went to a central training location in each province. At each location a professional counsellor gave them a scripted training for a day and a half. During the counselling training, participants learned how to give four scripted 45-min lessons to grade 7 students on career planning skills.	Dropout, Learning	5 months	Cluster-RCT	11633 students
Dinkleman *et al.*, 2011	Chile Urban (and possibly per-urban)	The intervention provides students with information about how effort and good grades in school open up opportunities later on for further study, primarily by making it possible to apply for scholarships and government loans. This information was provided in a DVD called “Open the Box” (“Abre la Caja”) developed by the researchers and distributed to the students. This DVD collected the tertiary education experiences of 13 adults who grew up in poor	Enrolment, attendance, learning	5 months	Cluster-RCT	3313 students

		families in urban Chile. Their studies enabled them to become (among other things) civil engineers, graphic designers, chefs, social workers, lawyers and TV commentators. These life stories informed students about the existence of academic scholarships and student loans.				
Nguyen 2008	Madagascar Rural primary schools.	The intervention delivered information about the returns to education. Both parents and children were invited to a teacher-parent meeting, the content of which differed according to which treatment group they were in. The first, “statistics”, treatment, provided participants with simple national statistics on the average returns to education. Parents also received an information card displaying mean earnings by gender and education. The second, “role model” treatment in this intervention involved role models delivering a presentation to participants on their background, educational experiences, and current job and standards of living. There were three treatment groups involving role models, each differing by background and level of achievement as follows: Low to Medium, Low to High, or High to High. The third treatment group received a combination of the first two intervention groups.	Attendance, learning	5 months	Cluster-RCT	88 schools

4.4.2 Synthesis of findings

Due to the small number of studies we identified and heterogeneity in outcome measures we were not able to conduct any meta-analysis for this intervention area. Instead we have provided a narrative description of results by study, with all estimates standardised.

Effects of providing information to children on access and learning outcomes

Loyalka *et al.* (2013) included two treatment arms. For the counselling arm, which involved providing information to children over four sessions, including additional career planning activities and information on further-education funding options, the effect on dropout was 0.06 (95% CI [0.02, 0.09]), with slightly lower and less precise estimates in the information treatment arm where students received information on returns to education delivered during one session only (SMD=0.03, 95% CI [-0.01, 0.07]). The study also measured learning outcomes, using maths test scores as a measure with small effects observed for both treatments (Counselling: SMD=-0.04, 95% CI [-0.08, 0.00]; Information arm: SMD= 0.00, 95% CI [-0.04, 0.04]).

The evaluation of the Abre la Caja programme in Chile (Dinkleman and Martinez, 2011) found a larger reduction in recorded absence for children who were given the Abre la Caja DVD to take home and watch with their families (SMD=-0.08, 95% CI [-0.15, -0.02]), than students who watched the DVD at school (SMD=-0.06, 95% CI [-0.13, -0.00]). Effects on enrolment and test scores were small and not statistically significant for both treatment groups (see appendix H for full results).

The study of the programme in the Dominican Republic (Jensen, n.d) found a small positive effect on students completing the school year (SMD=0.08, 95% CI [0.02, 0.13]) and a small statistically insignificant effect on students completing secondary school (SMD=0.05, 95% CI [-0.01, 0.11]).

Finally, assessing the effect of a programme providing information about returns to education in Madagascar (Nguyen, 2008) found positive effects on school attendance for both the for the Statistics treatment (0.48, (95% CI [0.40, 0.55]) and the Role Model treatment (0.16, 95% CI [0.08, 0.24]) respectively. However, there was a negative effect in the combined treatment group (-0.2 [-0.28, -0.12]). Effects on learning outcomes were small and only significant for the statistics treatment arm (SMD=0.07, 95% CI [0.02, 0.12]; SMD = 0.02, 95% CI [-0.02, 0.06]; SMD = -0.03, 95% CI [-0.06, 0.01]). The findings also suggest that there were not major differences in effect for children with different socio-economic status.

4.4.3 Summary of findings

We identified four studies assessing the effect of providing information to children and parents, including two studies from Latin America and the Caribbean, one from East Asia and one from Sub-Saharan Africa. We were unable to conduct meta-analyses for any of these as no two studies reported on the same outcome measure. The limited number of studies does not allow for any strong conclusions about effects of programmes where information about returns to education is the primary input. The observed effects are mostly small and in a few cases negative, although the study of a programme providing information about returns to education to both children and parents in Madagascar found relatively large effects on school attendance.

5. Household level interventions

Household interventions are designed to improve access to education by reducing costs associated with attending school or highlighting the returns to education to parents or guardians of school-aged children. We reviewed the evidence on three types of interventions, reducing or eliminating user fees, cash transfers and provision of information on returns to education to parents. However, as we only identified one study assessing the provision of information to parents we present the findings of this study together with studies of interventions providing information to children in chapter 4. First we present the findings for interventions reducing or eliminating user fees, followed by cash transfers. Each sub-section starts with a description of the intervention type and its theory of change, followed by descriptive results and the findings addressing our research questions.

5.1 Reducing fees

Programmes reducing or eliminating school user fees aim to improve access to schooling. Direct user fees, including payments for tuition, uniforms, textbooks and parent-teacher association contributions are common in many L&MICs (Morgan *et al.,* 2012). Interventions to reduce or eliminate school user fees include removing all or some of these direct costs of schooling, for instance by providing school uniforms for free, or through the elimination of tuition fees, as has been done in many African countries over recent decades (Bentaouet-Kattan, 2006). Tuition fees may be universally removed, rolled out gradually or targeted towards particularly vulnerable groups (Morgan *et al*., 2012).

5.1.1 How may programmes reducing fees affect education outcomes?

Up to the 1980s, most schools in L&MICs used to charge school fees (Tomasevski, 2003). Following the World Education Forum in 2000, when the Education for All (EFA) movement was established, several governments started eliminating or reducing school fees. The EFA movement emphasised the importance of making good-quality education accessible to every child. Within this new context, school fees were seen as a major obstacle toward the realisation of inclusive schooling that all children could attend and where each child could learn regardless of their socio-economic background.

User fees for education are commonly thought to be a barrier to access in low-income households or for more marginalised children such as girls. The basic mechanism for fee reduction programmes is to reduce or eliminate costs that act as a barrier to access, and thereby increase participation (enrolment, attendance and retention) for students that would otherwise be unable to afford such costs (Evans, 2012). There are three different means that programmes in our included studies use to reduce user fees: eliminating school fees, reducing school fees and removing other costs (for example, the cost of uniforms).

Reduced or eliminated costs can be implemented at the student level, particularly if they aim to target subgroups of children such as girls, or cover costs other than fees, such as uniforms. Targeting students at the individual level singles them out and may result in increased morale and enthusiasm for attendance and attainment. Conversely, in schools where only some students are selected, the remaining students may feel disaffected and less inclined to attend or work hard at school (Evans, 2012).

An increase in enrolment is a key aim of removing or eliminating fees (Al-Samarrai *et al.,* 2007; Barrera *et al.,* 2007; Oketch *et al.,* 2010a). In order for fee reduction programmes to work effectively, they need to anticipate the increase in the number of students and ensure increased material, infrastructural and teaching support. Unless finance available to schools are increased, this can result in them not having sufficient resources to meet the new demand. This may result in increased pupil-teacher or student-material ratios and other reductions in school quality (Gajigo, 2012; Borkum, 2012). If there is a deterioration in school quality due to a lack of resources to cater for an increase in student numbers this may change both parents' and children's opinions of education (Borkum, 2012), resulting in dropout or decreases in enrolment. For example, in the case of fee removal in Kenya, the sudden increase in class sizes made teaching and learning conditions more challenging. As a consequence, many families decided to withdraw their children from public schools and send them to private schools (Tooley *et al.,* 2008). In this way, though user fee elimination may increase enrolment, this in turn may mean that children from families with more resources leave the public school system to join low-fee private schools, while the poorest remain in public schools, creating differential experiences of schooling (Bold *et al.,* 2010; Oketch *et al.,* 2010b).

5.1.2 Description of included studies

We included nine studies that evaluated the effects of programmes that reduced school user fees. These refer to nine unique programmes. The term 'study' is used here to refer to a unique evaluation of a programme, which may occasionally be described by several papers. The following section describes the studies' characteristics in more detail.

Setting

The included studies assessed programmes in Latin America and the Caribbean, Sub-Saharan Africa, South Asia and East Asia and the Pacific. In Latin America and the Caribbean there are two programmes, one in Ecuador (Hidalogo *et al.,* 2013) and one in Colombia (Barrera-Osorio *et al.,* 2007). Four of the included programmes were implemented in Sub-Saharan Africa, one each in Uganda (Grogan, 2009), South Africa (Garlick, 2013), Kenya (Evans *et al.,* 2012) and the Gambia (Gajigo, 2012). Two of the included programmes were implemented in East Asia and the Pacific, in Indonesia (Kharisma, n.d) and China (Hau, 2014) and one also in Nepal (Edmonds, 2014) in South Asia.

Five of the programmes reported in these studies were nationally implemented (Grogan, 2009; Garlick, 2013; Kharisma, n.d; Gajigo, 2012; Hau, 2014) and we assume that they were implemented across rural, urban and peri-urban areas. Two studies report on programmes implemented primarily in urban areas (Barrera-Osorio *et al*., 2007; Hidalgo *et al.,* 2013). Three studies reported on a programme implemented primarily in rural areas (Evans *et al.,* 2012; Hau, 2014; Edmonds, 2014).

Populations

Five of the programmes were targeted at the primary school level only, reporting on students in the first grade through to the sixth grade and students aged between five and 18 years old (Hidalgo *et al.,* 2013; Grogan, 2009; Evans *et al.,* 2012). Three programmes targeted both primary and secondary schools (Garlick, 2013; Barrera-Osorio *et al.,* 2007; Kharisma, n.d), reporting on students in grade 'zero' through to grade 12 and aged between seven and 20 years old. Barrera-Osorio and colleagues (2007) report that 47 per cent of the sample was female. Two studies (Gajigo, 2012; Edmonds, 2014) report on a programme implemented only in secondary schools. Three of the included studies evaluate programmes targeted at public schools only (Hidalgo *et al.,* 2013; Evans *et al.,* 2012; Gajigo, 2012) and three studies evaluate user fee elimination for both public and private schools (Garlick, 2013; Kharisma, n.d; Barrera-Osorio *et al.,* 2007). In two studies it is unclear which school provider was evaluated (Edmonds, 2014; Hau, 2014).

Interventions

All nine of the included programmes involve some form of school fee reduction or elimination. These include reductions of costs such as tuition fees, school uniforms, as well as indirect costs such as contributions to Parent Teacher Associations and field trips. All programmes take unique approaches to reducing or eliminating user fees and some incorporate additional components such as school grants or media campaigns. We describe them in more detail below, separating them into three main types: eliminating fees; reducing fees; and provision of school uniforms.

Eliminating fees

Four programmes removed tuition fees for certain categories of students. One study reports on universal elimination of fees for all primary school grades (Grogan, 2008), another one eliminated fees based on the socio-economic status of primary and secondary school neighbourhoods (Garlick, 2013). One eliminates fees for girls, introducing the policy regionally (Gajigo, 2012) and another pays schools fees to private schools that do not charge poor students (Adelman, 2015). Most of these programmes provide a per-student subsidy to offset the loss of revenue that the school would encounter due to fee elimination.

In the Universal Primary Education (UPE) Reform in Uganda (Grogan, 2009), the government of Uganda provided a per-student subsidy of 5,000 Ugandan shillings per annum for primary school grades one to three and 8,100 Ugandan shillings for primary grades four to seven. The subsidy provided by South Africa's No-Fee Policy evaluated by Garlick (2013) is based on the socio-economic status of the community in South Africa - a per-learner allocation is assigned to each community based on a table published annually by the department of Education. The Girls' Scholarship Programme in The Gambia uses a voucher redemption system (Gajigo, 2012). Participating schools could redeem vouchers from a fund administrator created by the Ministry of Education in The Gambia.

Of the programmes described in this section, three also incorporated additional components. The UPE Reform in Uganda (Grogan, 2009) abolished contributions to parent teacher associations (these had previously constituted 50% of schools' income). Furthermore, the government ended the requirement to wear school uniform, government spending in education was restructured and a media campaign was launched targeting early marriage in girls, aiming to increase female enrolment. Under South Africa's No Fee Policy, additional funding and training was provided to schools included in the programme to address areas of construction, grade expansion, student nutrition and safety at school, while training in financial management was provided to schools (Garlick, 2013). The Girls' Scholarship Programme included additional promotional activities including local media campaigns and regional workshops (Gajigo, 2012).

Reducing Fees

Four studies covered programmes that aimed to reduce the cost of tuition (Kharisma, n.d.,; Barrera-Osorio *et al.,* 2007; Hau, 2014; Edmonds, 2014). The Bantuan Operational Sekolah (School Operational Assistance) Programme (BOS) (Kharisma, n.d) reduced fees by providing funds to schools. In the Gratuidad fee reduction initiative (Barrera-Osorio *et al*., 2007) students in grade zero are exempt from all 'complementary' charges, such as report cards, school handbooks, ID cards, pedagogical materials, maintenance, and field trips (Barrera-Osorio, 2007). Students in grades one to nine with the lowest income status (level one) were completely exempt from 'complementary' charges. Students from households with the second lowest income level (level two) had their ‘complementary’ charges halved. For students in secondary school (grades 10-11) ‘complementary’ services and tuition fees were cut, again completely for those in households labelled as level one and by 50 per cent for those whose households were categorised as level two (Barrera-Osorio et al. 2007). The Three Tuition Policy Reform implemented three different reforms of tuition relief. The first, in 2001, introduced tuition control by setting a maximum chargeable amount for primary and junior high schools students in rural areas (Hou & Zhou, 2014). It also prevented any other fees being applied to students in addition to the controlled tuition. The second, in 2003, included free textbooks and a living stipend was provided in addition to a tuition waiver. The third, in 2006, introduced a tuition waiver across all rural students (Hou & Zhou, 2014). The Schooling Incentives Project Evaluation programme reimbursed or directly paid students’ school-related costs up to a maximum of NPR 3,950 per year (Edmonds, 2014).

Provision of school uniforms

Two studies report on programmes that provide school uniforms (Evans *et al.,* 2012; Hidalgo *et al.,* 2013). Both school uniforms and fees were provided in the NGO-run Kenya Child Sponsorship Programme (CSP) evaluated by Evans and colleagues (2012). The programme randomly assigned uniforms to students in schools and also paid their tuition fees. Students were also singled out to have their picture taken and sent to sponsors in the Netherlands. The participating schools also received additional benefits such as lessons in agriculture, a visiting nurse and funds for materials and classroom construction. The second study evaluated the Free Uniforms Programme in Ecuador (Hidalgo, 2013). The programme is a government initiative that provides uniforms to students in selected urban schools (rural schools had been covered by an earlier Free Uniforms Programme). The participating schools provided the Ministry of Education with details on student numbers, grades and sizes. The Ministry of Education then commissioned tailors to make the uniforms.

Comparisons

Seven of the included studies compared the effects of the intervention to business as usual (Hidalgo *et al.,* 2013; Grogan, 2009; Gajigo, 2012; Barrera-Osorio *et al.,* 2007; Kharisma, n.d; Hue, 2014; Edmonds, 2014). The study that evaluated fee elimination in South Africa (Garlick, 2013) compared schools qualifying to receive the intervention due to being below a school ‘poverty score’ cut-off, with schools just above the poverty score cut off. A study evaluating the provision of school uniforms and school fees in Kenya (Evans *et al.,* 2012) uses a within school comparison; students were randomly assigned to treatment or comparison groups, with parents paying for uniforms in the comparison group. All students had access to other elements of the program including lessons in agriculture, a visiting nurse and funds for materials and classroom construction in the school.

Outcomes

The included studies reported on a wide range of education outcomes. Seven of the nine studies report on enrolment outcomes (Barrera-Osorio *et al.,* 2007; Gajigo, 2012; Garlick, 2013; Grogan, 2009; Hidalgo *et al.,* 2013; Hau, 2014; Edmonds, 2014). Four of these studies conduct sub-group analysis by baseline socio-economic status and gender (Barrera-Osorio *et al.,* 2007; Grogan, 2008; Edmonds, 2014; Hau, 2014). Three studies report on attendance outcomes (Evans *et al.,* 2012; Hidalgo *et al.,* 2013; Edmonds, 2014). Three studies report on dropout outcomes, with two measuring dropout (Garlick, 2013; Kharisma, n.d), while one measured retention rates (Evans *et al.,* 2012). One study reports on completion, as measured by rate of repetition (Garlick, 2013), with subgroup analysis by baseline socioeconomic status. Two studies include attainment, measured as the highest level of education attained (Evans *et al.,* 2012) and number of years of schooling attained (Gajigo, 2012), with Gajigo (2012) analysing attainment by school grade. Finally, one study reports on learning in the form of standardised final exam scores (Edmonds, 2014).

Study types

The included studies use a wide variety of study designs and estimation strategies. One of the included studies is a cluster randomised controlled trial (Hidalgo *et al.,* 2013) and one study is a randomised controlled trials (Edmonds, 2014). Two studies use a regression discontinuity design (RDD) (Barrera-Osorio *et al.,* 2007, Grogan, 2008)., four studies use a controlled before-after study design with difference-in-differences estimation approach (Garlick, 2013; Kharisma, n.d.; Gajigo, 2012; Hau, 2014) and one study is a natural experiment with instrumental variables and ordinary least squares estimation approaches (Evans *et al.,* 2012).

Length of follow-up

The shortest follow-up period is six months (Hidalgo *et al.,* 2013), while two studies have a follow-up period of 12 months (Garlick, 2013 and Barrera-Osorio *et al.,* 2007) and one has a follow-up period of between 12 and 24 months (Gajigo, 2012). Three studies have a follow-up period of between 24 (Kharisma, n.d) and 48 months (Grogan, 2008), with Edmonds (2014) having a follow-up period of 30 months. The two longest follow-up periods are 72 months (Hau, 2014) and 96 months (Evans *et al.,* 2012).

Table 5.1 a: Characteristics of included studies – reducing or eliminating user fees

Included study	Setting	Intervention summary	Included outcomes	Follow-up	Study design	Sample size
Hidalgo *et al.*, (2013)	Ecuador, urban public primary schools	**The Free Uniforms Programme** in Ecuador is a government initiative that provides uniforms to students in selected schools. The school provides the Ministry of Education with details on student numbers, grades and sizes who commission tailors to make the uniforms.	Attendance; Enrolment	6 Months	Cluster- RCT	197 schools, 9,851 students
Grogan, (2008)	Uganda, urban and rural primary schools	**Universal Primary Education (UPE) Reform**: The government of Uganda universally abolished fees for all primary grades. A per student subsidy was provided to schools; 5,000 Ugandan shillings per annum for primary grades 1-3 and 8100 Ugandan shillings for primary grades 4-7. The programme also abolished contributions to parent teacher associations (these had previously constituted 50% of schools' income), the government ended the requirement to wear school uniform, government spending in education was restructured and a media campaign was launched targeting early marriage with the aim of increasing female enrolment.	Enrolment	48 months	Regression Discontinuity Design (RDD)	10,496 students

Included study	Setting	Intervention summary	Included outcomes	Follow-up	Study design	Sample size
Garlick, (2013)	South Africa, rural, urban and peri-urban public and private primary and secondary schools	**South Africa's No- Fee Policy**: the government policy abolished fees based on the socio-economic status of neighbourhoods in primary and secondary schools. A per-student subsidy to offset the loss of revenue was implemented with allocation assigned to communities based on their socio-economic status. Additional funding and training was provided to schools, funding was aimed at addressing areas of construction, grade expansion, student nutrition and safety at school while training was given in financial management to schools.	Drop-out Enrolment Completion	12 months	Controlled before - after with Difference in Difference (DID) estimation	40,940 students
Evans *et al.*, (2012)	Kenya, rural public primary schools	**Kenya Child Sponsorship Programme (CSP)**: Through the CSP children sponsored by donors in the Netherlands and elsewhere have their school fees covered and receive school uniforms. Children had their photo taken and sent to sponsors upon enrolment. The schools also benefited from visits from nurses, agricultural extension officers and grants for classroom construction, desks and books. These were school wide benefits and not restricted to sponsored students.	Drop out Attainment Completion	96 months	Natural experiment with Instrumental Variables and OLS regressions.	1,152 students

Included study	Setting	Intervention summary	Included outcomes	Follow-up	Study design	Sample size
Kharisma, (n.d.)	Indonesia, rural, urban and peri-urban public and private primary and junior secondary schools	**Bantuan Operational Sekolah (School Operational Assistance) Programme (BOS):** The BOS programme reduced fees by providing funds to schools. Funds were allocated on a per student basis, with 235,000 Indonesian Rp allocated per student in primary schools and 324,500 Rp per student in junior secondary schools. The funds were for the general costs of running a school with the expectation that a reduced amount would be charged for attendance. The funds could be used to cover the costs of registration of new students, textbooks, and reading books, stationery, test, development and training of teachers, school maintenance, transportation costs for poor students, salaries of teachers, as well as the costs of electricity, water and telecommunications.	Dropout	Approximately 24 months	Controlled before after study with Difference in Difference (DID) estimation	7,244 households
Gajigo, (2012)	The Gambia (rural, unclear: urban and peri-urban) public secondary schools	**The Girls' Scholarships Programme** abolished school fees for girls, introducing the policy to the country by region. Schools could redeem vouchers from a fund created by the Ministry of Education in The Gambia. Additional elements of the programme included promotional activities such as local media campaigns and regional workshops.	Enrolment, Attainment	Approx. 12-24 months. Programme was rolled out in different areas over time.	Controlled before-after study using Difference in Difference (DID) estimation	4,493 school-aged children

Included study	Setting	Intervention summary	Included outcomes	Follow-up	Study design	Sample size
Barrera-Osorio *et al.,* (2007)	Colombia, urban public and private primary and secondary schools	The **Gratuidad** fee reduction initiative aims to reduce the costs of attending school based on the student's grade and their household's income level. Households are graded into 1 of 6 income levels based on census data. Students within households graded as level 1 receive complete removal of 'complementary' charges. Students in households graded as level 2 have their 'complementary' charges halved. For students in secondary school (grades 10-11) 'complementary' services and tuition fees are cut, again completely for those in households labelled as level one and by 50% for those in level 2.	Enrolment	12 months	Regression Discontinuity Design (RDD)	604,169 students
Hau, C., 2014	China, rural primary schools	**Tuition waivers reform**: This included three key reforms: (i) tuition control (whereby schools were limited to a regulated tuition fee), (ii) tuition waivers, free textbooks, and living stipends for children from poor families, (iii) tuition waivers for all other rural children. The reforms followed a certain order, with tuition control typically being the first to be implemented, followed by the two waivers.These reforms were typically implemented first in rural and poorer counties and then extended to other counties and cities.	Enrolment	72 months	Difference in Difference (DID) estimation	Approx. 2,200 students

Included study	Setting	Intervention summary	Included outcomes	Follow-up	Study design	Sample size
Edmonds, E., 2014	Nepal, rural secondary schools	**Schooling Incentives Project Evaluation (SIPE) -** Scholarship treatment arm: For children in this treatment group, the Nepal Goodweave Foundation (NGF) NGO reimbursed or paid each child's schooling-related costs up to a maximum of NPR 3,950 per year. This assistance could include all schooling-related costs such as fees, tuition, uniforms, books and other supplies. Once the scholarship funds were exhausted, no additional support was available.	Enrolment Attendance Completion Learning	30 months	RCT	660 children

Table 5.1 b: Intervention Design Features of included studies

Author	Country	Programme name	Implementer(s)	Type of user fee reduction	Additional components	Target group	Conditionality	Amount received by school or student
Hidalgo *et al.*, (2013)	Ecuador	Free Uniforms Programme	The government of Ecuador	Provision of uniforms	N/A	Primary school students	N/A	Uniform only
Grogan, (2008)	Uganda	Universal Primary Education (UPE) Reform	The government of Uganda, Ministry of Education	Elimination of fees	Abolished parent teacher association contributions and requirement to wear school uniform. Government spending in education restructured. Media campaign targeting early marriage in girls.	Primary school age children, with focus on gender equity.	N/A	For schools: 5,000 Ugandan shillings for students in primary grades 1-3, 8,100 Ugandan shillings for those in primary grades 4-7
Garlick, (2013)	South Africa	No- Fee Policy (school fee elimination intervention)	The Government of South Africa: Ministry of Education, National and Provincial Level	Elimination of fees	N/A	Socio-economic status of region	N/A	Per-students allocation.
Evans *et al.*, (2012)	Kenya	Child Sponsorship Programme CSP	International Christelijk Steunfonds-Africa	Provision of uniforms and payment of fees by implementer	N/A	N/A	Attendance on first day of term	N/A
Kharisma, (n.d.)	Indonesia	Bantuan Operational Sekolah (BOS) Programme	Government of Indonesia, Ministry of Education	Reduction of fees	N/A	Unclear: possibly targeted at poor students	N/A	Per-student allocation of 235,000 Indonesian Rp for primary and

Author	Country	Programme name	Implementer(s)	Type of user fee reduction	Additional components	Target group	Conditionality	Amount received by school or student
		(School Operational Assistance)						324,500 Indonesian Rp for junior secondary school students.
Gajigo, (2012)	The Gambia	Girls' Scholarship Programme	The Government of The Gambia, Ministry of Education	Elimination of fees	Local media campaigns and regional workshops	Female students	N/A	Unclear, voucher redemption system
Barrera-Osorio *et al.,* (2007)	Colombia	Gratuidad fee reduction initiative	Bogotá's municipal government	Reduction of fees	N/A	Socio-economic status of household	N/A	N/A
Hau, C., (2014)	China	Tuition waivers reform	Government of China	Tuition fee waivers. Tuition control (whereby schools were limited to a regulated tuition fee)	Free textbooks and living stipends for children from poor families	Students from poor or rural households	N/A	N/A
Edmonds, E. (2014)	Nepal	Schooling Incentives Project Evaluation (SIPE) - scholarship treatment arm	Nepal Goodweave Foundation NGO	Reimbursement of school related expenses	N/A	Socio-economic status of household	NA	Up to NPR 3,950 per year.

5.1.3 Synthesis of findings

The results of our synthesis are presented in two sections. First, we present the detailed findings of the meta-analysis on the effects of fee reduction programmes on primary and secondary outcomes of interest, before we conclude the section and summarise the results.

Effects of user fee reduction interventions on enrolment, attendance, dropout rates, completion and attainment

This section reports the results of the meta-analysis of the effects of fee reduction interventions. We structure the presentation of results according to the 'ideal type' theory of change, starting with reported primary outcomes (enrolment, attendance, drop out, completion) followed by secondary outcomes (attainment).

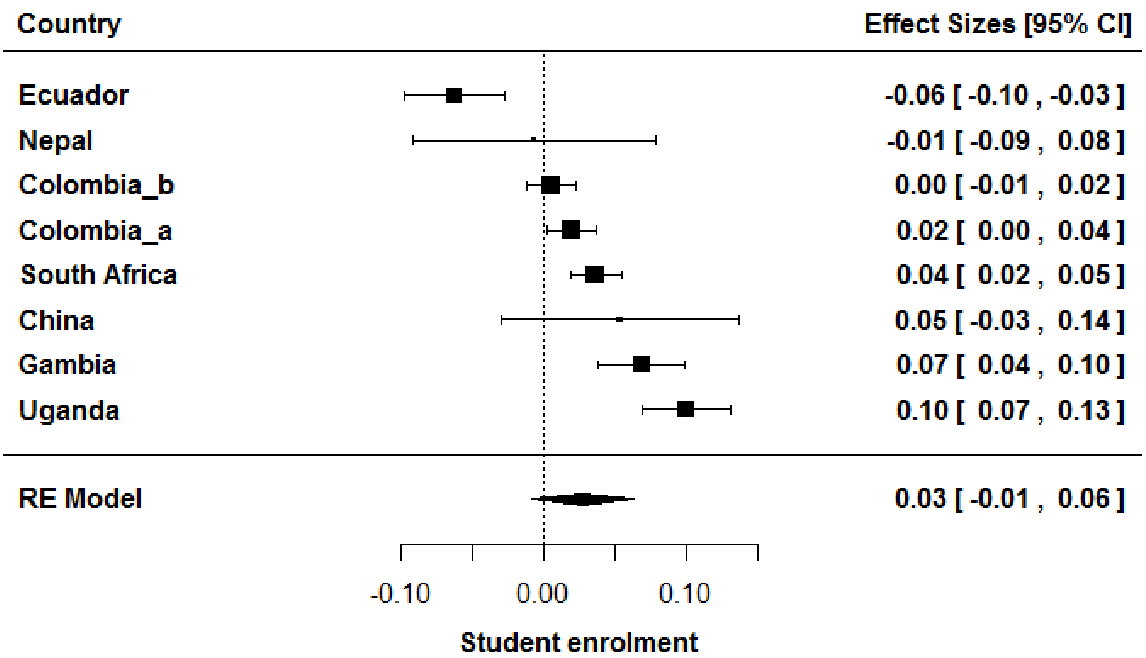
A total of nine studies provided data for meta-analysis, but none of the studies reported on all outcomes. The number of effect sizes that our included studies provided for each outcome ranged from two for attendance and attainment, to seven for enrolment. The included studies report outcomes at diverse follow-up periods range from six months to 96 months. If one study reported more than one follow- up period, we selected the one that was most similar to the other studies included in each meta- analysis. For Hau *et al.*'s (2014) study in China we included the effect size from the second no fee reform (tuition waivers plus free textbooks, and living stipends) as the intervention design was most similar to the other no fee interventions that we include. All effect estimates not included directly in the meta-analyses are reported narratively if they provide substantively different findings than those included in the meta-analysis. Estimates that are not substantively different than those already included in the meta-analysis are reported in technical Appendix H, section 5.2.

All effect sizes are expressed as standardised mean difference (SMD), interpreted as the magnitude of the number of standard deviation changes in the outcome for the intervention group as compared to students in non-programme schools.

Enrolment

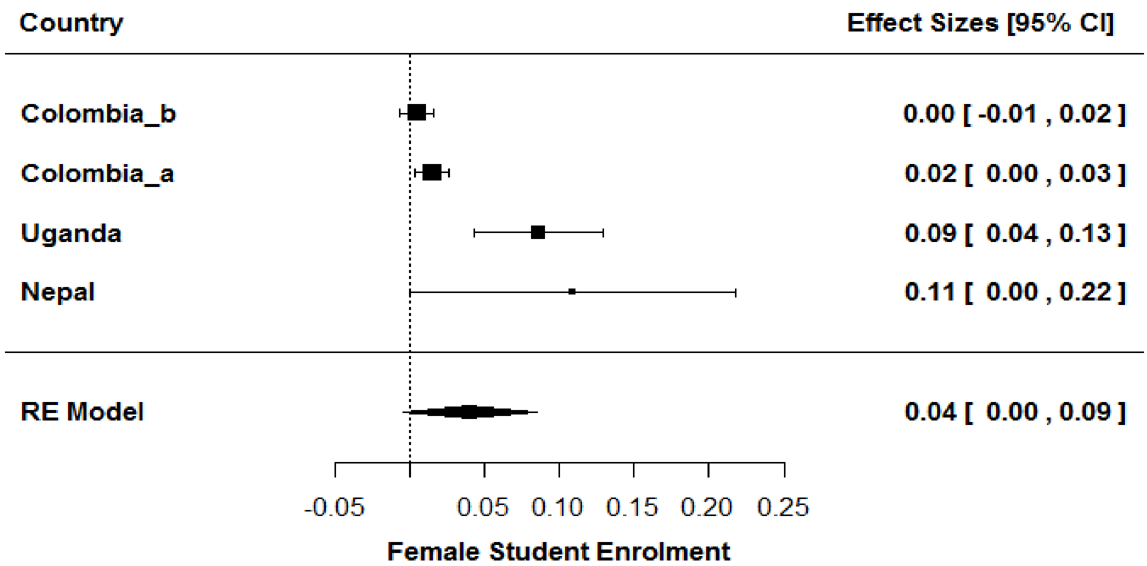
The overall average effect of user fee reduction on enrolment is 0.03, 95% CI [-0.01, 0.06], calculated under a random-effects model. The assessment of homogeneity suggests a large amount of between-studies variability (I^2 = 92.64%, τ^2 = 0.0023, Q (df = 7) = 62.55, p = <0.0001) indicating the effects did not all arise from the same population**.** Figure 5.1a supports the presence of heterogeneity. The effect sizes range from -0.06, 95% CI [-0.1, -0.03] in Ecuador (Hidalgo et al., 2013), to 0.1, 95% CI [0.50, 0.73] in Gambia (Gajigo, 2012).

Figure 5.1 a: Enrolment[9]



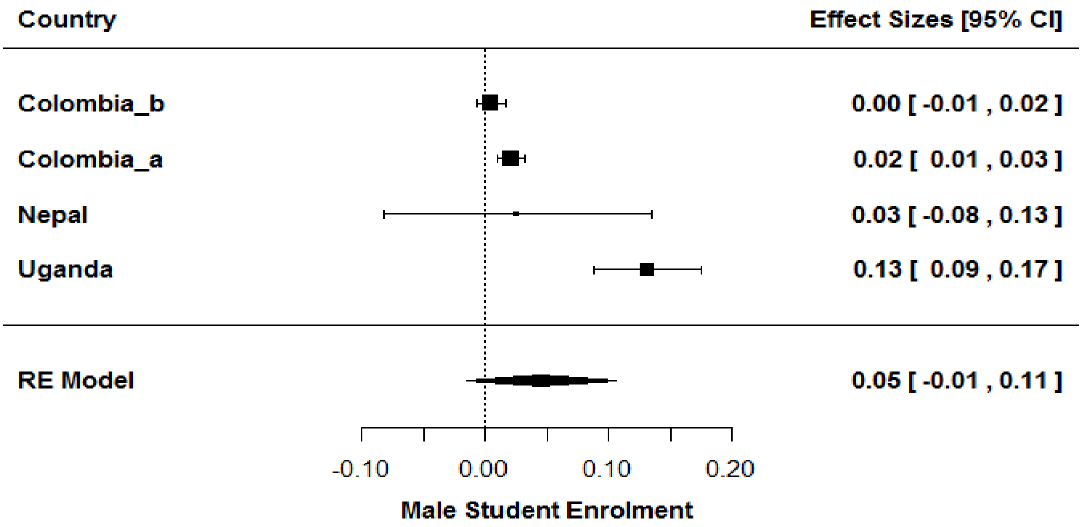
Three studies from Colombia, Nepal Uganda and Haiti report enrolment by gender sub-groups (Edmonds *et al.,* 2014, , Barrera- Osorio *et al*., 2007, Grogan, 2008). As suggested by figures 5.1b and 5.1c the magnitude of the effect is slightly larger for both boys and girls, but this appears to be a feature of the studies with sub-group analysis. Adelman *et al.* (2015) also provides estimates for different grades, and the results for grades five and six are substantively different from the main sample, with zero and small effects for girls and boys respectively (95% CI [-0.13, 0.13]; SMD=0.05, 95% CI [-0.09, 0.17]).

Figure 5.1 b: Enrolment: Female student subgroups[10]



[9] China: Hau, C., 2014; Colombia_a: Barrera-Osorio *et al.,* 2007; grades 1-9; Colombia_b: Barrera-Osorio *et al.,* 2007; grades 10-11; Ecuador: Hidalgo *et al.,* 2013; The Gambia: Gajigo, 2012; Nepal: Edmonds, E. 2014; South Africa: Garlick, 2013; Uganda: Grogan, 2008
[10] Colombia_a: Barrera-Osorio *et al.,* 2007; grades 1-9; Colombia_b: Barrera-Osorio *et al.,* 2007; grades 10-11 Nepal: Edmonds, E. 2014; Uganda: Grogan, 2008

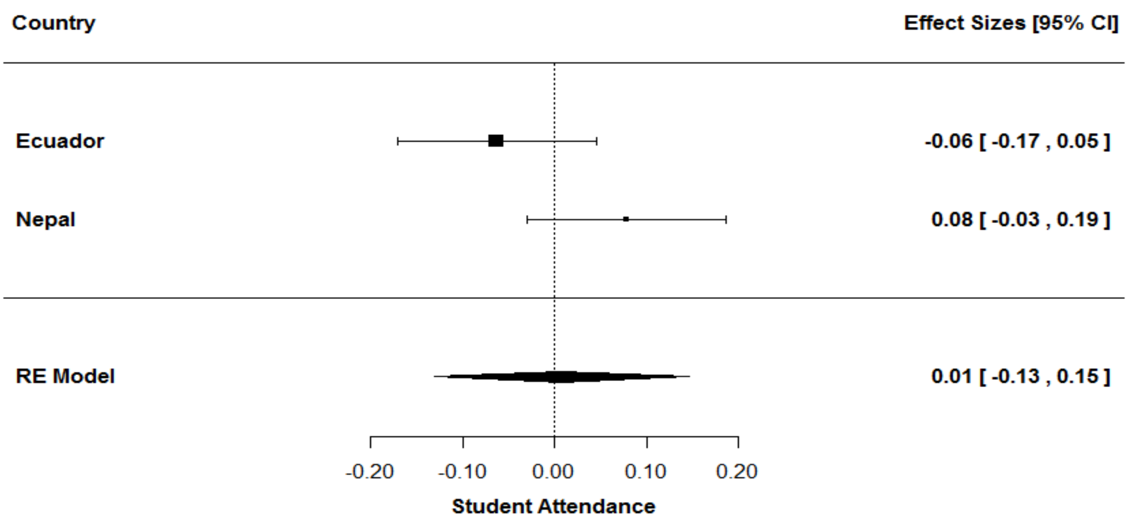
Figure 5.1 c: Enrolment: Male student subgroups[11]



Attendance

The overall average effect of eliminating fees on attendance is 0.01, 95% CI [-0.13, 0.15], calculated under a random effects model. The assessment of homogeneity suggests a moderate amount of between study variability (I^2 = 69.28%, τ^2 = 0.01, Q (df = 1) = 3.26, p = 0.0712). Figure 5.1d presents the forest plot with the results of the individual studies and the overall estimate. As expected, the confidence intervals overlap for these studies.

Figure 5.1 d: Student Attendance[12]



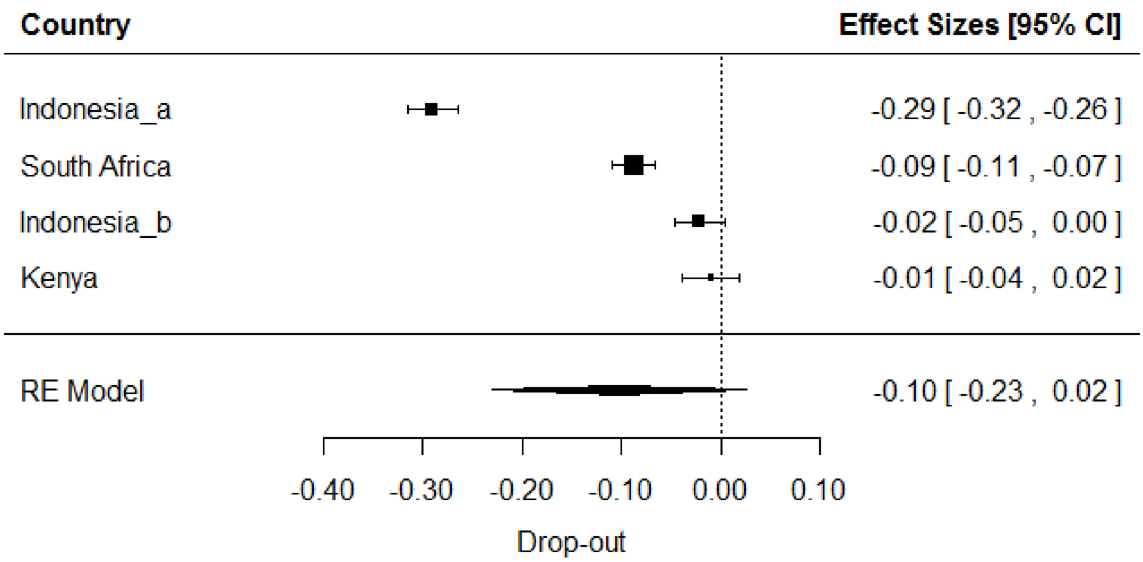
Dropout

The overall average effect of fee reduction on dropout is -0.10, 95% CI [-0.23, 0.02], calculated under a random-effects model. The assessment of homogeneity suggests a large amount of between-studies variability (I^2 = 99.03%, τ^2 = 0.0166, Q (df = 3) = 286.70, p = <0.0001), indicating the effects did not all arise from the same population.

[11] Colombia_a: Barrera-Osorio *et al.,* 2007; grades 1-9; Colombia_b: Barrera-Osorio *et al.,* 2007; grades 10-11;; Nepal: Edmonds, E. 2014; Uganda: Grogan, 2008
[12] Ecuador: Hidalgo et al., 2013; Nepal: Edmonds, E. 2014

Figure 5.1e supports the presence of heterogeneity. The effect sizes range from -0.29, 95% CI [-0.32, -0.26] for age group 7-15 in Indonesia (Kharisma, n.d.), to -0.01, 95% CI [-0.04, 0.02] in Kenya (Evans et al., 2012). The confidence intervals between most of the included studies do not overlap. As expected the results are sensitive to the inclusion of the effect size for age group 7-15 from the study in Indonesia (Kharisma, n.d.). Removing this study results in a smaller average effect of -0.04 (95% CI [-0.08, 0.01]). However, even after removal of this study, the degree of heterogeneity due to between-study variation remains high (I^2 = 90.93%, τ^2 = 0.0016).

Figure 5.1 e: Dropout13

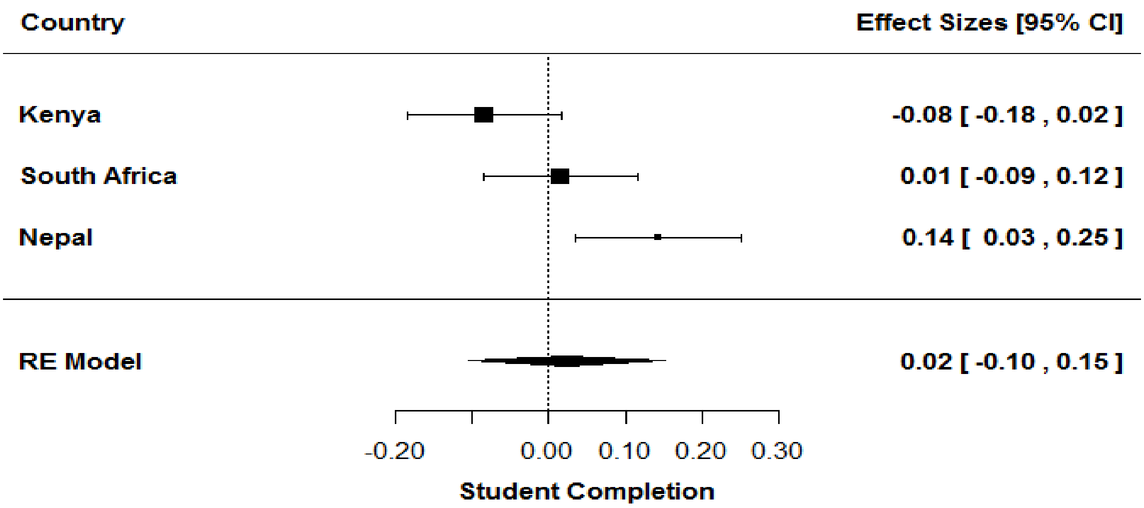


Completion

The overall average effect of user fees on completion is 0.02, 95% CI [-0.11, 0.15] under a random-effects model. The assessment of homogeneity suggests a large amount of between-studies variability (I^2 = 78.30%, τ^2 = 0.1, Q (df = 2) = 9.05, p = =0.01) indicating the effects did not all arise from the same population. Figure 5.1f supports the presence of heterogeneity. One of the included effect estimates is negative and statistically significant, the other two are positive. Results are sensitive to the inclusion of Schooling Incentives Project (SIPE) in Nepal. Removing this study results in a negative pooled effect estimate of – 0.03 (95% CI, [-0.13, 0.06]).

[13] Kenya: Evans et al., 2012; Indonesia_a: Kharisma, (n.d.); age group 7-15; Indonesia_a: Kharisma, (n.d.); age group 16-20; South Africa: Garlick, 2013;

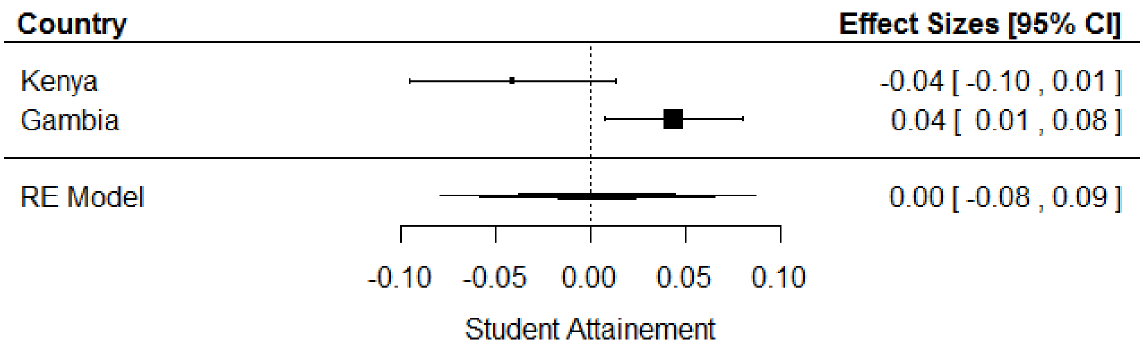
Figure 5.1 f: Student Completion[14]



Attainment

Two studies from Kenya and the Gambia assessed student attainment outcomes, as measured by the highest level of education attained (Evans *et al.,* 2012) and years of schooling attained (Gajigo, 2012). The overall average effect of fee reduction on student attainment is zero, 95% CI [-0.08, 0.09], calculated under a random effects model. The homogeneity tests (I^2 = 84.81%, τ^2 = 0.0031, Q (df = 1) = 6.58, p = 0.0103) suggest a large amount of between-study variability, indicating the effects did not arise from the same population. Figure 5.1g supports the presence of heterogeneity.

Figure 5.1 g: Student Attainment[15]



Composite test scores

Edmond *et al.*'s study (2014) evaluating a scholarship programme designed to cover direct, out-of-pocket schooling expenses in Nepal is the only study that reported on learning outcomes. It measures composite test scores of students who took the end of year exam and did not fail, finding a small negative, but not statistically significant, effect on test scores (SMD= -0.06, 95% CI [-0.17, 0.05]).

---

[14] Kenya: Evans *et al.,* 2012;; Nepal: Edmonds, E. 2014; South Africa: Garlick, 2013
[15] The Gambia: Gajigo, 2012.
Kenya: Evans *et al.,* 2012

5.1.3 Summary of findings and discussion

We identified ten studies of programmes eliminating or reducing school user fees. None of the studies reported on all outcomes and we were able to conduct meta-analysis for enrolment, attendance, dropout, completion and attainment. The effects range from 0.01 for student attendance (95% CI [-0.13, 0.15]) to 0.15 for completion (95% CI [-0.11, 0.42]). For all outcomes the confidence intervals of the average effect estimate cross the line of no effect. There is a large amount of heterogeneity, suggesting substantively different effects in different contexts.

Table 5.2 a: Descriptive findings: Process and implementation

Descriptive findings: Process and implementation	Context	Citation (Info. type)
Gaming behaviour from elected officials households (Garlick 2013 and Castaneda 2005).	South Africa and Columbia	Garlick 2013 (Impact Evaluation) Castaneda 2005 (Process Evaluation)
The targeting process may have resulted in the mis-targeting of schools and a failure to include schools from poor communities.	Indonesia South Africa	Soharyo 2006 (Project Document) Garlick, 2013 (Impact Evaluation) Giese et al. 2009 (Mixed Methods Study)
Change in accountability from parents to those implementing fee reduction or removal caused three primary issues; poor management in the provision of goods or funds, misuse of funds and insufficient funds to schools.	South Africa Ecuador	Giese et al. 2009 (Mixed Methods Study) Mestry, 2013 (Project Document) Hidalgo et al, 2013 (Impact Evaluation) Juan 2007 Marishane 2013 (Qualitative Study)
A failure to monitor and inspect schools resulted in poor policy enforcement and the continuation of fee charges in programme schools.	South Africa	Setoaba 2011 (Qualitative Study)
Reducing user-fees can reduce the quality of education due to less parental involvement in schools, increased student-classroom and student-resource ratios and increased delinquent behaviours.	South Africa Uganda	Garlick (2013) Impact Evaluation Nkosi, 2011 (Project Document) Ministry of Education and Sports of Uganda, 2005 (Project Document)
Misleading public messaging about the programmes meant that some parents did not pay for any associated costs of education, or reduced their fee-paying behaviour even if their children were in non-programme schools.	Indonesia South Africa Uganda	Soharyo, 2006 (Project Document) Giese et al. 2009 (Mixed Methods Study) Ekaju, 2011 (Project Document)
Limited funds and time caused poor training of implementers and schools, hampering programme implementation. Misunderstanding of the programme exacerbated unclear public messaging and inadequately prepared schools for implementing the programme.	Indonesia South Africa	Soharyo, 2006 (Project Document) Giese et al. 2009 (Mixed Methods Study) Nkosi, 2011 (Project Document) Marishane, 2013 (Qualitative Study) Setoaba 2011 (Qualitative Study)

Table 5.2 b: Descriptive findings: Contextual Factors

Descriptive findings: Contextual factors	Context	Citation (info. type)
The South African schooling system reflects inequalities regarding race, class and gender. Perceived low returns for education are considered a key causal factor in the poor performance of students compared to other African countries.	South Africa	Mestry 2013 (Project Document) Borkum 2009 (Project Document)
Strong political will for the implementation of the Universal Primary Education (UPE) Reform, resulted in its implementation 6 months after the original government pledge. This resulted in insufficient physical and human resources to support the programme.	Uganda	Ministry of Education and Sports of Uganda 1999 (Project Document)
Pre-existing gender inequalities and the cost of education is significant barrier to education in China, though many rural families recognise the long-term value of education.	China	Hau and Bo, 2014 (Impact Evaluation)

5.2 Cash Transfers

Cash transfers are social safety-net programmes that provide a direct transfer of cash to households, mothers or children. The primary objective of such programmes is to increase the human capital of the poor. Cash transfer programmes are typically classified into two main categories. Conditional Cash Transfers (CCTs) transfer money to households or children conditional on certain behaviour, such as school enrolment and attendance above a certain rate or visits to health centers. Unconditional Cash Transfers (UCT), also include a money transfer, but do not come with any explicit conditions (Baird *et al.,* 2010). However, Gaarder (2012) and Baird *et al.* (2013) argue that in practice cash transfers exists on a continuum from transfer programmes with no explicit conditions or label, via programmes with implicit directions on use (labelled programmes), programmes with explicit conditions but no enforcement to programmes with conditions that are monitored and enforced.

5.2.1 How may programmes providing cash transfers affect education outcomes?

The main mechanism through which cash transfers are thought to affect education outcomes is through the removal of financial barriers to education (Miller *et al.,* 2012). By providing families with additional funds, cash transfers aim to increase school enrolment and attendance, while reducing the risk of children dropping out by decreasing the direct (uniform, textbooks, etc.) and indirect costs (loss of income) of schooling.

There are several factors that might prevent households from investing in education. These include a lack of information about the potential returns to education, or community/familial norms regarding education. Families that already invest in education may not invest enough to have the necessary wider social impact that cash transfers attempt to achieve (Baird *et al.,* 2010). Under these circumstances it is thought that conditions linked to education may result in increased attendance and higher productivity (Attanasio *et al.,* 2006). Conditionalities are assumed to be particularly effective in cases where parents place low priority on investing in education (Akresh *et al.,* 2013). Conditionality is also expected to have an effect on child labour as both the parents and the child are incentivised to remain in school, rather than joining the labour force (Mo *et al.,* 2011). Conditional cash transfers do however have a higher cost attached to them because of the monitoring and enforcement of conditionalities (Benhassine *et al.,* 2013). As a result the extent of monitoring and enforcement of conditions vary between programmes.

Labelled transfers are unconditional and the label has no impact on how recipients can spend the money. As suggested in behavioural economics, in cases where a transfer is labelled as an 'education' transfer parents may mentally process the money for the purpose it is labelled. It is expected to indicate the value of education to parents (Benhassine *et al.,* 2013), who may then change their behaviour towards supporting the education of their children. Labelled transfers may therefore affect education outcomes in two ways; additional household funds toward education and changed parental behaviours.

There are a number of other components of cash transfer design that may influence final outcomes. Some cash transfers target marginal students, like girls, that are likely to be out of school or are at risk of dropping out. Other marginal students include low ability children who are often thought to have lower returns from education and as a result have reduced educational investment.

The size of the transfer amount may also influence whether education outcomes are improved. The opportunity cost of attending school increases with age as students have increased likelihood of finding work or getting married. Some transfer programmes increase the value of the transfer with the age of the child to compensate for this (Baulch, 2011).

Who receives the transfer is also a factor that is thought to influence the effects of the programme. For instance, one of the theories underlying many cash transfer programmes is that mothers are more likely to use money to improve their child's welfare. As a result many transfer programmes are designed to transfer money to mothers, with the expectation they will use the funds to improve child wellbeing, including through increased spending on education (Baulch, 2011).

Finally, contextual factors such as the quality of services available influence the extent to which final outcomes such as learning are improved. If there is a lack of supply of high quality of schools and teachers then increased participation is unlikely to lead to improved learning outcomes.

5.2.2 Description of included studies

We included fifty studies that evaluated the effect of cash transfer programmes. These refer to thirty eight unique programmes. The term 'study' is used to refer to research from an author team, which is occasionally reported across several papers. In many cases we also include several studies of the same programme, with authors reporting on different outcomes, participants and time periods. Table 5.4a provides an overview of the characteristics of included studies and these are also summarised below.

Setting

The included studies covered programmes in Latin America and the Caribbean, Sub-Saharan Africa, East Asia and Pacific, the Middle East and North Africa and Europe and Central Asia. In Latin America and the Caribbean we identified studies of eighteen different programmes. Four of these programmes were implemented in Brazil (de Oliveira and Kassouf, 2012; de Janvry *et al.,* 2012; de Brauw *et al*., 2014; Pianto and Soares, 2004). Two programmes were implemented in Columbia (Attanasio *et al*., 2006 and Barrera-Osorio *et al.,* 2011), Nicaragua (Macours and Vakis 2009; Maluccio *et al.,* 2009) and Honduras respectively (Benedetti *et al.,* 2015 and Glewwe; Olinto 2004). We also identified one study each of programmes in Argentina (Heinrich 2005), Ecuador (Edmonds and Schady 2011), Chile (Galasso 2006), El Salvador (De Brauw and Gilligan 2011), Uruguay (Amarante *et al.,* 2011), Jamaica (Levy and Ohls 2007), Paraguay (Ribas *et al.,* 2011) and Mexico (Schultz 2004).

We identified seven studies evaluating programmes in Sub-Saharan Africa. Two of these programmes were implemented in Malawi (Baird *et al., y* 2010; Luseno 2013) and one each in Lesotho (Pellerano *et al.,* 2014), South Africa (Eyal and Woolard 2014), Tanzania (Evans *et al.*, 2014), Zimbabwe (Robertson *et al.,* 2013) and Burkina Faso (Akresh et al., 2013). Six included studies assessed programmes implemented in East Asia and Pacific. Three of these were implemented in Cambodia (Filmer and Schady 2006; Filmer and Schady 2014; Barrera-Osorio and Filmer 2014). One programme was implemented in Indonesia (Sparrow 2007), one in the Philippines (Chaudhury et al., 2013) and one in China (Mo *et al.*, 2011). Five studies assessed programmes implemented in South Asia, with three of these being implemented in Bangladesh (Ferré and Sharif 2014; Khandker *et al.*, 2003; Baulch 2011) and one each in Nepal (Edmonds and Shrestha 2014), Pakistan (Alam and Baez 2011), Morocco (Benhassine *et al.*, 2013) and Turkey (Ahmed *et al.*, 2007).

Thirteen programmes were implemented in both rural and urban areas. Eight of these were in Latin America and the Caribbean.[16] Three were implemented in Sub-Saharan Africa (Baird *et al.*, 2010; Pellerano *et al.*, 2014; Eyal and Woolard) and two in South Asia (Baulch 2011; and Ferré and Sharif 2014). Five programmes were implemented in rural, urban and peri-urban areas, three of these being from Latin America and the Caribbean (de Janvry *et al.*, 2012; de Brauw *et al.*, 2014; Levy and Ohls 2007) and one each from East Asia and Pacific (Sparrow 2007) and Sub-Saharan Africa (Robertson *et al.,* 2013). One programme was implemented in urban areas alone (Barrera-Osorio *et al.,* 2011).

Eleven programmes were implemented in rural areas only. Five of these programmes were in Latin America and the Caribbean[17]. Two rural programmes were implemented in South Asia (Khandker *et al.*, 2003; Chaudhury and Parajuli, 2006; Alam and Baez, 2011), two in Sub-Saharan Africa (Evans *et al.*, 2014; Akresh *et al.*, 2013). One rural programme was implemented in Middle East and North Africa (Benhassine *et al.*, 2013) and one in East Asia and Pacific (Mo *et al.,* 2011). Two programmes were implemented in rural and peri-urban areas (Benedetti *et al.,* 2015; Chaudhury *et al.*, 2013). In six programmes the context of implementation was unclear[18].

Populations

Nineteen of the included cash transfer programmes aimed to benefit both primary and secondary school children. Eleven of these programmes were implemented in Latin America and the Caribbean, where the student's age ranged between 6 and 17 and grades ranged between 1 and 9 (see table of characteristics for details). Four of these programmes were implemented in Sub-Saharan Africa; students age ranged between 7 and 13 (grade ranges were not reported in these studies)[19]. Two programmes implemented in both primary and secondary schools were in East Asia and the Pacific (Sparrow, 2007 and Chaudhury *et al.*, 2013). The age range of students was between 10 and 11 in Sparrow (2007) and all grades were included with exception of grades 1-3. Student's age and grade is not reported in Chaudhury *et al.* (2013). One programme in South Asia is targeted at primary and secondary school children (Edmonds and Shrestha 2014). The average age of students in this study is

[16] Attanasio *et al.*, 2006; Pianto and Soares, 2004; Kassouf, 2012; Heinrich, 2005; Edmonds and Schady, 2011; Galasso, 2006; De Brauw and Gilligan, 2011; Amarante *et al.,* 2011.
[17] Glewwe and Olinto, 2004; Schultz, 2004; Maluccio *et al.,* 2009; Ribas *et al.,* 2011; Macours and Vakis, 2009
[18] Filmer and Schady, 2009, 2011, 2014; Barrera-Osorio and Filmer, 2014; Luseno, 2013; Edmonds and Shrestha, 2014; Ahmed *et al.,* 2007; Filmer and Schady, 2006
[19] (Pellerano *et al.,* 2014); (Eyal and Woolard, 2014); (Luseno, 2013) (Robertson *et al.,* 2013).

12. One programme in Europe and Central Asia is targeted at primary and secondary school children (Ahmed *et al.,* 2007); there are no details of participant characteristics for this study.

Eight included programmes were targeted at secondary school students only. Three of these were in East Asia and Pacific (Filmer and Schady, 2006; Filmer and Schady, 2009; Mo *et al.*, 2011). Grades range between 7 and 9 across these studies, with one study (Mo *et al.,* 2011) reporting the average age of students as 12 years. Two secondary school programmes are implemented in Latin America and the Caribbean (Heinrich, 2005; Barrera-Osorio *et al.*, 2011). Grades range between 6 and 11 in these studies and the average age is between 14 and 15. Two secondary school programmes are implemented in South Asia (Khandker *et al.,* 2003 and Alam and Baez 2011). No participant characteristics are reported in Khandker *et al.* (2003), students in grades 5 to 9 are included in Alam and Baez (2011). One secondary school programme is implemented in Sub-Saharan Africa (Baird *et al.*, 2010); the average student age in this study is 15.

Nine cash transfer programmes were implemented in primary schools. Three of these were in Latin America and the Caribbean (De Brauw and Gilligan, 2011; Glewwe and Olinto, 2004 and Maluccio *et al.* 2009). The age range of students is 5 to 13 in these studies, no grades are reported. Two primary school level programmes are in South Asia (Baulch, 2011; Ferré and Sharif, 2014). The age range of students in these studies is between 6 and 15, grade range is not reported. Two primary school level programmes are in Sub-Saharan Africa (Akresh *et al.,* 2013 and Evans *et al.*, 2014). The average age reported in Akresh *et al.* (2013) is 10; no other population characteristics are reported for these studies. One programme in the Middle East and North Africa is a primary school cash transfer (Benhassine *et al.,* 2013), the grade range of students in the study is 1 to 6 and the average age is 9. One programme in East Asia and Pacific is a primary school cash transfer (Barrera-Osorio and Filmer, 2014), grades 4 to 6 were included in this study.

In two cash transfer programmes it is unclear in which schools they were implemented. The first, reported in Macours and Vakis (2009) had students ranging from 7 to 18 years old and can be assumed to include both primary and secondary schools. The second (Kassouf, 2012), had students ranging between 10 and 15 years old.

In the majority of included programmes it was unclear if they were implemented in public or private schools. Five programmes were clearly implemented in public schools[20]. Two are clearly implemented in public and private schools (Baird *et al.,* 2010 and Sparrow, 2007). Two cash transfer programmes are implemented in public and non-governmental schools (Ferré and Sharif, 2014 and Khandker *et al.*, 2003).

Interventions

The thirty eight included programmes represent different types of cash transfer programmes. The programmes vary across a range of dimensions, including the intensity of conditions, the size of the cash transfer and the recipient of the transfer. The table of characteristics include detailed descriptions of all programmes and we summarise these below.

Conditionality

We followed Baird *et al.* (2012) who classified programmes according to the extent of conditionality on a scale from 0 to 6 as follows: Programmes that are unconditional and not targeted at children such as old aged pensions (0); unconditional programmes that have the aim of improving education (1); labelled transfers where participants are explicitly told that

[20] Benhassine *et al,.* 2013; Heinrich, 2005; Chaudhury *et al.,* 2013; Schultz, 2004; Ahmed *et al.,* 2007); Alam and Baez, 2011.

they are for use on education, but without any conditions (2); conditional transfers where conditions are not monitored or enforced (3); conditional transfers where conditions are monitored imperfectly and with little enforcement (4); conditional transfers where school enrolment condition monitored and enforced (5); and finally, conditional transfers where school attendance condition is monitored and enforced (6)

One of thirty eight programmes is classified as unconditional and not targeted at children (0), the BPC programme (Kassouf, 2012) in Brazil. One programme was classified as an unconditional programme that had the aim of improving education (1), the Child Support Grant (CSG) in South Africa (Eyal and Woolard, 2014). One programme was a labelled transfer programme where participants were explicitly told that funds were for use on education, but without any conditions (2), namely the Child Grants Programme in Lesotho (Pellerano *et al.,* 2014).

Nine of the included programmes are classified as cash transfers with clear conditions, but these are neither monitored nor enforced (3). Three of these programmes had conditions of attendance at school (Malawi Social Cash-Transfer; the Punjab Female Stipend Program; Bolsa Escola). One programme had conditions of attendance and participation in an extended day at school (Programa de Erradicacao do Trabalho Infantil (PETI). Three programmes had conditions of attendance at both school and health centres for children (Shombhob project; Atención a Crisis; Bono de Desarrollo Humano). Two final programmes had conditions of attendance at school and health centres for children, but also medical check-ups for pregnant women (Plan Nacional de Atencion a la Emergencia Social (PANES); the Social Risk Mitigation).

Seven included programmes were conditional transfers where conditions were monitored imperfectly and with little enforcement (4). One of these cash transfers is conditional of attendance at school with annual grade progression (Programa Nacional de Becas Estudiantiles). The Female Secondary Stipend Programme (FSSP) was specifically targeted at girls, conditional on attendance and maintaining a 45 per cent in test scores, as well as remaining unmarried: (Khandker *et al.,* 2003). The conditions attached to Chile Solidario (Galasso, 2006) were linked to participation in the wider programme. Families had to work with a social worker for a period and were provided with psycho-social support. The remaining four programmes in this category had conditions linked to enrolment and attendance in school and health clinics (Bono 10,000; Bolsa Familia; Progresa; Tekopora Conditional Cash Transfer Programme). Three included programmes were conditional with monitoring and enforcement of a school enrolment condition (5). The CESSP Scholarship Programme (CSP) and Jaring Pengaman Sosial (JPS) required enrolment with passing grades for students. In addition to enrolment Pantawid Pamilya required the attendance of children and pregnant mothers at health centres, and the attendance of a family member at training sessions.

Finally, eleven included programmes were conditional and had monitoring and enforcement of a school attendance condition (6). Two of these programmes had monitoring and enforcement of attendance in class (Schooling Incentives Project Evaluation (SIPE); an un-named programme in Mo *et al.* (2011)). Two programmes were conditional on both enrolment and attendance (Subsidios Condicionados a la Asistencia Escolar; Comunidades Solidarias Rurales). Three programmes were conditional on enrolment, attendance and maintaining a passing grade (Cambodian Primary Scholarships Pilot; the Primary Education Stipend (PES) Programme (Baulch 2011); the Japan Fund for Poverty Reduction). Four programmes had both education and health conditions, but these were separate conditions for two transfers participants were eligible for (Familias en Accion: unnamed programme evaluated in Evans *et al*. (2014); Glewwe and Olinto, 2004; Red de Proteccion Social).

In addition, five included programmes had both conditional and unconditional arms. Tayssir's (Benhassine *et al.,* 2013) unconditional arm was categorised as an education support programme, while the conditional arm had monitoring and enforcement of school enrolment and attendance conditions (6). The conditional arm of the Nahouri Cash Transfers Pilot Project (Akresh *et al.,* 2013) similarly had school enrolment and attendance conditions which were monitored and enforced (6). The Programme of Advancement Through Health and Education (PATH) (Levy and Ohls, 2007) had three different intervention arms. The conditional arms (health and education) are labelled as 6, while the social assistant grant was unconditional. The conditional arms of both the Zomba Cash Transfer Programme (Baird *et al*., 2010) and the Manicaland HIV/STD Prevention Project (Robertson *et al.,* 2013) had conditions of enrolment and attendance but there was no clear monitoring or enforcement of those conditions (3).

Recipient

The majority of cash transfer programmes are described as being received by either households or parents. Nine programmes are however paid directly to mothers[21]. Four of the programmes provided funds directly to students, thereby bypassing caregivers as decision makers over the use of funds.[22] One included programme, The BPC program (Kassouf 2012), provided cash to pension aged persons within the household.

Outcomes

The included studies reported on a wide range of education outcomes, with the majority focusing on school participation outcomes. Twenty seven of the fifty studies measured attendance outcomes, with twelve studies reporting attendance by gender subgroup. Twenty nine studies assessed enrolment outcomes, with eleven of these studies reporting enrolment by gender subgroup. Twenty nine studies provided some measure of completion, with eight of these also reporting estimates for gender subgroups. Ten studies measured dropout, with four studies also reporting dropout by gender subgroup.[23] Fewer studies assessed learning outcomes. Ten studies reported on maths (four by gender sub-group), nine on language arts (four by gender sub-group) and three on composite score (two by gender sub-group). Finally, two studies reported on cognitive test score outcomes.

Study types

The included studies used a variety of study designs and estimation strategies. Nineteen studies were cluster randomised studies and two studies were randomised at the individual level. Thirteen studies used a regression discontinuity design (RDD), while twelve studies used a controlled before-after study design with statistical methods such as propensity score matching and difference in difference to control for selection bias. Finally, four were natural experiments.

[21] Familias en Accion (Attanasio *et al,* 2006); Atención a Crisis (Macours and Vakis, 2009); Bolsa Escola (de Janvry *et al.,* 2006 and 2012); Bono de Desarrollo Humano (BDH) (Edmonds and Schady, 2011); Chile Solidario (Galasso, 2006); Child Support Grant (CSG) (Eyal and Woolard, 2014); Programa de Erradicacao do Trabalho Infantil (PETI) (Pianto and Soares, 2004); Progresa (Schultz, 2004); an unnamed programme evaluated in Evans et al. (2014)

[22] Cambodian Primary Scholarships Pilot (Barrera-Osorio and Filmer, 2014); Female Secondary Stipend Programme (FSSP) (Khandker *et al.,* 2003); Jaring Pengaman Sosial (JPS) (Sparrow, 2007); and The Japan Fund for Poverty Reduction (Filmer and Schady, 2006).

[23] A table with the break-down of outcomes with subgroup analysis by study is available upon request.

5.2.3 Synthesis of findings

Below we present the findings of the meta-analysis on the effects of cash transfer programmes on primary outcomes. We have structured the presentation of results according to the causal chain, starting with effects on school participation outcomes, followed by the learning outcomes.

Fifty studies provided data for meta-analysis, but none of the studies reported on all outcomes. We included 48 effect sizes for enrolment. Several of these effect sizes represent independent subsamples from the same studies, such as different grades or different genders. Three studies include different treatment arms (Barrera-Osorio *et al.*, 2011; Akresh *et al.*, 2013; Baird *et al.*, 2010), with Barrera-Osorio *et al*. (2011) having three treatment arms, and the other two studies having two different treatment arms. We included all of these estimates in the meta-analysis as we had a sufficient number of studies to use robust variance estimation, which accounts for dependencies between effect sizes.

We included 38 effect sizes for attendance. As for enrolment, several of these effect sizes represent independent subsamples from the same studies, such as different grades or different genders. The studies with multiple treatment arms included in the meta-analysis for enrolment also reported on attendance and all treatment arms are included in the meta-analysis, with robust variance estimation applied as above. For completion outcomes we included 27 effect sizes. As with enrolment and attendance, several of these represent independent sub-samples from the same studies. Two of the studies with multiple treatment arms are included in the analysis (Barrera-Osorio *et al.,* 2011; Baird *et al.,* 2010) and, as above, we applied robust variance estimation.

For maths, language arts, dropout and all sub-group analyses, we did not have sufficient number of studies to use robust variance estimation in the meta-analysis. Therefore, we had to select which effect sizes to include when there were dependent effect sizes. In the case of studies with multiple treatment arms, we selected the CCT treatment arm as this was most similar to the other studies in the sample. Benhassine *et al.*'s study of cash transfers in Morocco includes four different treatment arms, but only one of the treatment arms is compared to a business as usual comparison group and we included effect sizes for this comparison in all the analyses. Most of the included studies report follow up periods that ranged between 12-24 months, so when there was a choice of follow up periods, we selected the one that fell within this period. If a study reported both 12 and 24 months follow up, we included the longest follow up period in the meta-analysis. If a study only included a follow up period which fell outside of this range we still included it in the meta-analysis.

All effect sizes are expressed as standardised mean difference (SMD), interpreted as the magnitude of the number of standard deviation changes in the outcome for the intervention group as compared to students in non-programme schools. SMD scores are interpreted as the number of standard deviation changes in the outcome.

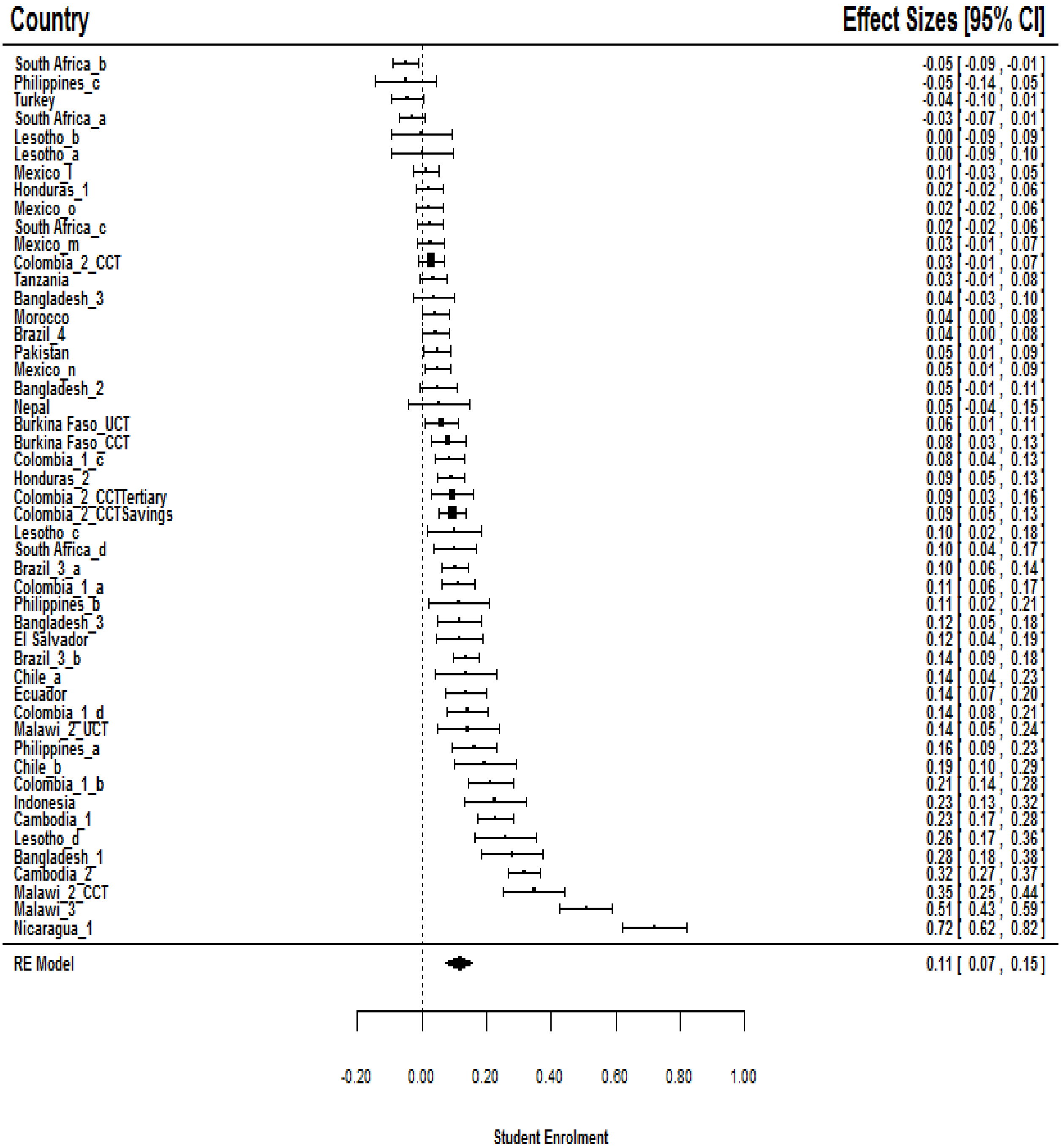
Effects of Cash transfer interventions on enrolment, attendance, dropout rates, completion and test scores

Enrolment

The overall average effect of cash transfer on enrolment, based on 48 comparisons, is 0.11, 95% CI [0.07, 0.15], calculated under random-effect model with robust variance estimation. The assessment of homogeneity suggests a large amount of between-studies variability (I^2 = 92.85%, Q(df = 48) = 671.7829, p = <0.0001) indicating the effects did not all arise from the same population.

Figure 5.4a supports the presence of heterogeneity. The effect sizes range from -0.05, 95% CI [-0.09, -0.01] in South Africa_b (Eyal and Woolard 2014) to 0.72, in Nicaragua_1 (Maluccio *et al.*, 2009), (95% CI, [0.62 0.82]). A handful of studies show zero or small negative effects, but the majority of studies suggest an improvement in enrolment. The results of sub-group analysis by gender suggest effects of slightly smaller magnitude for both groups (girls: SMD=0.07, 95% CI [0.03, 0.11]; boys: SMD=0.10, 95% CI [0.05, 0.16]), however the studies that report on gender sub-groups are likely not a representative sample and the results should therefore be interpreted with caution.

Figure 5.4 a: Enrolment



The moderator analysis does not suggest any statistically significant differences in effects due to the size of cash transfers or the presence and enforcement of conditions provided in the included studies. We do however find a statistically significant relationship between the size of the effect and programmes where parents or households are the recipients, but not when students or mothers are the recipients. The moderator analysis rely on few observations however and should therefore be interpreted with caution (see appendix H for full results).

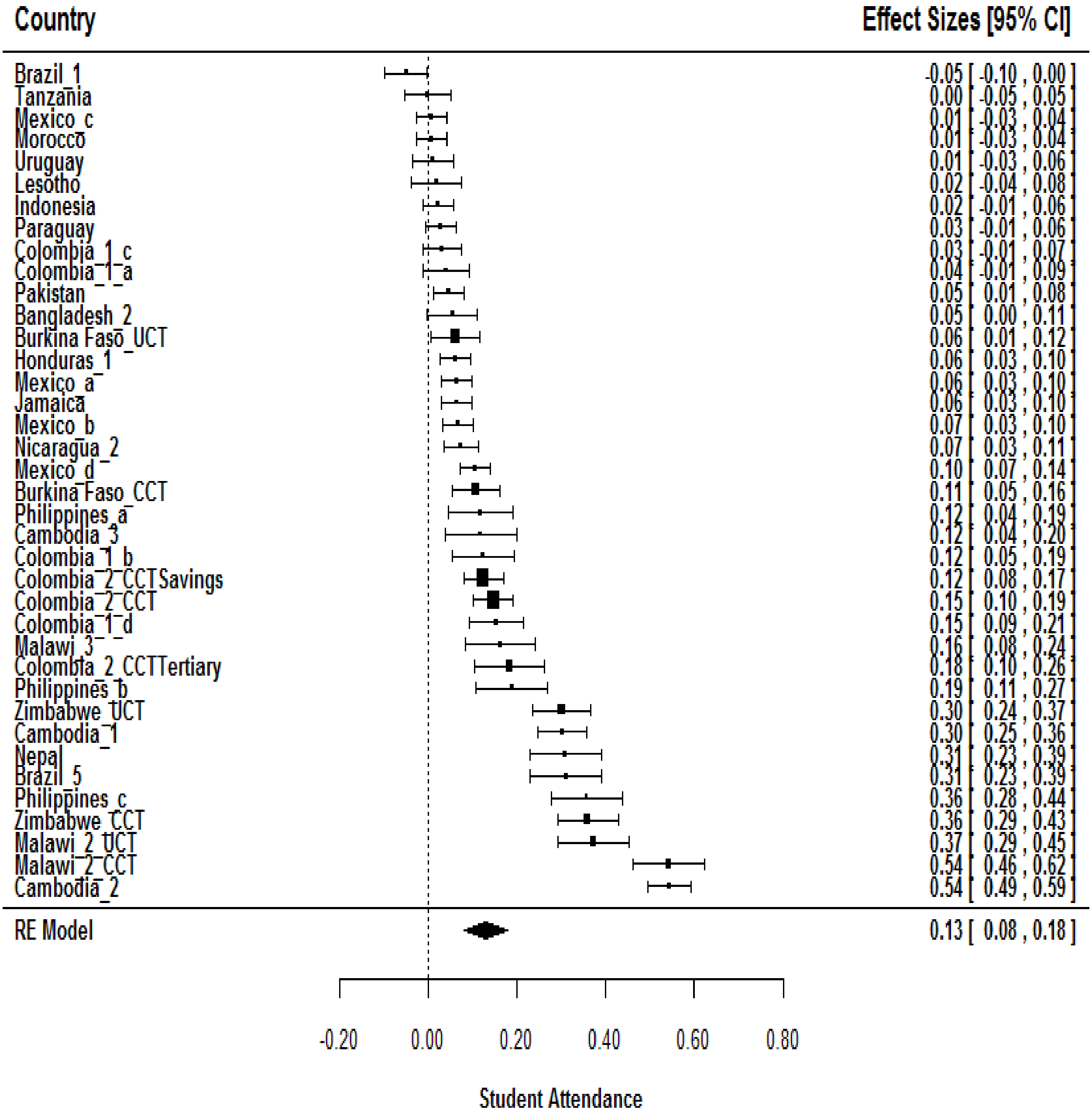
Attendance

The overall average effect of cash transfer on attendance is 0.13, 95% CI [0.08, 0.18], calculated under a random-effect model with robust variance estimation. The assessment of homogeneity suggests a large amount of between-studies variability (I^2 = 96.069%, Q (df = 37) = 941.3885, p = <0.0001), indicating the effects did not all arise from the same population.

Figure 5.4b supports the presence of heterogeneity. The effect sizes range from -0.05, 95% CI [-0.10, 0.00] in Brazil_1 (Kassouf, 2012) to 0.54, in Cambodia_2, (95% CI [0.49 0.59]) and 0.54 in Malawi_2_CCT (95% CI [0.46, 0.62) (Shady, 2014; Baird et al, 2010). As with enrolment, a handful of studies suggest small negative or zero effects, but the results from the majority of studies suggest a beneficial effect of cash transfers. As with enrolment the results of sub-group analysis by gender suggest effects of slightly smaller magnitude for both groups, but these estimates should be interpreted with caution.

We find a statistically significant relationship between the size of the effect and the intensity of conditions, suggesting a larger effect on attendance for programmes with tighter monitoring and enforcement of conditions. As with student enrolment we also observe a statistically significant relationship between the size of the effect and programmes where parents or households are the recipients, but not for students or mothers as recipients. The analysis does not indicate a statistically significant relationship between the magnitude of the effect and the size of the cash transfer. The moderator analysis rely on few observations however and should therefore be interpreted with caution.

Figure 5.4 b: Attendance

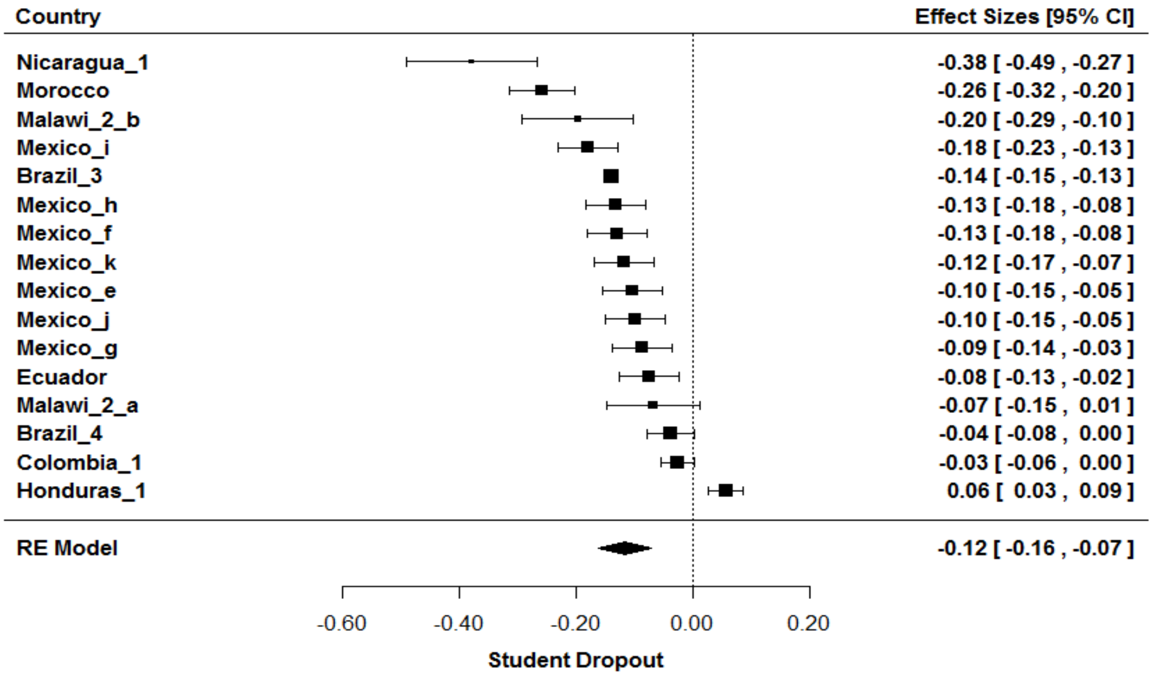


Dropout

The overall average effect of cash transfer on dropout is -0.12, 95% CI [-0.16, -0.07]. The assessment of homogeneity suggests a large amount of between-studies variability (I^2 = 92.70%, Q (df =15) = 192.1435, p = < .0001) indicating the effects did not all arise from the same population.

The effect sizes range from -0.38, 95% CI [-0.49, -0.27] in Nicaragua_1 (Maluccio *et al.*, 2009) to 0.06, in Honduras_1, (95% CI [0.03, 0.09]) (Glewwe and Olinto 2004), and Figure 5.2b supports the presence of heterogeneity. While one study suggests an increase in dropout rates, the majority of the studies find a reduction in dropout rates. The results are robust to the removal of any single study. The results from the studies with sub-group analysis by gender suggest slightly larger overall effects, especially for girls. The estimates are however less precise and not from a representative sample of studies.

Figure 5.4 c: Dropout

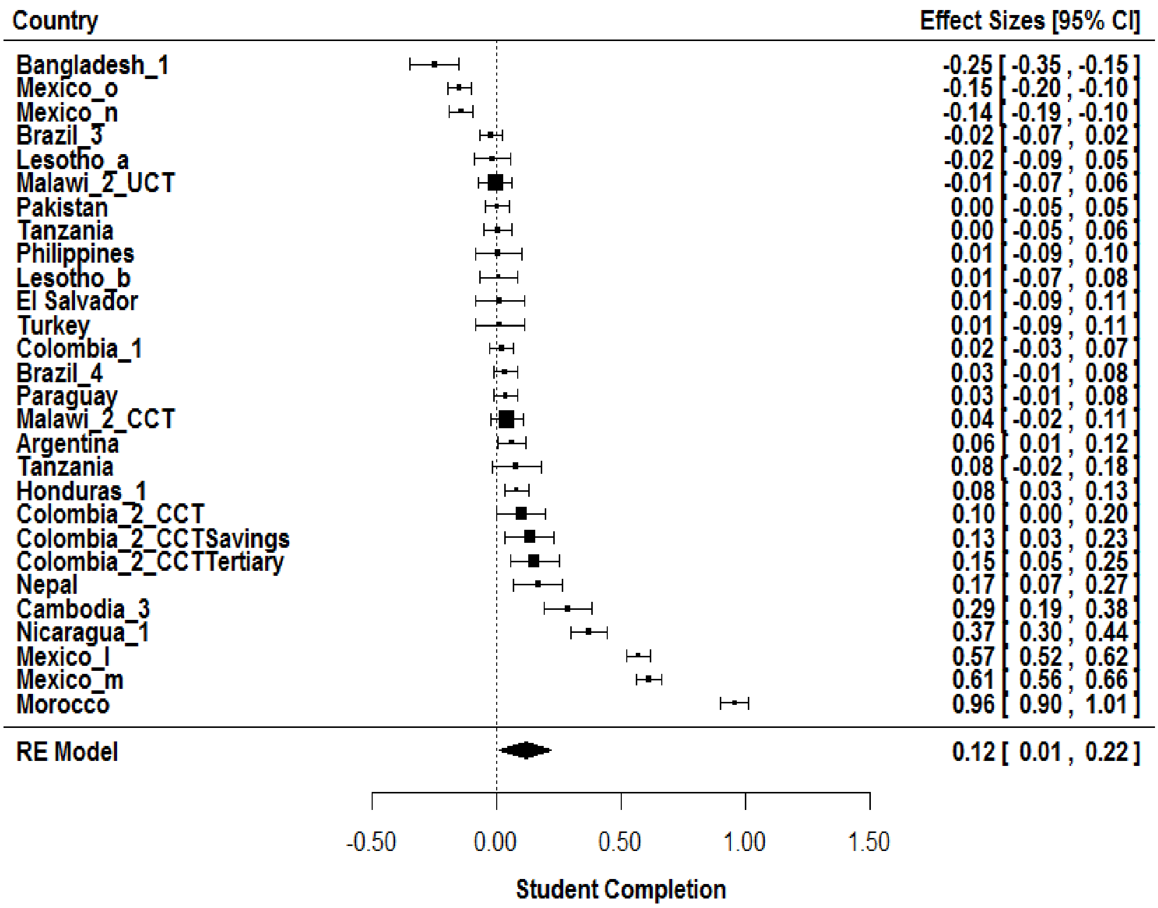


Completion

We include 28 effect estimates from 20 different programmes in the meta-analysis of student completion outcomes. The overall average effect of cash transfers on student completion is 0.12, 95% CI [0.01, 0.22], calculated under a random-effect model with robust variance estimation.

There is a large amount of between study variability, as indicated by the tests of homogeneity (I^2 = 98.72%, Q(df = 27) = 2118.8108, p-value < .0001). This is supported by the forest plot in figure 5.4d. The effect sizes range from -0.25 SMD, 95% CI [-0.35,-0.15] for the Primary Education Stipend (PES) programme in Bangladesh up to 0.96 SMD, 95% CI [0.90, 1.01] for the Tayssir cash transfer programme in Morocco. There are several clusters of studies without overlapping confidence intervals. This indicates that in some contexts cash transfers have had adverse effects on completion rates, in others zero or small effects, and yet in others large beneficial effects.

Figure 5.4 d: Student Completion

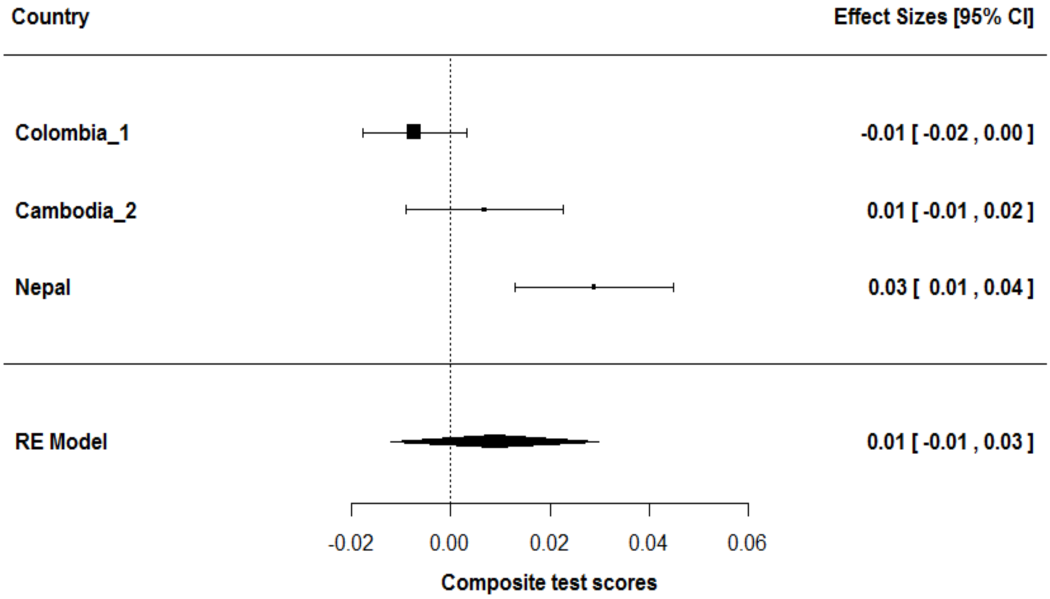


We were able to undertake sub-group analyses for boys and girls for student completion, although we only identified eight effect sizes for each of seven different programmes and so the results are not representative of the full sample of programmes. The results for boys were similar, but less precise than the results for the full sample (SMD = 0.11, 95% CI [-0.12, 0.33). The results for girls were larger in magnitude than that of the main sample (0.18 SMD, 95% CI [-0.03,0.40]), although the confidence interval is crossing the line of no effect. As for the main sample heterogeneity is high ($I^{2=}$ 98.19%, (Q(df = 7) = 492.1081, p-value= < .0001). The results are sensitive to the inclusion of the two observations from Mexico (Mexico_b: completion for secondary school girls and Mexico_a: completion for primary school girls). When we remove Mexico_a the average effect for girls is reduced (SMD= 0.09, 95% CI [-0.05, 0.22]), while when we remove Mexico_b the average effect is larger and more precise (SMD=0.24, 95% CI [0.03, 0.45]).

Composite Test Scores

The overall average effect of CT on composite test scores is 0.01, 95% CI [-0.01, 0.03] under random-effects model. The assessment of homogeneity suggests a large amount of between-studies variability (I^2 = 84.99%, τ^2 = 0.02, Q (df = 2) = 13.95, p = 0.0009) indicating the effect did not all arise from the same population. As can be seen from figure 5.4e the results range from -0.01 (95% CI [-0.02, 0.00]) to 0.03 (95% CI [0.01, 0.04]).

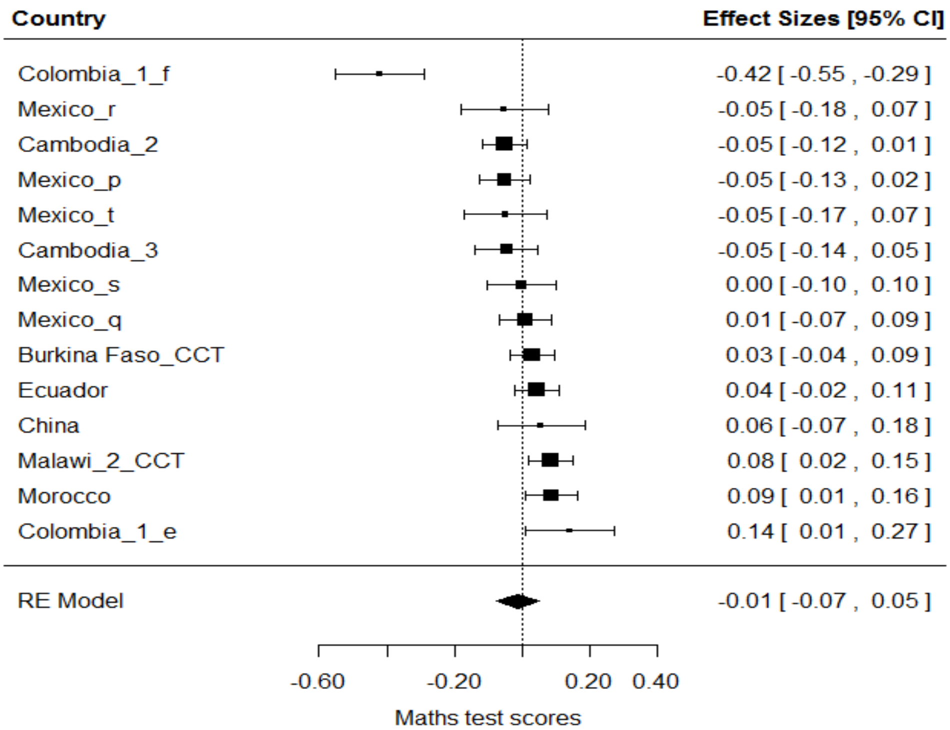
Figure 5.4 e: Composite Test Scores



Math Test Scores

The overall average effect of cash transfers on maths test scores is -0.01, 95% CI [0.07, 0.05]. The assessment of homogeneity suggests considerable between-studies variability (I^2 = 86.43%, τ^2 = 0.0118, Q (df = 13) = 65.5800, p = 0.0001), indicating the effects did not all arise from the same population. Figure 5.2f supports the presence of heterogeneity. The results are sensitive to the inclusion of one of the estimates from Colombia_1_f (Garcia and Hill 2009). Removing this study changes the direction of the overall effect, although it remains statistically insignificant (SMD = 0.02, 95% CI [-0.01, 0.01]) (see Appendix H for results of all sensitivity analyses).

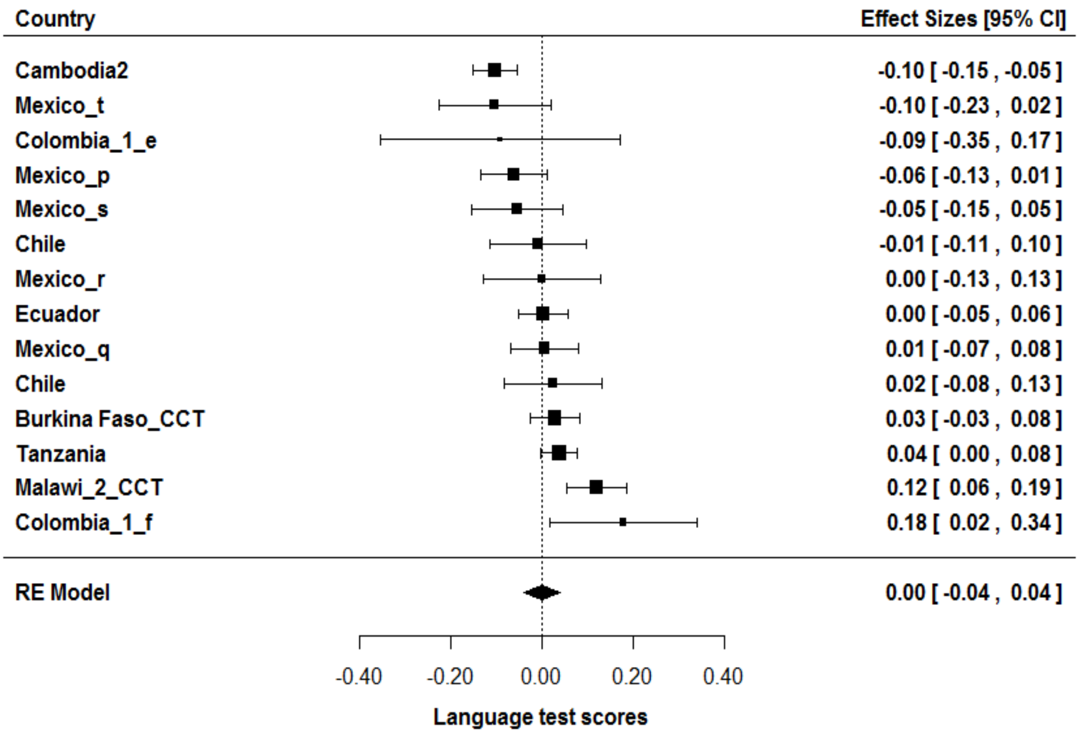
Figure 5.4 f: Maths Test Scores



Language Arts

The overall average effect of cash transfers on language arts test scores is zero, 95% CI [-0.04, 0.04], as calculated under a random-effects model. The assessment of homogeneity suggests a significant amount of between-studies variability (I^2= 72.19%, τ^2 = 0.0039, Q (df = 13) = 46.8803, p = 0.0001) indicating the effect did not all arise from the same population. Figure 5.2g supports the presence of heterogeneity. The effects range from -0.10 (95% CI [-0.15, -0.06]) in Cambodia, to 0.18 (95% CI [0.02, 0.34]) in one of the sub-samples in Colombia. There are several clusters of studies without overlapping confidence intervals, indicating both beneficial and adverse effects have been observed in different contexts. Nevertheless, the overall effect is not sensitive to the inclusion of any one study.

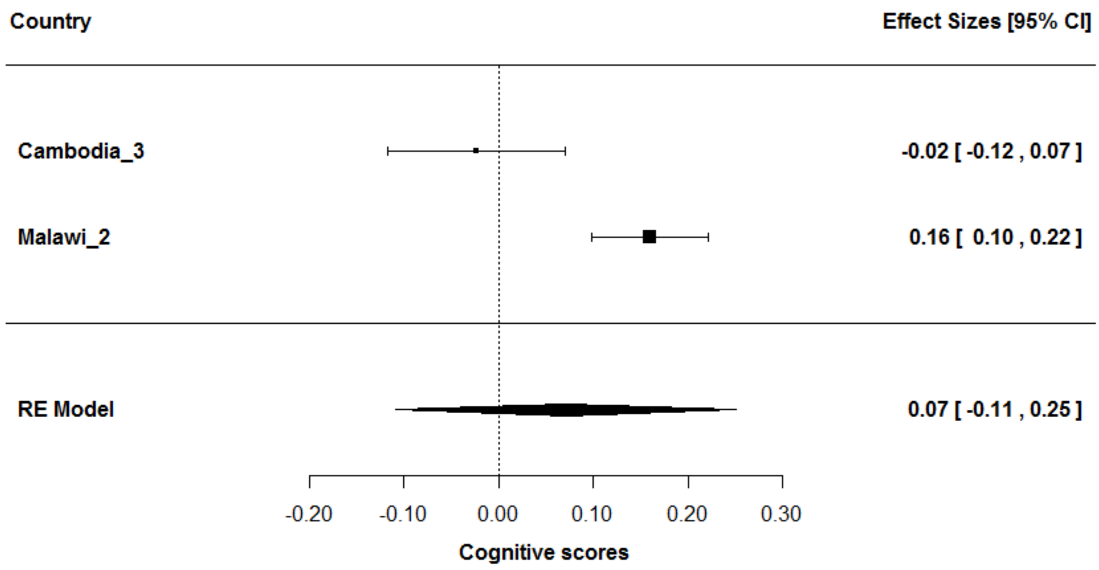
Figure 5.4 g: Language Arts Test Scores



Cognitive Test Scores

Only two studies reported on the effect of CT programmes on cognitive scores. The overall average effect on cognitive scores is 0.07, 95% CI [-0.11, 0.25], calculated under a random effect model. The assessment of homogeneity suggest a large amount of variation between these two studies (I^2 = 90.37%, τ^2 = 0.0152, Q (df = 1) = 10.38, p = 0.0013) and this is also indicated graphically in the forest plot in figure figure 5.4h.

Figure 5.4 h: Cognitive Test Scores



5.2.4 Summary of findings and discussion

We included 50 studies that evaluated the effect of 38 unique programmes in Latin America and the Caribbean, Sub-Saharan Africa, East Asia and Pacific, the Middle East and North Africa and Europe and Central Asia. The results suggest consistently positive effects of cash transfers across all school participation outcomes. Effects range from 0.11, 95% CI [0.07, 0.15] for enrolment, where we included 48 effect sizes, to 0.13, 95% CI [0.08, 0.18] for attendance, where we included 38 effect sizes. The average effect for dropout also suggests a reduction in dropout rates (-0.12, 95% CI [-0.16, -0.07]) and we also see an average improvement in school progression and completion rates (0.12, 95% CI [0.01, 0.22]). The results of the moderator analysis suggest the intensity of conditions is associated with an increase in the magnitude of effects on attendance, but not enrolment. Programmes where the receipient of the cash transfer is the household, rather than mothers or students, are also associated with a larger effect.

Cash transfers do not appear to improve student learning outcomes however. The effect on maths test scores is 0.01, 95% CI [-0.07, 0.05] and no different from zero for language arts test scores (0.00, 95% CI [-0.04, 0.04]). However, these findings are based on fewer studies than the results on school participation and should therefore be interpreted with some caution. The overall results are robust, but there is considerable heterogeneity and we observe effects that are both substantially larger and smaller than the average pooled effects.

Table 5.4 a: Characteristics of included studies for Cash Transfers

Study ID and Country ID	Setting	Intervention summary	Included outcomes	Follow- Up*	Study design	Sample Size
Filmer and Schady, 2009, 2011, 2014 **Cambodia_2**	Cambodia, lower secondary schools. Students in 7th, 8th and 9th grade.	CESSP Scholarship Programme (CSP): Primary “feeder” schools were mapped to each CSP-eligible secondary schools. Within the primary feeder schools, all students in the last year of primary school, filled out an application form for the CSP scholarship programme. In large CSP schools (enrolment above 200) 50 students were offered a scholarship for 7th, 8th, and 9th grade; in “small” CSP schools (total enrolment below 200 students), 30 students were offered the scholarship. Two-thirds of the scholarships were given to girls. The amount of the CT depended on the dropout-risk score that the child would get. There is no monitoring on how the money is spent.	Attendance Enrolment Learning: language arts, maths and composite	Between 18-60 months	RDD	3225 Participants
Zavakou ND **Colombia_1**	Colombia, urban and rural, primary and secondary schools.	The Familias en Acción has two main components: education and nutrition. Education: The largest component of the programme is the education one, which is targeted at families with children aged 7 to 17. The subsidy doubles for secondary school in comparison to primary. The subsidy is provided regardless on the number of children per household. Nutrition: A flat-rate monthly monetary supplement is provided to mothers of all beneficiary families with children aged 0 through to 6.	Completion Dropout	48 months	CBA	9,415 Participants
Garcia and Hill, 2009 **Colombia_1**	Colombia, urban and rural, primary and secondary schools. Student’s average age is 10.	The Familias en Acción has two main components: education and nutrition. Education: The largest component of the programme is the education one, which is targeted at families with children aged 7 to 17. The subsidy doubles for	Learning: Maths and language arts	12 months	CBA	237 Participants

		secondary school in comparison to primary. The subsidy is provided regardless on the number of children per household. Nutrition: A flat-rate monthly monetary supplement is provided to mothers of all beneficiary families with children aged 0 through to 6.				
Baez and Camacho, 2011 Colombia_1	Colombia, urban and rural, primary and secondary schools. Student's average age is 12.	The Familias en Acción has two main components: education and nutrition. Education: The largest component of the programme is the education one, which is targeted at families with children aged 7 to 17. The subsidy doubles for secondary school in comparison to primary. The subsidy is provided regardless on the number of children per household. Nutrition: A flat-rate monthly monetary supplement is provided to mothers of all beneficiary families with children aged 0 through to 6	Completion Learning: Maths, language arts and composite	up to 108 months	RDD	624,028 Participants
Attanasio and Gómez, 2004; Attanasio *et al.,* 2006 Colombia_1	Colombia, urban and rural, primary and secondary schools. Student's average age is 11.	The Familias en Acción has two main components: education and nutrition. Education: The largest component of the programme is the education one, which is targeted at families with children aged 7 to 17. The subsidy doubles for secondary school in comparison to primary. The subsidy is provided regardless on the number of children per household. Nutrition: A flat-rate monthly monetary supplement is provided to mothers of all beneficiary families with children aged 0 through till 6.	Attendance Enrolment	12 months	Natural Experiment	4,718 Participants
Ferré and Sharif, 2014 Bangladesh_2	Bangladesh, rural and urban primary schools. Students age range is 6-15 years.	Shombhob project: Households eligible one of three types of CCT dependant on their demographic composition. (1) a per household transfer if the household included one or more children aged 0 to 36 months (2) a per household transfer if eligible families had one or more children going to primary school (3) a per household transfer for households with children under 36 months and children going to primary school.	Attendance Enrolment	15 months	CBA	2348 Households

		Payments made using electronic transfers to accounts in post-offices. To ensure that households knew conditions, each household received a booklet detailing the purpose of the project and its timeline, eligibility criteria, beneficiary conditions, and the payments and grievance redress mechanisms. A campaign to advertise the CCT was done. Participants needed to apply and when they were selected (on poverty data) they were required to enrol in the project.				
Benhassine et al., 2013 Morocco	Morocco, rural primary schools. Grade rage of students is 1-6, their average age is 9.	Tayssir: Small cash transfer made for school-aged (6-15) children. Four treatment groups: Unconditional; Conditional on school attendance; Father beneficiary; Mother beneficiary. Parents had to enrol students into the programme. Students enrolled in the first year were automatically enroled in the second year. It was made clear to parents that the transfers were from the ministry of education and part of an effort to promote education. Advertisements included flyers given to schools with children sitting at desks with education slogans. Post Office distribution: recipients received transfer at post-offices, otherwise a mobile cashier was used. All transfers could be withdrawn at once if recipient wanted to wait until end of the year.	Dropout Enrolment Learning: Maths Attendance	24 months	Cluster RCT	5998 Participants
Baird et al., 2010 Malawi_2	Malawi, urban and rural secondary schools. Student average age: 15.	Zomba Cash Transfer Programme: CCT and UCT groups. CCT: Made to households on once a month. School fees were also paid in full directly to the school. Fee payment provided to public schools and up to the cost of attending public schools for private schools. If the condition of attendance was not met each month the following month was withheld. UCT: Identical to conditional programme with no conditions. The cash payment points were chosen to take place at centrally	Learning: Maths Learning: English language test Learning: Cognitive score Attendance	Dropout: 12 months Enrolment: self reported: 24 months Enrolment: teacher reported: 28 months Attendance: register: 28 months Attendance: Average between 3 measures: between 12-23 months	Cluster RCT	2057 Participants

		located and well-known places, such as churches or schools.	Dropout Enrolment	Cognitive Test Score: 24 months Learning: language arts: 24 months Learning: maths: 24 months		
Macours and Vakis, 2009 Nicaragua_2	Nicaragua, rural schools. Age range of students is 7 to 18.	Atención a Crisis: Three arms: Basic CCT; CCT plus scholarship for occupational training; CCT plus grant for productive investments. All selected beneficiary households received the basic CCT component. One third of the beneficiary households also received a scholarship that allowed one household member to choose from a number of vocational training courses. One third of the beneficiary households also received a grant for productive investments aimed at encouraging recipients to start a small non-agricultural business activity. This package included technical assistance and training in basic commercial skills. Beneficiary women were chosen to act as leaders. Their role was to provide information, monitoring and motivation to ensure compliance with various programme requirements and conditionalities.	Attendance	9 months	Cluster RCT	5176 Participants
Heinrich, 2005 Argentina	Argentina, urban and rural, secondary schools. Average student age is 14. Student's grade ranges between 8 and 9.	Programa Nacional de Becas Estudiantiles: a conditional cash transfer programme that targets youth ages 13-19 years, entering their 8th and 9th years of study in public schools and are at risk of leaving school. Students must be from households that have a monthly income of less than 500 pesos and do not receive any other benefit.	Completion	12-48 months depending on outcome	CBA	2586 Participants
Glewwe and Kassouf, 2012 Brazil_3	Brazil, rural, urban and peri urban primary and secondary schools.	Bolsa Escola: Transfer given to mothers of households with a monthly per capita income of less than 90 reais (around US$40) with children aged 6-15. The programme is implemented by municipalities, who are	Enrolment	Glewwe and Kassouf, 2012: 60 months	CBA	699,255 Participants

	Students grade ranges between 1 and 8.	responsible for identifying beneficiaries and monitoring and enforcing conditionalities. In 2004, the programme was incorporated in the current Bolsa Familia programme.				
de Janvry *et al.*, 2006 and 2012 Brazil_3	Brazil, rural, urban and peri urban primary and secondary schools.	Bolsa Escola: Transfer given to mothers of households with a monthly per capita income of less than 90 reais (around US$40) with children aged 6-15. The programme is implemented by municipalities, who are responsible for identifying beneficiaries and monitoring and enforcing conditionalities. In 2004 the programme was incorporated in the current Bolsa Familia program.	Completion Dropout	De Janvry, 2006 & 2012: 36 months	CBA	182,191 Participants
de Brauw *et al.*, 2014 Brazil_4	Brazil, rural, urban and peri urban primary and secondary schools. Students age ranges from 6 to 17. Students grades ranges between 1 and 11.	Bolsa Familia: a CT composed of three different transfers, one of which is unconditional. (1) A conditional variable payment per child aged 0 to 15 years, for up to three children, to "poor" households below a per capita income threshold. (2) An additional unconditional transfer to "extremely poor" households below a lower per capita income threshold. (3) BVJ programme added payments and a schooling conditionality for children aged 16 and 17. Any household in Bolsa Família with children aged 16 or 17 was automatically part of BVJ.	Enrolment Completion Dropout	Approximately 48 months	RDD	6507 Participants
Benedetti *et al.*, 2015 Honduras_2	Honduras, rural and peri-uban primary and secondary schools. Average age of students is 8.7, in grades 1 to 9.	Bono 10,000: A transfer of 10,000 made to households if they had at least one 6-18 year old but had not completed 9th grade, 5,000 was given for households that had children under six, but only in absence of children aged 6-18. Transfers were made by programme staff to households.	Enrolment	21 months	CBA	6573 Households

Ponce and Bedi, 2008 Ecuador	Ecuador, rural and urban primary and secondary schools.	Bono de Desarrollo Humano: Unconditional transfer that provides a flat amount per household per month. There was a social marketing campaign in radio and newspapers that stressed the importance of the BDH. Benefit distribution started without commitments on the length of benefits or when eligibility would be reassessed. To receive the transfer eligible persons had to visit any branch office of largest network of private banks or the National development Bank on pre-arranged days. They could also arrange to have it transferred monthly to their account.	Learning: Maths and language arts	Approx. 18 months	RDD	2589 Participants
Edmonds and Schady, 2011 Ecuador	Ecuador, rural and urban, primary and secondary schools. Students average age 13.	Bono de Desarrollo Humano: Unconditional transfer that provides a flat amount per household per month. There was a social marketing campaign in radio and newspapers that stressed the importance of the BDH. Benefit distribution started without commitments on the length of benefits or when eligibility would be reassessed. To receive the transfer eligible persons had to visit any branch office of largest network of private banks or the National development Bank on pre-arranged days. They could also arrange to have it transferred monthly to their account.	Dropout Enrolment	Enrolment: Approx. 18 months dropout: 36 months	RCT	2876 Participants
Barrera-Osorio and Filmer 2014 Cambodia_3	Cambodia, Primary schools. Grades 4 to 6.	Cambodian Primary Scholarships Pilot: offered two types of cash transfer, one on merit, the other on socio-economic status. The socio-economic status is evaluated here. There is moderate enforcement of the conditionality. Students absent for many days are followed up by school officials and if they return to school would remain eligible for the scholarship. If a student is absent for too many days they would be classified as having dropped out and no longer be eligible for the scholarship.	Attendance Completion Learning: Maths Learning: Cognitive test scores	24 months	Cluster RCT	883 Participants
Pellerano *et al.*, 2014 Lesotho	Lesotho, urban and rural primary and secondary	Child Grants Programme: unconditional transfer provided a regular transfer of between M360 and	Enrolment	24 months	Cluster RCT	2,396 participants

	schools. Average age of students 9.5.	M7506 every quarter households with children aged 0-17. At the start of the programme there was a monthly flat rate per child of M120, as of April 2013 this changed to M360 for one to two children, M600 three to four children, M750 for five or more children. The person who will collect the grant can be male or female, usually the head of the household or a parent. Although theoretically a UCT, the recipients receive effective messaging that the transfer should be spent on children.	Attendance Completion			
Santana, 2008 South-Africa	South-Africa, rural and urban, primary and secondary schools.	Child Support Grant (CSG): Socio-economically targeted child assistance grant. The grant was 100 rands per child (approximately US$17) when the programme was initially introduced and has increased to 200 rands in 2007. The minimum age for child eligibility has increased, in 2003 children under nine were eligible, in 2004 it became children under 11 years and in 2005 children under 14.	Enrolment	48 months	Natural experiment	10,279 participants
Eyal and Woolard, 2014 South-Africa	South-Africa, rural and urban primary and secondary schools. Students age range between 7 and 13.	Child Support Grant (CSG): Socio-economically targeted child assistance grant. The grant was 100 rands per child (approximately US$17) when the programme was initially introduced and has increased to 200 rands for in 2007. The minimum age for child eligibility has increased, in 2003 children under 9 were eligible, and in 2004 it became children under 11 years and in 2005 children under 14.	Enrolment	132 months	CBA	1,826 participants
Galasso, 2006 Chile	Chile, rural and urban, primary and secondary schools.	Chile Solidario: A two part programme. The first provides a social worker to households in extreme poverty. Households sign a contract committing them to working with the social worker on 53 social, economic and legal issues. The contract is for two years, the cash transfer reduces in size every six months over this period. After two years, households are ensured a direct cash transfer and preferential access to assistance programmes for an additional	Enrolment Attendance Learning: language arts	Two follow up periods: 12 and 24 months	CBA	693 Households

		period of three years. The second part works on the supply side by ensuring coordination among different programmes. It aims to provide bundles of programmes that are tailored to meet the specific needs of households that are hard to reach.				
Barrera-Osorio *et al.*, 2011 Colombia_2	Colombia, urban secondary schools. Average student age, 15. Grade range between 6 and 11.	Subsidios Condicionados a la Asistencia Escolar: programme had three treatment arms called Basic, Savings and Tertiary. Basic intervention: participants receive 30,000 pesos per month as long as conditions are met. Saving treatment: varied the timing of the distributions to students' families. Students were paid 20,000 pesos on a bi-monthly basis, while the remaining 10,000 pesos was held in a bank account. The accumulated funds were made available to students' families during the period in which students prepare to enrol for the next school year. If the conditions were met this treatment would make 100,000 pesos available to them in December. Tertiary treatment: changes the outcomes upon which students are being incentivized. This treatment also provides an incentive to graduate and matriculate to a higher education institution. The monthly transfer is 20,000 pesos and there is no remaining third being held in an account. However, upon graduating the student earns the right to receive a transfer of 600,000 pesos, amounting to 73 per cent of the average cost of the first year at a vocational school. If the student enrols in a tertiary institution, they receive the transfer immediately; if they fail to enrol, they receive the transfer after a year has passed.	Attendance Enrolment Completion	Attendance: Approximately 6-8 months All others: 12 months	Cluster RCT	7721.66 Participants
Khandker *et al.*, 2003 Bangladesh_3	Bangladesh, rural secondary schools.	Female Secondary Stipend Programme: introduced a uniform stipend and tuition subsidy programme for each girl attending a secondary school in rural areas who meets the conditionalities. When in the	Enrolment	Not clear.	CBA	1967 Participants

		programme, all female students meeting conditions receive a specified amount of stipend and other allowances as prescribed for each grade. The girl's school is directly paid all of her tuition by the project. The stipend was expected to cover as much as 50 per cent of the costs of textbooks, uniforms, stationary, transportation, exam fees, and miscellaneous direct educational expenses. The stipend directly to an account in the girl's name. Other components include: curriculum reform, instructional materials development, teacher training, recruitment female teachers, improved infrastructure, awareness programmes at community level and institutional building capacity.				
Fusades, 2009 **El Salvador**	El Salvador, urban and rural primary schools. Age range of students between 7 and 12 years old.	Red Solidaria: operates in three strategic areas (1) (Support to family) conditional cash transfers, (2) Basic services, infrastructure of education, health and nutrition services; (3) financing of micro-credit business projects as a tool to support small farmers. Cash Transfer: Households with children five years of age and under were targeted for a health transfer. Households with children between the ages of six and 15 who had not completed primary school were targeted for an education transfer. From 2009, a pension was given to all eligible elderly (70+ years) living in the targeted areas.	Completion	12 months	RDD	752 municipalities
De Brauw and Gilligan, 2011 **El Salvador**	El Salvador, urban and rural primary schools. Student age range between 7 and 12.	Red Solidaria: operates in three strategic areas (1) (Support to family) conditional cash transfers, (2) Basic services, infrastructure of education, health and nutrition services; (3) financing of micro-credit business projects as a tool to support small farmers. Cash Transfer: Households with children five years old and under were targeted for a health transfer. Households with children between the ages of 6 and 15 who had not completed primary school were targeted	Enrolment	12 months	RDD	1534 Participants

		for an education transfer. From 2009, a pension was given to all eligible elderly (70+ years) living in the targeted areas.				
Sparrow, 2007 Indonesia	Indonesia, rural, urban and peri-urban primary and secondary (junior and higher) schools. Student age range from 10 to 18 years. All grades except 1 to 3.	Jaring Pengaman Sosial (JPS): Scholarships were given to primary and secondary public/private school students and the amount varied according to the HH income and to the enrolment level. It goes from 7 to 18 per cent of average per capita HH consumption. The scholarships were monthly transfers and students had full discretion on how to use the funds. The JPS programme included also budgetary support schools.	Enrolment Attendance	12 months approx.	Natural Experiment	113,187 participants
Luseno, 2013 Malawi_1	Malawi, secondary school and primary schools. Average student age, 11.	Malawi Social Cash-Transfer Scheme (SCTS): beneficiaries of this programme receive payments and are told that children should attend school, but there is no monitoring or penalties. The value of the transfer depends on the size of the household and the number of school-aged children.	Enrollment Attendance	12 months	Cluster RCT	1184 participants
Evans *et al.,* 2014 Tanzania	Tanzania, rural primary school.	No specific name: education and health Community-based CCT scheme that provides cash to families with children and elderly. The Education transfer targeted at children 7–15 years old. Conditionalities are monitored. The Health transfer is targeted at children 0–5 years old and elderly (60+). Conditionality is monitored. If beneficiaries failed to comply in the first period of monitoring they were issued with a warning, if this continued to the second period, the payments were reduced by 25 per cent and a second warning issued. The payments were stopped with the third warning, participants could return to the programme with a review from the community's programme implementer.	Enrolment Attendance completion Learning: language arts	24 months	Cluster RCT	4823 Households
Amarante *et al.*, 2011 Uruguay	Uruguay, rural and urban secondary school and primary schools. Age	Plan Nacional de Atencion a la Emergencia Social: Two main components: (1) Cash Transfer, which was found to correspond to approximately 50 per cent of the average pre-programme HH self-reported income. (2)	Attendance	24-36 months	RDD	3621 participants

	range of students between 6 and 17.	HHs with children or pregnant women were also entitled to a food card, an in-kind transfer that operated through an electronic debit card. Although initially conditional, conditionalities were not enforced. (3) Additional but smaller components included a workfare programme, job training, adult educational interventions and health care subsidies.				
Chaudhury *et al.*, *2*013 Philippines	Philippines, rural and peri-urban secondary school and primary schools.	Pantawid Pamilya: two conditional cash transfers provided to poor households every two months, the cash amount depends on the number of eligible children. (a) Education Grants. The education grant is aimed at children 6-14 years. The transfer is given per child, up to a maximum of three children for a period of 10 months/year. b) Health Grants. The health grant is aimed at children 0-14 years old and/or pregnant women. Households receive a lump sum per month.	Enrolment Attendance Completion	30 months	Cluster RCT	1570 Participants
Glewwe and Olinto, 2004 Honduras_1	Honduras, rural primary schools. Average student age is 8.	PRAF II: Two conditional components HEALTH: monetary transfers to pregnant women and to mothers of children under three years of age. A voucher is provided to women. Families may receive up to two vouchers per month (one woman and one child, or up to two children). Monetary transfers to primary health care teams, formed of community members and health care workers (nurses and, when available, doctors). EDUCATION: conditional payments to families for each child age 6-12. US$5 is provided per month for each eligible child up to a maximum of three children per family (in addition to the health transfer). Payments to the Parent Teacher Associations (PTAs) associated with each primary school. PTA's obtained legal status and to prepared plans to improve the quality of the	Attendance Enrolment Dropout Completion	Glewwe: 12 and 24 months	Cluster RCT	12,065 Participants

		education with a budget for the educational materials and equipment (selected from items approved by PRAF II).				
Baulch, 2011 Bangladesh_1	Bangladesh, rural and urban primary schools. Student age range between 6 and 12.	Primary Education Stipend (PES) Programme: households with receive payments per month based on the number of children; BDT100 per month for one child and BDT125 for more than one child. Stipends were disbursed at local bank branches temporary distribution points established within 5 kilometres of the school. Only Students of rural government and NGO (those that provide full range of grades) run primary schools eligible. The transfer had education conditions for the student and the school.	Completion Enrolment	72 months	CBA	1534 Households
Pianto and Soares, 2004 Brazil_5	Brazil, urban and rural, primary and secondary schools. Student age range between 7 and 15.	Programa de Erradicacao do Trabalho Infantil (PETI): Transfers to families with school-aged children (7 to 14) conditional on school attendance and their participation in a sub-programme of after-school activities. Transfer amounts are higher in urban areas than in the countryside, vary from one state to another and sometimes depend on the number of children per family. The programme has a psychosocial component whose aim is to alter the beliefs and attitudes regarding the education and child labour.	Attendance	24 months	CBA	72 Municipalities
Levy and Ohls, 2007 Jamaica	Jamaica, urban, rural and peri-urban primary and secondary schools. Student age range between 6 and 17.	Programme of Advancement Through Health and Education: Health grants provided to children aged 0-6 and education grants are provided to children aged 6-17. Social assistance grants provided poor adults that are pregnant or breastfeeding, elderly (over age 65), disabled or homeless. Households were required to apply to be eligible for the programme. Recipients could benefit from another programme, a government waiver of education and health fees. Including annual tuition fee that students must pay, and the cost of a visit to a health centre.	Attendance	Approx. 24-29 months	RDD	6,790 Participants

Behrman *et al.*, 2000 Mexico	Mexico, rural primary and secondary schools.	Progresa: Grants were provided for children enrolled in grades 3 through till 9, or the last four years of elementary school and the next three years of junior secondary school. The size of the grants increase several fold at the higher grades. A premium for girls was introduced in junior secondary school. Every 6 months, the grants were adjusted upward to compensate for inflation. There are additional supports for eligible families. A transfer payment for school materials was paid to the mother for each school term in which her child is enrolled in the programme-subsidized grades. Finally, a ''food'' transfer is provided the mother. Pregnant and breastfeeding women, and children under two years of age were given nutritional supplement, as were other young children who were not growing at an acceptable rate.	Learning: Maths and Language arts	6 and 18months	Cluster RCT	1617 Participants
Dubois *et al.*, 2011 Mexico	Mexico, rural primary and secondary schools.	Progresa: Grants were provided for children enrolled in grades 3 through 9, or the last four years of elementary school and the next three years of junior secondary school. The size of the grants increased several fold at the higher grades. A premium for girls was introduced in junior secondary school. Every six months, the grants were adjusted upward to compensate for inflation. There are additional supports for eligible families. A transfer payment for school materials was paid to the mother for each school term in which her child is enrolled in the programme-subsidized grades. Finally, a ''food'' transfer is provided the mother. Pregnant and breastfeeding women, and children under two years of age were given nutritional supplement, as were other young children who were not growing at an acceptable rate.	Completion	6 months	Cluster RCT	1739 Participants
Raymond and Sadoulet, 2003 Mexico	Mexico, rural primary and secondary schools.	Progresa: grants were provided for children enrolled in grades 3 through 9, or the last four years of elementary school and the next 3 years of junior secondary school.	Dropout	Approximately 48 months	Cluster RCT	2934 Participants

		The size of the grants increased several fold at the higher grades. A premium for girls was introduced in junior secondary school. Every six months, the grants were adjusted upward to compensate for inflation. There are additional supports for eligible families. A transfer payment for school materials was paid to the mother for each school term in which her child is enrolled in the programme-subsidized grades. Finally, a ''food'' transfer is provided the mother. Pregnant and breastfeeding women, and children under two years of age were given nutritional supplement, as were other young children who were not growing at an acceptable rate.				
Skuffious and Parker, 2011 Mexico	Mexico, rural primary and secondary schools.	Progresa: grants were provided for children enrolled in grades 3 through 9, or the last four years of elementary school and the next three years of junior secondary school. The size of the grants increased several fold at the higher grades. A premium for girls was introduced in junior secondary school. Every six months, the grants were adjusted upward to compensate for inflation. There are additional supports for eligible families. A transfer payment for school materials was paid to the mother for each school term in which her child is enrolled in the programme-subsidized grades. Finally, a ''food'' transfer is provided the mother. Pregnant and breastfeeding women, and children under two years of age were given nutritional supplement, as were other young children who were not growing at an acceptable rate.	Attendance	Skuffious, 2011: Attendance: 6, 12 and 18 months	Cluster RCT	7182 Participants
Schultz, 2004 Mexico	Mexico, rural primary and secondary schools. Student age range between 5 and 16. Student grades between 3 and 9.	Progresa: grants were provided for children enrolled in grades 3 through 9, or the last four years of elementary school and the next three years of junior secondary school. The size of the grants increased several fold at the higher grades. A premium for girls was introduced in junior secondary school. Every six months, the grants were adjusted upward to compensate for	Schultz, 2004: Enrolment	Schultz, 2004: Up to 36 months	Cluster RCT	55396 Participants

		inflation. There are additional supports for eligible families. A transfer payment for school materials was paid to the mother for each school term in which her child is enrolled in the programme-subsidized grades. Finally, a ''food'' transfer is provided the mother. Pregnant and breastfeeding women, and children under two years of age were given nutritional supplement, as were other young children who were not growing at an acceptable rate.				
Maluccio *et al.*, 2009 Nicaragua_1	Nicaragua, rural primary schools. Student age range between 5 and 13.	Red de Proteccion Social: Two conditional cash transfers provided to eligible households. Food Security and Nutrition Transfer and an Education Transfer. Both were bimonthly fixed per household transfer. An additional annual per student transfer to the household received for school supplies. A teacher incentive element aimed at teachers that had additional reporting duties or increased class size due to the programme. They were given a 'small' transfer, which was given to the beneficiary child who then delivered it to the teacher. This was intended to substitute informal fees.	Enrolment Dropout Completion	Enrolment: 12 months 24 months Dropout: 12 months 24 months Completion: 12 months, 24 months, 36 months	Cluster-RCT	6602 Participants
Edmonds and Shrestha, 2014 Nepal	Nepal Primary and Secondary. Average student age is 12.	Schooling Incentives Project Evaluation (SIPE): children received a scholarship and an additional stipend conditional on regular attendance. Scholarship: paid or reimbursed child's school related cost. Stipend for food rations distributed through local stores given in the form of store credit. Every child that received the stipend was given an identity card with a picture of the child and their guardian used to identify beneficiary.	Attendance Enrolment Learning: composite score Completion	28 months	RCT	655 Participants
Ahmed *et al.*, 2007 Turkey	Turkey, primary and secondary schools.	Social Risk Mitigation Project: designed to provide support to the poorest 6 per cent of the Turkish population conditional on improved use of basic health and education services. Households can be eligible either for the education or health transfer. Education CCT are given to eligible households (Families having	Enrolment Completion	24 months	RDD	2905 Participants

		no social insurance, expecting babies, having children at 0-6 age and/or school-age children).				
Ribas *et al.*, 2011 Paraguay	Paraguay, rural primary and secondary schools.	Tekopora Conditional Cash Transfer Programme: a monthly grant delivered via a cash distribution system provided on several conditions. The amount provided is dependent on the number of children and pregnant women per household. It is a per child/pregnant woman premium up to a maximum of four. Support provided by to households by social workers on issues including obtaining identification cards, budget planning, cultivation of vegetables gardens, health and hygiene. Social workers also monitor conditionalities.	Attendance Completion	15-19 months	Natural experiment	6404 Participants
Kassouf, 2012 Brazil_1	Brazil, rural and urban schools. Students between 10 and 15 years old.	The BPC programme: a non-contributory pension scheme which provides a minimum wage for elders and people with disabilities who are unable to live on their own or work. To be eligible, the person must be over 64 years old or prove to be incapable to work and have a per capita family income no greater than 25 per cent of the current minimum wage. They must also have no social security aid or any other retirement plan fund. There may be more than one applicant per family, the pension of the first applicant is not included in the income calculation of the family (until 2004). In 1988, this age of eligibility was reduced to 67 years old, and in 2003, to 65 years old.	Attendance	60 months	RDD	3285 Households
Filmer and Schady, 2006 Cambodia	Cambodia, secondary schools. Students in grade 8 in the programme.	The Japan Fund for Poverty Reduction: conditional cash transfer. Scholarship recipients agree to use funds toward education, but no attempt was made to enforce this agreement. The transfer is almost exactly equivalent to average household spending per student in lower secondary school. Girls apply for the transfer in primary school and it is provided in secondary school. Girls already in secondary school could apply for the transfer if they were at risk of dropping out.	Attendance Enrolment	Approx. 12 months	RDD	2545 Participants

Robertson *et al.*, 2013 Zimbabwe	Zimbabwe, rural, urban and peri-urban secondary and primary schools.	The Manicaland HIV/STD Prevention Project: two programme arms: Unconditional Cash Transfer and Conditional Cash Transfer. UCT: Every household collected a flat rate and an additional per child transfer (up to a maximum of three children) from designated pay points every two months. CCT: Households in CCT clusters could receive the same amount but were monitored for compliance with several conditions. If households had a good reason for not meeting conditions (child missing school due to illness) committees judged these on a case-by-case basis. Spot checks were done in schools and health clinics to verify attendance.	Attendance	14-17 months	Cluster RCT	1752 Participants
Akresh *et al.*, 2013 Burkina Faso	Burkina Faso, rural primary schools. Average age of students is 10.	Nahouri Cash Transfers Pilot Project (NCTPP): CCTs given to fathers; CCTs given to mothers; UCTs given to fathers; UCTs given to mothers. For CCTs, the mother or father received a stipend per child if the conditions were met. Each child was provided with a programme booklet in which school attendance or health clinic visits were recorded. For UCTs, the mother or father received a stipend per child with no conditionalities. Both CCT and UCT households were told they could use the funds as they wished.	Attendance Enrolment Learning: Maths and language arts	24 months	Cluster RCT	2864 Participants
Mo *et al.*, 2011 China	China, rural, secondary schools (junior secondary). Average age of students 12. Students in grade 7.	No programme name: a CCT intervention for 7th grade poor students attending Junior high school. The parents of the children were given cash for each semester if the child had met the conditionalities. The amount was about one month's wage if a student dropped out and found a job in a coastal factory working as a migrant worker. The CCT was monitored by unannounced attendance checks throughout the semester.	Dropout Learning: Maths	12 months	Cluster-RCT	300 Participants

Chaudhury and Parajuli, 2006 Pakistan	Pakistan, rural and urban middle schools. Students in grades 5 to 9.	The Punjab Female Stipend Programme: the programme was targeted at girls in grades 6 - 8. In 2006 it was extended to grades 9 and 10. Around 46 per cent of the stipends were distributed by postal-carrier or postal agent and the rest were distributed through schools. The Other elements: improving physical infrastructure, limited access, poor service delivery, financial management, limited capacity of staff and system, low community and parental involvement and a raise of private sector involvement with low standards.	Enrolment Attendance	Approx. 12 months	RDD	10,560 Participants
Alam and Baez, 2011 Pakistan	Pakistan, rural and urban middle schools. Students in grades 6 to 17.	The Punjab Female Stipend Programme: the programme was targeted at girls in grades 6 through to 8. In 2006, it was extended to grades 9 and 10. Around 46 per cent of the stipends were distributed by postal-carrier or postal agent and the rest were distributed through schools. The Other elements: improving physical infrastructure, limited access, poor service delivery, financial management, limited capacity of staff and system, low community and parental involvement and a raise of private sector involvement with low standards.	Completion Attendance	Approx. 60 months	RDD	20,826 Participants
*** Follow-up was calculated as the time between the start of the intervention and data collection, in months.**						

Table 5.4 b: Overview of effect size labels

Effect size label	Description
Argentina	Heinrich, 2005
Bangladesh_1	Baulch, 2011
Bangladesh_2	Ferré and Sharif, 2014
Bangladesh_3	Khandker *et al.,* 2003
Brazil_1	Kassouf, 2012
Brazil_3	Glewwe and Kassouf, 2012; de Janvry *et al.,* 2006 and 2012
Brazil_3_a	Glewwe and Kassouf, 2012; Subsample: grades 1-4
Brazil_3_b	Glewwe and Kassouf, 2012; Subsample: grades 5-8
Brazil_4	de Brauw *et al.,* 2014
Brazil_5	Pianto and Soares, 2004
Burkina Faso_UCT	Akresh *et al*., 2013; UCT treatment arm
Burkina Faso_CCT	Akresh *et al*., 2013; CCT treatment arm
Cambodia_1	Garcia and Hill, 2009; Zavakou ND; Baez and Camacho, 2011; Attanasio and Gómez, 2004/ Attanasio *et al*., 2006
Cambodia_2	Filmer and Schady 2009, 2011, 2014
Cambodia_3	Barrera-Osorio and Filmer, 2014
China	Mo *et al*., 2011
Chile	Galasso, 2006
Chile_a	Galasso, 2006; Urban subgroup
Chile_b	Galasso, 2006; Rural subgroup
Colombia_1	Zavakou ND; Garcia and Hill, 2009; Baez and Camacho, 2011; Attanasio and Gómez, 2004/ Attanasio *et al.,* 2006
Colombia_1_a	Attanasio *et al*., 2006; subgroup: urban, age 8 -13,
Colombia_1_b	Attanasio *et al*., 2006; subgroup: Urban, age 14-17 years
Colombia_1_c	Attanasio *et al*., 2006: subgroup: rural, age 8-13 years
Colombia_1_d	Attanasio *et al*., 2006: subgroup: rural, age 14-17 years
Colombia_1_e	Garcia and Hill, 2009; Baez and Camacho, 2011: subgroup grade 9
Colombia_1_f	Garcia and Hill, 2009; Baez and Camacho, 2011; subgroup: Grade 5
Colombia_1_g	Attanasio *et al*., 2006; subgroup 8-13, urban, boys
Colombia_1_h	Attanasio *et al*., 2006; subgroup 14-17, urban, boys
Colombia_1_i	Attanasio *et al*., 2006; subgroup 8-13, urban, girls
Colombia_1_j	Attanasio *et al*., 2006; subgroup 14-17, urban, girls
Colombia_1_k	Attanasio *et al*., 2006; subgroup 8-13, rural, boys

Colombia_1_l	Attanasio *et al*., 2006; subgroup 14-17, rural, boys
Colombia_1_m	Attanasio *et al*., 2006; subgroup Sub: 8-13, rural, girls
Colombia_1_n	Attanasio *et al.,* 2006; subgroup 14-17, rural, girls
Colombia_2	Barrera-Osorio *et al*., 2011
Colombia_2_CCT	Barrera-Osorio *et al.,* 2011; CCT treatment arm
Colombia_2_CCTTertiary	Barrera-Osorio *et al.,* 2011; Tertiary treatment arm
Colombia_2_CCTSavings	Barrera-Osorio *et al.,* 2011; Savings treatment arm
Ecuador	Edmonds and Schady, 2011
El Salvador	De Brauw and Gilligan, 2011
Honduras_1	Glewwe and Olinto, 2004
Honduras_2	Benedetti *et al.,* 2015
Indonesia	Sparrow, 2007
Jamaica	Levy and Ohls, 2007
Lesotho	Pellerano *et al.,* 2014
Lesotho_a	Pellerano *et al.,* 2014; Subsample: 6- 8 years
Lesotho_b	Pellerano *et al.,* 2014; Subsample: 9-12 years
Lesotho_c	Pellerano *et al.,* 2014; Sample: 14- 17 years
Lesotho_d	Pellerano *et al*., 2014; Sample: 18- 19 years
Malawi_2_UCT	Baird *et al.,* 2010; UCT treatment arm
Malawi_2_CCT	Baird *et al.,* 2010; CCT treatment arm
Malawi_2_a	Baird *et al.,* 2010: subgroup: Primary school
Malawi_2_b	Baird *et al.,* 2010: subgroup: secondary school
Malawi_3	Luseno, 2013
Mexico_a	Dubois *et al*., 2011/ Raymond and Sadoulet, 2003: subgroup: Girls, Primary school
Mexico_b	Dubois *et al.,* 2011/ Raymond and Sadoulet, 2003: subgroup: Girls, Secondary school
Mexico_c	Dubois *et al.,* 2011/ Raymond and Sadoulet, 2003: subgroup: Boys, Primary school
Mexico_d	Dubois *et al.,* 2011/ Raymond and Sadoulet, 2003: subgroup: Boys, Secondary school
Mexico_e	Raymond and Sadoulet, 2003: subgroup: Grade 3, Primary school
Mexico_f	Raymond and Sadoulet, 2003: subgroup: Grade 4, Primary school
Mexico_g	Raymond and Sadoulet, 2003: subgroup: Grade 5, Primary school
Mexico_h	Raymond and Sadoulet, 2003: subgroup: Grade 6, Primary school
Mexico_i	Raymond and Sadoulet, 2003: subgroup: Grade 1, Secondary school
Mexico_j	Raymond and Sadoulet, 2003: subgroup: Grade 2, Secondary school

Mexico_k	Raymond and Sadoulet, 2003: subgroup: Grade 3: Secondary school
Mexico_l	Schultz, 2004; Subgroup: primary school, girls
Mexico_m	Schultz, 2004; Subgroup: primary school boys
Mexico_n	Schultz, 2004; Subgroup: secondary school, girls
Mexico_o	Schultz, 2004; Subgroup: secondary school, boys
Mexico_p	Behrman et al., 2000; Y2, Grade 4, Primary
Mexico_q	Behrman et al., 2000; Y2, Grade 6, Primary
Mexico_r	Behrman et al., 2000; Subgroup: Y2, Grade 1, Secondary
Mexico_s	Behrman et al., 2000; Subgroup: Y2, Grade 2, Secondary
Mexico_t	Behrman et al., 2000; Subgroup: Y2, Grade 3, Secondary
Mexico_u	Skuffious and Parker, 2011; subgroup Girls, 8-11 age
Mexico_v	Skuffious and Parker, 2011; subgroup Girls, 12-17 age
Mexico_w	Skuffious and Parker, 2011; subgroup Boys, 8-11 age,
Mexico_x	Skuffious and Parker, 2011; subgroup Boys, 8-11 age,
Morocco	Benhassine *et al.,* 2013
Nepal	Edmonds and Shrestha, 2014
Nicaragua_1	Maluccio *et al.,* 2009
Nicaragua_2	Macours and Vakis, 2009
Pakistan	Alam and Baez, 2011; Chaudhury and Parajuli, 2006
Paraguay	Ribas *et al.,* 2011
Philippines	Chaudhury *et al.,* 2013
Philippines_a	Chaudhury *et al.,* 2013: age 6-11
Philippines_b	Chaudhury *et al.,* 2013: age 12-14
Philippines_c	Chaudhury *et al.,* 2013: age 15-17
South Africa_a	Santana, 2008; Eyal and Woolard, 2014: Sample: age 7-8
South Africa_b	Santana, 2008; Eyal and Woolard, 2014: Sample: age 9-10
South Africa_c	Santana, 2008; Eyal and Woolard, 2014: Sample age 11- 13
South Africa_d	Santana, 2008; Eyal and Woolard, 2014: Sample: age 15-19
Tanzania	Evans *et al.,* 2014
Turkey	Ahmed *et al.,* 2007
Uruguay	Amarante *et al.,* 2011
Zimbabwe_UCT	Robertson *et al*., 2013; UCT treatment arm
Zimbabwe_CCT	Robertson *et al.,* 2013; CCT treatment arm

6. School level interventions

Education interventions taking place at the school level aim to improve the quality of the learning environment in classrooms. They include interventions providing physical inputs, or changes in the content, instructional approach, time or organisation of classes. Each programme may include a combination of some or all of these components.

We have categorised interventions according to their main component and defined seven distinct intervention types. These include structured pedagogy interventions, computer assisted learning (CAL), remedial education, programmes providing extra time for learning by lengthening the school day, interventions providing materials to schools, programmes providing new school buildings or infrastructure rehabilitation and programmes grouping students by ability.

The chapter provides the findings of our synthesis of the 65 included studies evaluating the effect of a range of interventions aiming to improve the classroom environment. We have presented the results by intervention type. Each sub-section starts with a description of the intervention type and its theory of change, followed by descriptive results and the findings addressing our research questions.

6.1 Structured pedagogy

There is an emerging body of research that indicates that it may not be enough to provide schools with new instructional materials, but that interventions should also look to improve the content and quality of instruction (Abeberese, Kumler, Linden, 2011; Lucas *et al.*, 2014). As a result, a range of interventions designed to adapt or improve educational content and/or the methods by which students are taught have been implemented. The fundamental objective of these programmes is to change existing classroom practices.

The ‘structured pedagogy’ interventions covered here typically seek to introduce new content and instructional approaches by developing structured lesson content and providing teachers with training in delivering such material, often together with materials for both teachers and children are also provided. A few of these programmes are designed to be applied across topics, while others are subject-specific, typically covering language arts or maths. In addressing core skills, such programmes might also provide knowledge that could be transferrable to other subjects (Dixon, Schagen and Seedhouse, 2011).

6.1.1 How might structured pedagogy programmes affect education outcomes?

Many teachers in L&FMICs have had a relatively low level of initial training and knowledge, influencing the standards of instruction and children’s learning outcomes (He, Linden and MacLeod, 2007). Figure 6.1a provides an ideal type programme theory for how structured pedagogy interventions may improve learning. The central element of many of these interventions is to develop evidence based curricula, content and lesson plans, together with training teachers so that they gain the knowledge and skills necessary to deliver new content. Innovative resources are often also a key feature, both to help teachers understand new concepts and plan lessons, but also to engage students and provide new tools for learning.

Figure 6.1 a: Structured pedagogy programme theory

Inputs	Activities	Outputs	Intermediate Outcomes	Final Outcomes
Evidence based content or methods, and budget to develop curriculum/plan; budget for training or hiring teachers, providing materials and any ongoing support, supervision and feedback to teachers; time and human resources to monitor and support introduction of new curriculum; time in school day set aside for intervention	(1) Teacher training (2) Ongoing teacher support, supervision and feedback (3) Provision of teacher-oriented resources or materials (4) Provision of classroom learning materials	(1) Teachers are trained or external teachers provided (2) Planned materials and resources provided (3) Supervision, support and feedback	(1) Teachers gain new skills and knowledge (2) Classroom practices change (the new content and methods are introduced) (3) New practices lead to student behaviours conducive to learning	(1) Improved student learning and achievement;
Assumptions: (1) The government or NGO/agency has the ability to administer the program, in a targeted area; (2) Educational system, schools and teachers are receptive to the intervention	*Assumptions:* (1) Trainers are available and of sufficient quality (2) Teachers will 'buy-in' to the new content and methods	*Assumptions:* (1) Teachers attend training (1) Training is of sufficient duration and quality (2) Teachers understand training (3) Relevant materials delivered in right quantities, on time and are fit-for-purpose (4) Supervision, support and feedback is provided	*Assumptions:* (1) Teachers understand training (2) Teachers and students react positively to new content and methods (3) Teachers apply new skills and knowledge in classroom (4) New materials are used	*Assumptions:* (1) Time-on-task is sufficient (2) Teachers attend school (3) New content and methods are more effective than previous

For pedagogical interventions to be successful, various factors were likely to play an important part. Relevant materials and equipment need to be of sufficient quality, must be provided in the right quantities and on time, and should function as intended. Training needs to be of sufficient duration and quality and teachers need to be able to understand it and 'buy-in' to it. Ultimately, the biggest test for such interventions is whether the innovative content and methods they aim to introduce are actually applied in classrooms, with a reasonable degree of fidelity and over a sufficient time period to make a difference.

Teachers may need on-going support over the lifecycle of the programme, with many interventions finding innovative ways of promoting continued teacher commitment and learning (Piper and Korda, 2011; Jukes and Dubeck, 2015). Many pedagogical interventions aim to encourage active student learning and greater participation in classes (Spratt, King and Bulat, 2013; Berlinski and Russo, 2013). The intervention's reception by students is also likely to be an important factor in determining its effectiveness. A series of contextual factors, such as basic classroom resources, pupil-teacher ratios and the language of instruction, are all possible determinants of intervention effectiveness.

6.1.2 Description of included studies

We included 21 studies, reported in 31 different papers. In two cases, a single paper reported on two separate programmes. In the first case, the Pratham PicTalk programme in India was implemented in different locations and the programme itself varied between year one and year two (He, Linden and MacLeod, 2007). In the second case, the Reading to Learn Intervention (RtL) was implemented in two different countries, Kenya and Uganda (Lucas *et al.*, 2014). We have used the term 'study' to refer to a unique evaluation of a programme. Table 6.1b provides an overview of the characteristics of all included studies, summarised below.

Population

All but one of the interventions targeted primary schools, with the exception being Berlinski and Russo's (2013) study of the Sa Aklat Sisikat (SAS) programme in the Philippines, which targeted secondary-school students. Studies also typically focussed on the earlier grades of primary school, with the vast majority of them targeting children from the first two grades. Those studies that reported it typically showed that there was a fairly equal split between male and female students included in the sample. Nearly all of the interventions were undertaken in public schools, with a few targeting private low-cost schools (Dixon, Schagen and Seedhouse, 2011; Pallante, 2013; Piper and Mugenda, 2014) and one targeting both public and community schools (Spratt, King and Bulat, 2013).

Setting

The included interventions were undertaken around the world, though the majority (eleven) were in Sub-Saharan-Africa; four were in South Asia, three in East Asia and the Pacific and three in Latin America. Of the Sub-Saharan-African studies, five were in Kenya, two in Uganda and one in each of South Africa, Liberia, Mali and Sudan. The South Asian studies were all in India, while the East Asian studies were in the Philippines (two) and Cambodia (one). The three Latin American studies were in Brazil, Chile and Costa Rica.

Interventions

The interventions grouped into this category include different combinations of some key intervention components, as illustrated in Table 6.1a. They all share a core element in that they are designed to introduce new content and instructional approaches into schools. The main intervention characteristics are summarised below.

Topic

The majority of interventions (n=14) were designed to introduce new pedagogy for language arts. Three were maths interventions and two combined both maths and language arts (Piper and Mugenda, 2014; RTI International, 2015). Two interventions were designed to introduce new pedagogical approaches that were not specific to a given topic (Tan, Lane and Lassibille, 1999; Leme, 2010).

Teacher training

In almost all cases, teachers were trained in how to use these new resources and/or apply the new curriculum. In a couple of cases the training seems to have been provided purely through teacher self-study (Leme, 2010; San Antonio *et al.,* 2011). The other 16 interventions provided training directly to teachers. Not all studies reported the duration of training provided, but for those that did, there was a range from two days (Abeberese, Kumler and Linden, 2011) to three weeks (Mouton, 1995), with the majority of interventions falling in the lower-to-middle portion of this range. Most programmes also provided some form of post-training monitoring, supervision or feedback to ensure that teachers were in classrooms teaching and that they were applying new concepts correctly. There were two cases in which external 'peripatetic' teachers were trained and then used to deliver the intervention to multiple schools (He, Linden and MacLeod, 2007: Year 1; Dixon, Schagen and Seedhouse, 2011).

Resources

Many of the interventions provided teacher resources such as lesson plans, activity guides and materials for making teaching aids. Typical student resources were flash-cards, wall-charts, textbooks, workbooks, storybooks or technology.

Table 6.1 a: Pedagogy Intervention Components

Study ID	Topic	Content	Teacher resources	Student resources	Teacher training	Monitor, mentor or feedback	Other
Abeberese *et al.,* 2011	Language arts	✓		✓	✓	✓	
Berlinski and Russo, 2013	Maths	✓	✓	✓	✓		
Dixon *et al*., 2011	Language arts	✓	✓	✓	✓		✓
He *et al.,* 2007 (Y1)	Language arts	✓	✓	✓	✓	✓	✓
He *et al.,* 2007 (Y2)	Language arts	✓	✓	✓	✓	✓	✓
He *et al*., 2009	Language arts	✓		✓	✓	✓	✓
Irwing *et al*., 2008	Maths	✓		✓	✓		✓
Jukes and Dubeck, 2015	Language arts	✓	✓	✓	✓	✓	
Kerwin and Thornton, 2015	Language arts	✓	✓	✓	✓	✓	✓
Leme, 2010	General	✓	✓	✓	✓	✓	
Lucas *et al*., 2014 (Uganda)	Language arts	✓		✓	✓	✓	✓
Lucas *et al*., 2014 (Kenya)	Language arts	✓		✓	✓	✓	✓
Mouton, 1995	Language arts	✓	✓	✓	✓	✓	
Nonoyama-Tarumi 2009	Language arts	✓	✓	✓	✓	✓	✓
Pallante, 2013	Language arts	✓	✓	✓	✓	✓	✓
Piper and Korda, 2011	Language arts	✓	✓		✓	✓	✓
Piper and Mugenda, 2014	Language arts; maths	✓	✓	✓	✓	✓	✓
RTI International 2015	Language arts; maths	✓	✓	✓	✓	✓	✓
San Antonio *et al.,* 2011	Maths		✓		✓	✓	
Spratt, King and Bulat, 2013	Language arts	✓	✓	✓	✓	✓	
Tan, Lane and Lassibille, 1999	General		✓	✓	✓		

Additional components

Several interventions incorporated additional components. Some interventions included mechanisms for monitoring student performance through assessment (Nonoyama-Tarumi, & Bredenberg, 2009; Spratt, King and Bulat, 2013) and report cards (Kerwin and Thornton, 2015). Other components included community libraries (He, Linden and Macleod, 2009; Lucas *et al.,* 2014) and maintenance support for the technology involved (He *et al.,* 2007). Some interventions also included on-going contact with teachers via email (Piper and Mugenda, 2014) or text message tips and motivation for lesson plans with small monetary incentives in the form of mobile phone credit (US$1) as an incentive for further communication (Jukes and Dubeck, 2015).

A few more large-scale programmes included a range of wider components. For example, the Early Grade Reading Assessment (EGRA) Plus programme in Liberia included student report cards designed to encourage parent engagement with their children's' education, radio outreach promoting reading, a regional reading competition and capacity building for Ministry of Education staff (Piper and Korda, 2011). A more comprehensive overview of intervention components is provided in Table 6.1a of Appendix J, section 6.1.

Targeting

Interventions typically targeted schools and students only in specified grades. Fourteen studies explicitly referred to this type of targeting, though given the age-specific nature of many of these interventions, it is likely that more or less all the interventions targeted a particular age-group. Three interventions explicitly referred to having deliberately targeted schools in areas with lower socioeconomic status (Tan, Lane and Lassibille, 1999; Dixon, Schagen and Seedhouse, 2011; San Antonio *et al.*, 2011) and one study targeted schools with a high dropout rate (Tan, Lane and Lassibille, 1999). The duration of interventions ranged from six months or less (n=6), with the shortest just one month long (Abeberese, Kumler and Linden, 2011). Four of the 20 interventions covered lasted for approximately a year. Five lasted two years or longer. The data suggests that the incidence of classes varied from daily n=4), to two (n=1) or three times per week (n=1).

Comparisons

Eleven of the included studies compared the effect of an intervention against a no-intervention comparison group (Abeberese, Kumler, Linden, 2011; Dixon, Schagen and Seedhouse, 2011, Irwing *et al.*, 2008; Kerwin and Thornton, 2015; Leme, 2010; Lucas *et al.*, 2014 (Kenya); Lucas *et al.,* 2014 (Uganda); Nonoyama-Tarumi, & Bredenberg, 2009; Pallante, 2013; Piper and Mugenda, 2014; RTI, 2015; Spratt, King and Bulat, 2013). There were also seven studies that evaluated multiple treatment arms (Berlinski and Russo, 2013, He, Linden and MacLeod, 2007 (Year 1), He, Linden and MacLeod, 2007 (Year 2), He, Linden, and Macleod, 2009; Jukes and Dubeck, 2015; Piper and Korda, 2011; Tan, Lane and Lassibille, 1999). For all such cases, we chose the business as usual comparison. A further two compared the treatment group to a comparison group receiving another education intervention, though these were small interventions designed to negate any Hawthorne-like effects (San Antonio *et al.*, 2011, Mouton, 1995)[24].

[24] In the case of Mouton (1995) control group teachers were given training in how to use interactive group techniques when they gave classes. In the case of San Antonio et al. (2011), the control group attended a seminar on the same topics covered by the treatment group. In the case of Piper (2009), control schools also received the DOE Reading Toolkit—a large but portable metal cabinet holding teaching and learning materials.

Outcomes

The included studies reported on a broad range of education outcomes. All studies included some measure of learning. In all cases, learning was measured through test scores. Two studies provided cognitive test scores. Eighteen studies provided a measure of language arts learning and a further fourteen for maths. We extracted data from studies relating to the grade of students tested, enabling us to undertake sub-group analysis by grade for language arts and maths. Two studies reported composite test scores aggregating both language arts and maths scores, and two reported the results of cognitive tests. In addition to the learning outcomes, a few studies also reported on teacher performance (n=4), teacher attendance (n=2), student attendance (n=6) and drop-out (n=1). The majority of studies measured outcomes over a relatively short time period. Nine studies looked at outcomes less than 12 months after the intervention start and another six looked at outcomes after exactly a year. Four studies looked at outcomes from between thirteen and twenty-four months, while two looked at longer term outcomes. The table of characteristics contains full details on follow-up.

Study Design

Seventeen of the included studies were cluster randomised controlled designs. The other four employed quasi-experiment designs with baseline and endline data collection and either covariate matching (Irwing *et al.,* 2008), multivariate analysis (Nonoyama-Tarumi, & Bredenberg, 2009; Pallante, 2013) or difference-in-difference (DID) (Leme, 2009) respectively. The table of characteristics contains details on the designs employed by each study.

Qualitative studies, process evaluations and project documents

We identified nine additional documents that presented qualitative, process and project information for five of the included programmes. These documents were used to provide additional background information and to inform our qualitative synthesis of intervention and implementation features associated with interventions' relative success and failure. This is a relatively small number of documents, probably due to most of our included interventions being small or medium-scale field trials rather than larger nationwide-programmes.

Table 6.1 b: Characteristics of included studies for pedagogy

Study ID	Setting	Intervention summary	Included outcomes	Follow-Up*	Study design	Sample Size
Abeberese, Kumler and Linden, 2011	Philippines, Tarlac province, public primary schools. Students were fourth graders and were, on average, nine years old	The **Sa Aklat Sisikat (SAS) programme** provided age-appropriate reading material, trained teachers to incorporate reading into their curriculum, and supported these changes through a 31 day reading marathon to encourage students to read. Teachers attended a two-day training session in advance. SAS monitored schools to ensure programme fidelity and support teachers' use of new books. SAS provided 60 age-appropriate storybooks in English and Filipino to each class. Teachers/classes kept all of the material at the end of the intervention.	Language arts; maths	4 and 7 months	Cluster randomised controlled trial	8,106 students at baseline
Berlinski and Russo, 2013	Costa Rica, urban and peri-urban areas, public secondary schools. Students were seventh graders and were, on average, fourteen years old	The maths reading intervention (no official name) randomly assigned 85 schools to five conditions: (T1) control (T2) new curriculum (T3) new curriculum and interactive white-board (T4) new curriculum and computer lab (T5) new curriculum and laptop for every child. T2 and T3 are assessed here. All intervention arms incorporated new pedagogical material and provided teachers' manuals and student workbooks, as well as school principal intervention manuals. For T2, the intervention relied on images or paper. T3 equipped classrooms with an interactive whiteboard, a desktop, a router and open-source mathematics software (GeoGebra), while all teachers received training and a laptop and	Teacher performance; maths	Approximately 4 months	Cluster randomised controlled trial	85 schools, 190 teachers and 4,830 students

		a virtual classroom was installed to support teachers' work.				
Dixon, Schagen and Seedhouse 2011	India, urban private primary schools in low-income areas of Hyderabad, Andhra Pradesh, Students were in grade one and aged seven on average	The **Synthetic Phonics** intervention aimed to improve reading and spelling by providing structured lessons guided by the 'Jolly Phonics' approach which teaches children to read using decoding and synthetic phonics skills. The Jolly Phonics package included: (1) worksheets (2) flash cards (3) blending cards (4) books (5) and reading books. Classes were taught by a peripatetic teacher trained by the researchers.	Language arts	6 months	Cluster randomised controlled trial	20 schools, 501 students
He, Linden and MacLeod, 2007 (Year 1)	India, Maharasthra, urban primary schools, students were aged seven on average	The **Pratham PicTalk programme (Year 1)** had two main components: the PicTalk Machine which lets children point at pictures with a stylus and hear the word pronounced aloud and a set of interactive activities designed around 440 flashcards. Pratham hired and trained 68 peripatetic teaching assistants who rotated between schools. Teachers went to weekly sessions for feedback and to prepare materials for the week ahead. Monitors supervised teacher attendance.	Attendance; language arts; maths	Approximately 12 months	Cluster randomised controlled trial	97 schools and 5,317 students
He, Linden and MacLeod, 2007 (Year 2)	India, Maharasthra, rural primary schools, students were aged seven on average	The **Pratham PicTalk programme (Year 2)** had the same two main components as Year 1. However, instead of peripatetic teachers, schoolteachers went to 5-day training. There were three intervention arms (machines only, activities only, and machines with activities). All schools also had regular access to	Attendance; composite; language arts; maths	Approximately 12 months	Cluster randomised controlled trial	181 schools and 9,644 students

		Pratham monitors who circulated amongst the schools to assist teachers as questions arose.				
He, Linden, and Macleod, 2009	India, Mumbai, public urban primary schools, students were in first grade	The **Shishuvachan programme** included two treatment arms, only one of which is examined here. It involved two main components: reading classes and provision of a library. Teachers received training and met supervision to ensure consistency in training and implementation. Both the class and library were offered in the community. In classes, teachers used story books, flash cards for word and letter recognition and charts.	Language arts	Approximately 1 month (6 weeks)	Cluster randomised controlled trial	67 schools and 2,679 students
Irwing *et al.*, 2008	Sudan, primary schools in rural Khartoum State, students were aged 7-11	The programme involved teaching maths using the abacus (no official name). Trained teachers provided two hours of classes to students each week for 34 weeks. Classes consisted of training in mental arithmetic including storing information in working memory while other mental operations were performed and then retrieved.	Cognitive	34 weeks, so approximately 7-8 months	Quasi-experiment, baseline and endline data collection & covariate matching	16 schools and 3,185 children
Jukes and Dubeck, 2015	Kenya, Kwale and Msambweni districts, rural primary schools, students, students were in grades one and two and aged 5-15	**Under the Health and Literacy Intervention (HALI)**, teachers received manuals containing sequential lessons and instructional materials. There was initial training and follow-up workshops to provide an opportunity to problem-solve, receive and share feedback, and introduce a set of new instructional materials. Teachers received weekly text messages with instructional tips for lesson plans. These required a response to receive a small amount of phone credit (US$1). Teachers were	Cognitive; language arts; maths	Approximately 24 months	Cluster randomised controlled trial	101 schools and 2,539 students

		asked to fill out a weekly summary sheet documenting lessons that worked well and suggestions for improvement.				
Kerwin and Thornton (2015)	Uganda, public primary schools, students were in grades one and two and were aged seven on average	Under the **Northern Uganda Literacy Project (NULP)**, children were taught in their local language rather than in English, often the de facto language of instruction in Ugandan primary schools. Teachers were given training and on-going support focusing on the uptake of practical and appropriate classroom skills. Classrooms were provided with slates that allowed each student to practice writing individually. Students were given textbooks. Teachers were provided with literacy lesson guides. Class content was introduced slowly with time for repetition and revision. The NULP also promoted parent and community engagement through parent-teacher meetings and parent training on how to interpret their child's literacy report card and use a simple home reading assessment tool.	Attendance, Enrolment; language arts; teacher attendance; teacher performance	Approxim ately 11 months	Cluster randomised controlled trial	1,900 students
Leme, 2010	Urban areas of Sao Paolo, Brazil, public primary schools, students were in the fourth and eighth grades	Based in Brazil's decentralised schooling system, where municipalities have some power to determine curriculum content and contract out educational services to private organisations. The **'structured methods'** varied according to contractual arrangements but typically encompassed curricular design, the provision of materials, teacher training, supervision and monitoring of implementation. There were often interactive websites with supplemental activities, texts, documents and education-related	Completion, language arts; maths	Started late 1990s and end-line data collected in 2007 (so at least 96 months)	Quasi-experiment with baseline & endline data collection & difference-in-	393 municipalities

		articles, as well as test-question banks. Some systems also offered online content for students.			difference analysis	
Lucas *et al.*, 2014 (Uganda)	Uganda, public primary schools, students were in grades 1 to 3	The **Reading to Learn (RtL)** intervention incorporated three key components: teacher preparedness and practice, school leadership, classroom learning environments. Teachers, head-teachers and school-management committees received training. Monthly in-class mentoring support for teachers was also provided. Classrooms received instructional materials, primarily books and stationery supplies to create visual aids. Teachers attended quarterly meetings to share ideas, observe model lessons and receive refresher training. Some schools also received an additional component establishing mini-libraries in communities.	Attendance; language arts; maths	Approximately 21-22 months	Cluster randomised controlled trial	109 schools, 6,861 students
Lucas, *et al.*, 2014 (Kenya)	Kenya, public primary schools, students were in grades 1 to 3	As above, but in Kenya	Attendance; language arts; maths	Approximately 21-22 months	Cluster randomised controlled trial	112 schools, 7,015 students
Mouton, 1995	South Africa, Kwa-Zulu Natal, rural and urban public primary schools, students were in grade five	The **English and Operacy programme (EOP)** was based on ‘suggestopedic’ methodology which is a communicative approach emphasising massive language input, attention to affect, and pair and group work. Teachers were given three weeks of training in the methodology, as well as English language and thinking skills. Some new materials were probably provided. Half the teachers received visits to provide support and motivate them.	Teacher performance; language arts; maths	Approximately 12 months	Cluster randomised controlled trial	2,200 students, 48 teachers

		Monitoring occurred though interviews with teachers and principals as well as classroom observation.				
Nonoyama -Tarumi, & Bredenberg, 2009	Cambodia, rural and urban primary schools, students were in grade one	**The School Readiness Programme (SRP)** modified curricular content to compensate for the lack of pre-schooling that is thought to cause high grade repetition. Programme components included: (1) development of special curriculum documentation (2) 14 day teacher training programme (3) a regular monitoring regimen to support teachers in their implementation (4) physical upgrading of classrooms (5) formalised student assessment for monitoring purposes.	Language arts	Approximately 2 and 10 months	Quasi-experiment with baseline & endline data collection & multivariate analysis	20 schools and 931 students
Pallante, 2013	Chile, urban and rural private primary schools. Students were in grade one and aged six on average	**The Collaborative Language and Literacy Instruction Project (CLLIP)** introduced a new curriculum designed around key components of language arts and literacy. Teachers received professional development over four workshops spread over the school year. They were trained in the theoretical basis for the content being taught, use of student assessment and strategies to increase achievement in reading and writing skill regardless of students' level. Teachers were given graphic organisers to help them prepare lessons. Materials (not described) were provided to promote best practice. Classes were monitored and teachers received coaching reports and sustained support. They were also trained to use student test results to identify children at risk, monitor progress and tailor small group teaching accordingly.	Language arts	Approximately nine months	Quasi-experiment with baseline & endline data collection with multilevel analysis	305 students

Piper and Korda, 2011	Liberia, primary schools, students were in grades two and three	**The Early Grade Reading Assessment (EGRA) Plus** had two treatment groups, of which only T1 is assessed here. Under T1, reading levels were assessed to continually assess student performance. Teachers were trained, continually assessed and supported by coaches. Schools were provided with resource materials, books, and parents and communities were informed of student performance. In the second year, outreach to communities via radio shows and reading competitions were introduced. Student performance and progress were also discussed with parents. There was semi-annual refresher training and materials were revised.	Language arts	Approximately 15-17 months	Cluster randomised controlled trial	76 schools and 2,988 students.
Piper and Mugenda, 2014	Kenya counties including Nairobi, Nakuru, Kiambu, Murang'a, and Kisumu. Urban, peri-urban and rural public and private primary schools, students were in grades one and two	The **Primary Math and Reading (PRIMR) initiative** included the following core activities: (1) curriculum-based content in Kiswahili, English, and mathematics; (2) lesson plans across all 3 subjects; (3) teachers and head teachers trained to implement lessons; (4) regular supervision and monitoring of teachers; (5) Reading and maths materials for students and teachers; (6) teachers trained to employ continuous assessment methods; (7) Programme materials revised and updated; (8) trained tutors and instructional coaches trained, monitored and coached teachers. Some programme adjustments were made for student cohort two, including the revision of materials, open-to-the-public reading and maths contests and teachers' monthly reflection meetings.	Language arts; maths	Approximately 12 months	Cluster randomised controlled trial	411 schools and 4,385 students

RTI International (2015)	Kenya, public primary schools, students were in grade one	**The Primary Math and Reading (PRIMR) Rural Expansion Programme** included most core PRIMR activities, though unlike the full PRIMR programme, teachers were not given teachers' guides and therefore training supported the development of lesson plans. Activities included: (1) curriculum-based content in Kiswahili, English and maths; (2) teachers trained to implement lessons and plan lessons; (3) regular supervision and monitoring of teachers; (4) reading and maths materials for students and teachers; (5) teachers trained to employ continuous assessment methods; (6) trained tutors and instructional coaches trained, monitored and coached teachers; (7) reading and maths exhibitions with parents and community invited to visit schools; (8) teacher monthly reflection meetings.	Language arts; maths	Approximately 18 months	Cluster randomised controlled trial	4,566 students
San Antonio *et al.*, 2011	Philippines, urban public primary schools in Luzon, students were in grade six	Grade six mathematics teachers were invited to volunteer, with participants given training through self-contained printed instructional packages or 'modules' on specific mathematics units. Teachers proceeded with teaching duties while studying the modules that were distributed to them simultaneously. Teachers received follow-ups from school-heads and supervisors. The control group attended a seminar on the same topics covered by the treatment group.	Teacher performance; maths	1 month (five weeks)	Cluster randomised controlled trial	50 teachers

Spratt, King and Bulat, 2013	Mali, urban and rural public and community primary schools, students were in grades one and two	**The Read-Learn-Lead (RLL)** offered students and teachers structured and systematic lessons, activities, and accompanying materials for instruction and practice on critical early reading skills in the local language medium during the first years of elementary school. RLL provided teachers with pre-service and in-service professional development, and support and monitoring visits. It also involved on-going assessment of children's reading performance. Materials included flash cards, levelled national-language arts readers, and related posters.	Language arts	12 months	Cluster randomised controlled trial	80 schools
Tan, Lane and Lassibille, 1999	Philippines, rural schools, primary school, students were in grade one	**The Dropout Intervention programme (DIP)** included two treatment arms, but only treatment 1 providing multi-level materials is assessed here. Under this arm, all teachers in the beneficiary schools received pedagogical materials to help them pace their teaching according to student ability. Prior to implementation, teachers attended a week-long training course on the use of materials.	Drop-out; composite test score; language arts; maths	Approximately 12 months	Cluster randomised controlled trial	14 schools and 1,929 students.

*** Follow-up was calculated as the time between the start of the intervention and data collection, in months.**

6.1.3 Synthesis of findings

The results of our synthesis are presented in two sections. First, we present the meta-analysis findings relating to the effects of pedagogy interventions on primary and secondary outcomes. Second, we provide a discussion of the overall findings, with reference to the qualitative synthesis of intervention and implementation features associated with relative success and failure in improving educational outcomes.

Effects of structured pedagogy interventions on intermediate, primary and secondary outcomes

This section reports the results of the meta-analyses of the effects of structured pedagogy interventions, addressing question 1a and 1b. We have structured the presentation of results according to the 'ideal type' theory of change (Figure 6.1a), starting with intermediate/ secondary outcomes (for example, teacher attendance), followed by education access outcomes (attendance, drop-out) and final outcomes (learning composite test scores, language test scores and maths test scores).

The studies include a range of different follow-up periods, with the majority of effect sizes after approximately 12 months. Therefore we selected these for the meta-analysis when available. Where studies reported estimates at multiple follow-up periods, we included the data-point closest to 12 months.

Some studies included multiple treatment arms. For outcomes relating to language arts and maths test scores, we were able to include all treatment arms as we used robust variance estimation that permits the inclusion of multiple dependent effects. However, for meta-analyses relating to other outcomes, we included the treatment arm most comparable to the other studies. The effect size included in the meta-analysis for the Pratham PicTalk programme year two (He, Linden and MacLeod, 2007) is for a treatment arm that received both the PicTalk machine and activities designed around flash cards[25]. We included effect sizes for the government treatment arm of the Northern Uganda Literacy Project (NULP) in preference to the NGO treatment arm (Kerwin and Thornton, 2015).

Where studies reported outcomes in multiple languages, we always chose the most widely spoken language for example selecting Kiswahili outcomes over English language outcomes for studies of Kenyan programmes. In all cases where we had to choose from multiple effect sizes for a single outcome, we have reported additional estimates narratively where they provided substantively different findings to those included in the meta-analysis. Estimates that were not substantively different from those already included in the meta-analysis have been reported in technical Appendix H, section 6.1.

There is just a single study that reports on dropout, but as many as 18 for language arts test scores. All effect sizes are expressed as standardised mean difference (SMD), interpreted as the magnitude of the number of standard deviation changes in the outcome for the intervention group as compared to students in comparison schools. SMD scores are interpreted as the number of standard deviation changes in the outcome.

[25] The PicTalk machine allows children to point at pictures with a stylus and hear the word pronounced aloud. Additionally there are touch points to receive instructions in Marathi as well as quiz questions that ask the children to identify words by pointing, then give auditory feedback.

Effects of structured pedagogy interventions on teacher performance and attendance

Teacher performance

Four studies provided a total of 10 measures relating to teacher performance. Because of the heterogeneity of measures meta-analysis was not possible. Kerwin and Thornton (2015) report a measure of teacher performance related to the amount of time spent teaching classes in a local language, finding an overall increase across two treatment arms (NGO: SMD=0.29, 95% CI [0.13, 0.45]; government: SMD=0.38, 95% CI [0.22, 0.54]). The other three studies report effects that are either small in magnitude or very imprecisise (see Appendix H for full results).

Teacher attendance

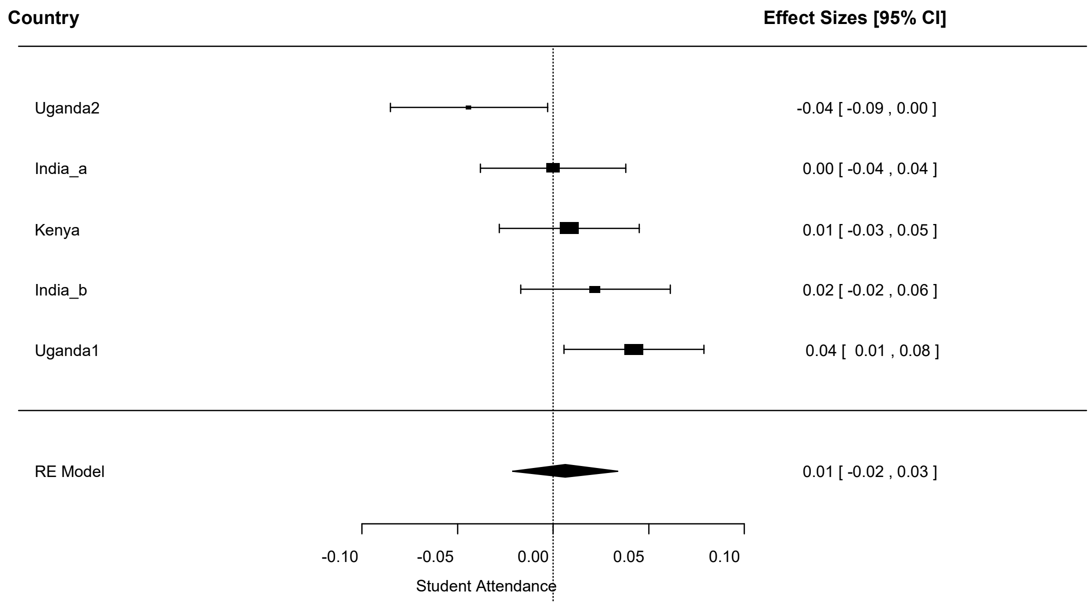
Kerwin and Thornton (2015) is the only study to report a measure of teacher attendance in the form of whether teachers reported having been absent over the past month. The effects for the NGO treatment arm suggest a reduction in teacher absences (SMD=-0.12, 95% CI [-0.52, 0.28]) and for the government the results suggest an increase in teacher absences (SMD=0.19, 95% CI [-0.22, 0.59]), however the confidence intervals of both estimates cross the line of no effect.

Effects of structured pedagogy interventions on student attendance, drop-out, completion and cognitive outcomes

Student attendance

The overall average effect of pedagogy interventions on student attendance is 0.01, 95% CI [-0.02, 0.03]. The homogeneity tests (I^2 = 61.18%, τ^2 = 0.0006, Q(df = 4) = 10.1726, p = 0.0376) indicate moderate heterogeneity. The effect sizes range from -0.04, 95% CI [-0.09, 0.00] In Uganda (Kerwin and Thornton, 2015) to 0.04, 95% CI [0.01, 0.08] also in Uganda (Lucas *et al.,* 2014). Confidence intervals for the four studies overlap. Sensitivity analysis indicates that removing any one study does not make a substantive difference to the overall average effect (see Appendix H for results of all sensitivity analyses).

Figure 6.1 b: Student attendance[26]



Kerwin and Thornton (2015) also include an estimate for the NGO implemented arm of their programme, which can be compared to the government treatment arm included in the meta-analysis. Unlike the government implemented treatment, the NGO version of the programme suggests an increase in student attendance (SMD=0.06, 95% CI [0.02, 0.10]).

Drop-out

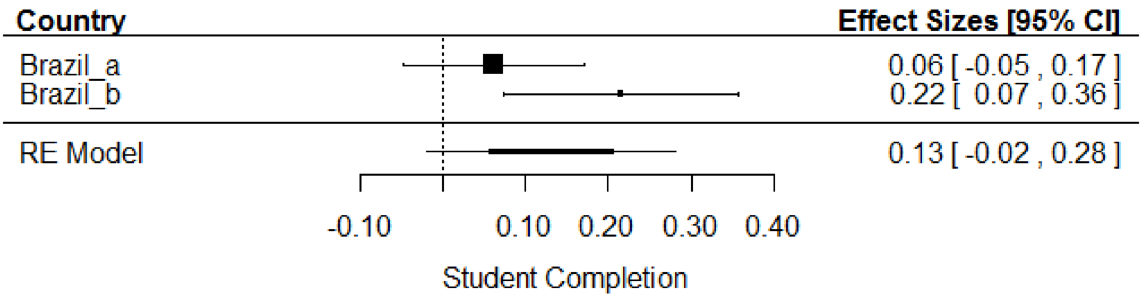
A single study, Tan, Lane and Lassibile (1999) measures this outcome, finding a substantive reduction in drop-out among the treatment group (SMD= -0.14, 95% CI [-0.25, -0.03]).

Completion

The overall average effect of pedagogy interventions on student completion is 0.13, 95% CI [-0.02, 0.28]. The homogeneity tests (I^2 = 65.33%, τ^2 = 0.0078, Q(df = 1) = 2.8840, p = 0.0895) indicate moderate heterogeneity. The effect sizes included here are both from the same study in Brazil (Leme, 2010) with separate estimates by grade. Both effect sizes are positive, though the one for Grade 4 (Brazil_a) is smaller in magnitude, suggesting the intervention was more successful in promoting completion for the older Grade 8 students than the younger Grade 4 students.

[26] India_a: He, Linden and macLeod, 2007: Year 1 Programme
India_b: He, Linden and macLeod, 2007: Year 2 Programme
Kenya: Lucas *et al.,* 2014
Uganda1: Lucas *et al*., 2014
Uganda2: Kerwin and Thornton, 2015

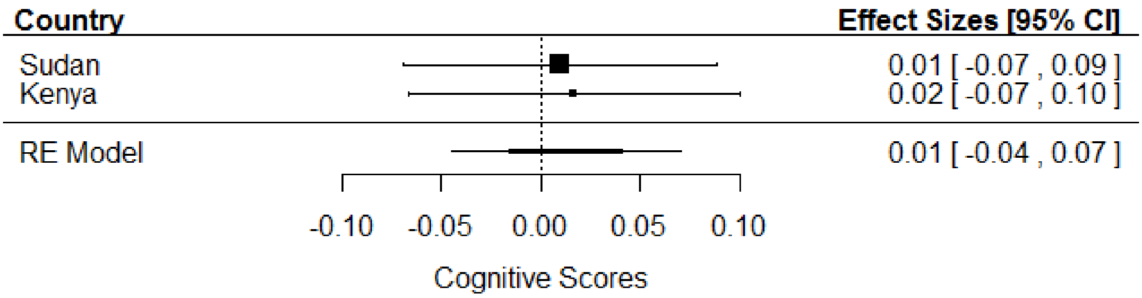
Figure 6.1 c: Student completion[27]



Cognitive outcomes

Two studies reported estimates for cognitive outcomes and the overall pooled effect is 0.01 and 95% CI of [-0.04, 0.07]. The homogeneity tests (I² = 0.00%, $\tau^2 = 0$, Q(df = 1) = 0.0148, p = 0.9030) indicate heterogeneity is low. The effect size for Kenya from James and Dubeck (2015) is a synthetic effect size representing grades from 7 through to 11.

Figure 6.1 d: Cognitive test scores28



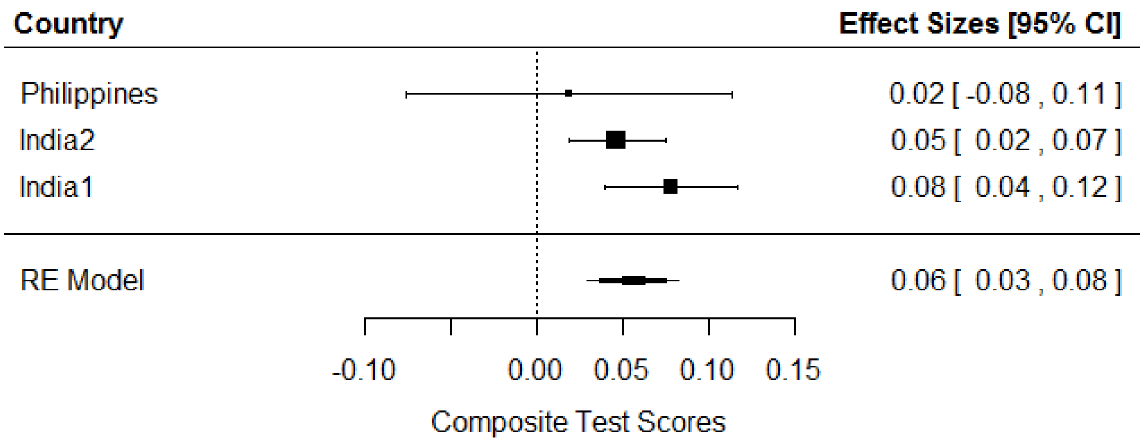
Effects of structured pedagogy interventions on composite, language arts and maths test scores

Composite test scores

Three studies reported estimates for composite learning outcomes. The overall average effect of pedagogy interventions on composite learning outcomes is 0.06 [0.03, 0.08]. Tests indicate heterogeneity is low (I² = 20.17%, $\tau^2 = 0.0001$, Q(df = 2) = 2.3084, p = 0.3153). The effect sizes range from 0.02, 95% CI [-0.08, 0.11] in the Philippines (Tan, Lane and Lassibille, 1999) to 0.08, 95% CI [0.04, 0.12] in India (He, Linden and MacLeod, 2007 [Year 2 programme]). Sensitivity analysis indicates that the substantive result is not sensitive to the removal of any one study from the analysis. Composite test scores by grade sub-groups do not provide substantively different findings.

[27] Brazil_a: Leme, 2010: Grade 4
Brazil_b: Leme, 2010: Grade 8
[28] Kenya: James and Dubeck (2015): grades 7-11
Sudan: Irwing *et al.*, 2008

Figure 6.1 e: Composite Test Scores[29]



Language arts test scores

The meta-analysis in Figure 6.1f contains 67 separate effect sizes from 18 different studies, though most studies provide multiple effect sizes representing different measures[30]. The overall average effect of pedagogy interventions on language arts test scores is 0.23, 95% CI [0.13, 0.34], calculated using robust variance estimation[31]. The effect sizes range from -0.14, 95% CI [-0.23, -0.05] to 0.90 95% CI [0.85, 0.94]. Homogeneity tests indicate a large amount of between study variability (I^2 = 98.98%, Q(df = 66) = 3537.8609, p-val < .0001).

The effects from Kerwin and Thornton's (2015) evaluation of the Northern Uganda Literacy Intervention (NULP) indicate that, for the most part, the NGO implemented treatment arm performed better than that implemented by the government. The PicTalk year two programme included three treatment arms[32] (He, Linden and MacLeod, 2007). The 'machines only' treatment arm performed best, followed by the 'activities only' and finally the 'machines and activities' arm, though the estimate was still positive and statistically significant and the differences between treatment arms were not especially large.

The results reported by Tan, Lane and Lassibille (1999) for the Dropout Intervention Programme (DIP) in the Philippines indicate that results for English were mostly larger than those for Filipino. Conversely, Jukes and Dubeck's evaluation of the Health and Literacy Intervention (HALI) in Kenya found that effects were larger for Kiswahili outcomes than for English, while RTI's (2011) study of the Primary Math and Reading (PRIMR) Rural Expansion Programme found no consistent difference between effects for Kiswahili and English outcomes.

[29] India1: He, Linden and MacLeod, 2007: Year 1 Programme; India2: He, Linden and MacLeod, 2007: Year 2 Programme; Philippines: Tan, Lane and Lassibille, 1999.
[30] So, for example, in Figure 6.1f, the entry for Mali reports on outcomes for the Read-Learn-Lead (RLL) programme (Spratt, King and Bulat, 2013). Each of the five effect sizes for this programme covers a different measure with examples including initial sound identification, listening comprehension and letter recognition.
N.B. We have not reported effect sizes relating to language arts from Piper and Mugenda's 2014 study of the Primary Math and Reading (PRIMR) initiative. This is because they reported extreme variation in effects ranging from SMD = 0.16 through to SMD = 24.16 and we were unable to obtain clarification from the authors when they were contacted.
[31] This method adjusts the standard errors to allow us to include multiple dependent effect sizes in a single meta-analysis.
[32] T1: PicTalk machines and activities; T2: machines only; T3: activities only. The PicTalk machine lets children point at pictures with a stylus and hear the word pronounced aloud. Additionally there are touch points to receive instructions in Marathi as well as quiz questions that ask the children to identify words by pointing, then give auditory feedback.

The results for the Read Learn Lead (RLL) intervention in Mali (Spratt, King and Bulat, 2013) indicate that the intervention produced larger effects across language arts outcome measures for the grade two students than for the grade one students. Finally, Nonoyama-Tarumi and Bredenberg's (2009) evaluation of the School Readiness Programme (SRP) in Cambodia indicates that effects were larger immediately after the end of the two month intervention than six months after it finished at the end of the school year, suggesting that the size of the effect may not be sustained in the long term.

Figure 6.1 f: Language Arts Test Scores (see table 6.1c for key)

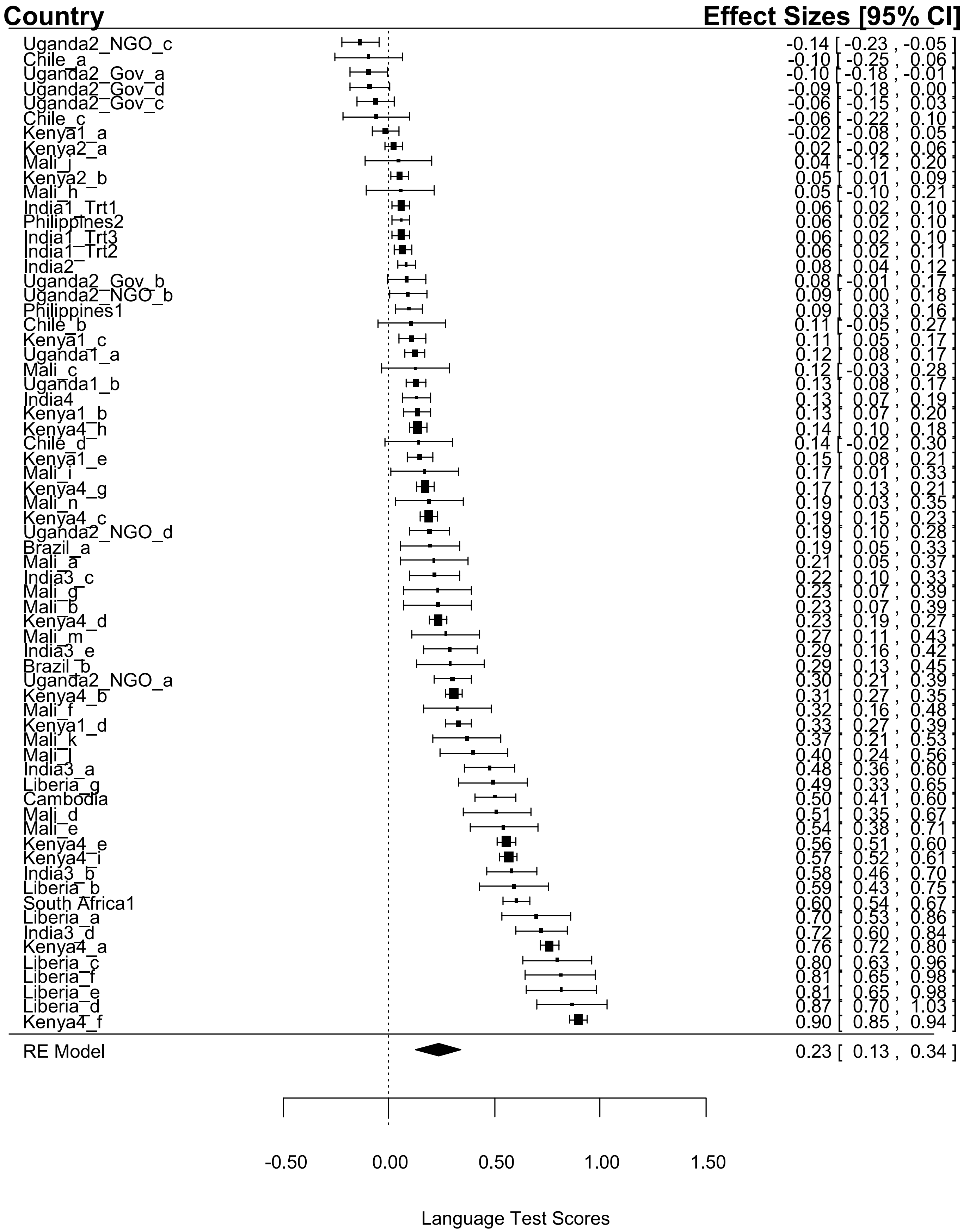


Table 6.1 c: Language Arts Test Scores: Key (ordered as in forest plot)

Effect label	Description
Uganda2_NGO_c	Kerwin and Thornton, 2015; NGO treatment arm; English word recognition
Chile_a	Pallante, 2013; Vocabulary
Uganda2_Gov_a	Kerwin and Thornton, 2015; Government treatment arm; English word recognition
Uganda2_Gov_d	Kerwin and Thornton, 2015; Government treatment arm; Writing Test Score
Uganda2_Gov_c	Kerwin and Thornton, 2015; Government treatment arm; Oral English Score
Chile_c	Pallante, 2013; Reading Comprehension
Kenya1_a	Jukes and Dubeck, 2015; English Letter Knowledge
Kenya2_a	Lucas *et al.*, 2014; Kenya programme; Written Literacy Exam
Mali_j	Spratt, King and Bulat, 2013; Listening Comprehension; Grade 2
Kenya2_b	Lucas *et al.,* 2014; Kenya programme; Oral Literacy Exam
Mali_h	Spratt, King and Bulat, 2013; Orientation to Print; Grade 2
India1_Trt1	He, Linden and MacLeod, 2007; Year 2 programme; Machines and Activities; English Test
Philippines2	Abeberese, Kumler and Linden, 2011; Reading; 7 month follow-up
India1_Trt2	He, Linden and MacLeod, 2007; Year 2 programme; Machines only; English Test
India1_Trt3	He, Linden and MacLeod, 2007; Year 2 programme; Activities only; English Test
India2	He, Linden and MacLeod, 2007; Year 1 programme; English Test
Uganda2_Gov_b	Kerwin and Thornton, 2015; Government treatment arm; EGRA Test Score
Uganda2_NGO_b	Kerwin and Thornton, 2015; NGO treatment arm; Oral English Score
Philippines1	Tan, Lane and Lassibille, 1999; Filipino Test
Chile_b	Pallante, 2013; Nonword Reading Fluency
Kenya1_c	Jukes and Dubeck, 2015; Swahili Passage Reading Comprehension
Uganda1_a	Lucas *et al.*, 2014; Uganda programme; Oral Literacy Exam
Mali_c	Spratt, King and Bulat, 2013; Listening Comprehension; Grade 1
Uganda1_b	Lucas *et al.*, 2014; Uganda programme; Lucas *et al.*, 2014; Oral Literacy Exam
India4	He, Linden and MacLeod, 2009; Normalised Reading Level (Hindi, Marathi and Urdu)
Kenya1_b	Jukes and Dubeck, 2015; Swahili Passage Reading Fluency (words per minute)
Kenya4_h	RTI International, 2015; Kiswahili High Reading Fluency
Chile_d	Pallante, 2013; Word Reading

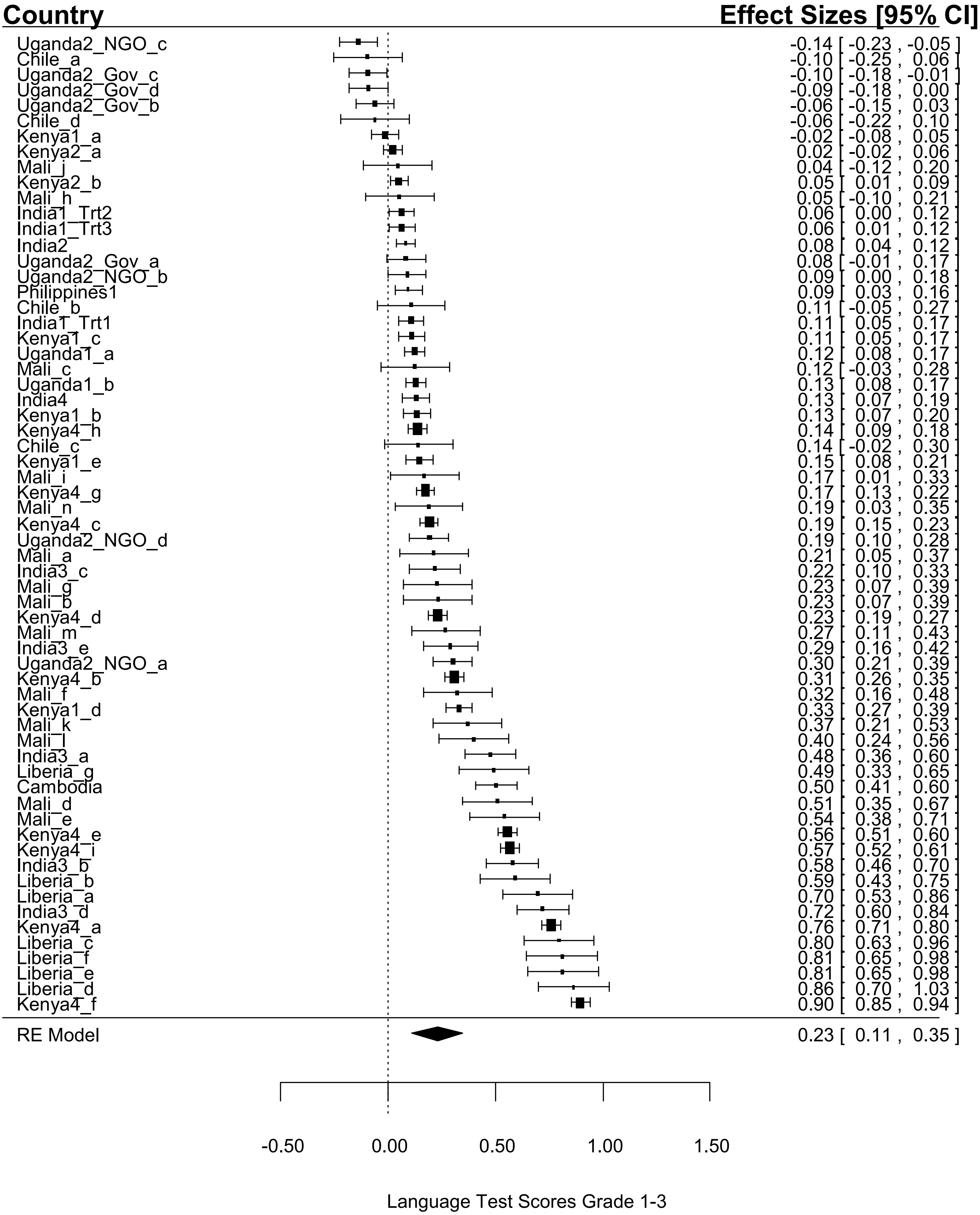
Kenya1_e	Jukes and Dubeck, 2015; Swahili Word Identification (words per minute)
Mali_i	Spratt, King and Bulat, 2013; Phonemic Awareness (initial sound identification); Grade 2
Kenya4_g	RTI International, 2015; Kiswahili Ability to Decode Words and Read Aloud to Listener
Mali_n	Spratt, King and Bulat, 2013; Oral Reading Fluency (connected text); Grade 2
Kenya4_c	RTI International, 2015; Kiswahili Decoding Fluency (correct words per min)
Uganda_2_NGO_d	Kerwin and Thornton, 2015; NGO treatment arm; Writing Test Score
Brazil_a	Leme, 2010; Portuguese Proficiency; Fourth Grade
Mali_a	Spratt, King and Bulat, 2013; Orientation to Print; Grade 1
India3_c	Dixon, Schagen and Seedhouse, 2011; Letter Matching Test
Mali_g	Spratt, King and Bulat, 2013; Oral Reading Fluency (connected text); Grade 1
Mali_b	Spratt, King and Bulat, 2013; Phonemic Awareness (initial sound identification); Grade 1
Kenya4_d	RTI International, 2015; Kiswahili Reading Fluency
Mali_m	Spratt, King and Bulat, 2013; Correct Invented Words per Minute; Grade 2
India3_e	Dixon, Schagen and Seedhouse, 2011; Dictation
Brazil_b	Leme, 2010; Portuguese Proficiency; Eighth Grade
Uganda_2_NGO_a	Kerwin and Thornton, 2015; NGO treatment arm; EGRA Test Score
Kenya4_b	RTI International, 2015; Kiswahili Syllable Fluency
Mali_f	Spratt, King and Bulat, 2013; Correct Invented Words per Minute; Grade 1
Kenya1_d	Jukes and Dubeck, 2015; Swahili Letter Sounds (letters per minute)
Mali_k	Spratt, King and Bulat, 2013; Correct Letters per Minute; Grade 2
Mali_l	Spratt, King and Bulat, 2013; Correct Familiar Words per Minute; Grade 2
India3_a	Dixon, Schagen and Seedhouse, 2011; Burt Reading Test
Liberia_g	Piper and Korda, 2011; Listening Comprehension
Cambodia	Nonoyama-Tarumi and Bredenberg, 2009; Khmer Language
Mali_d	Spratt, King and Bulat, 2013; Correct Letters per Minute; Grade 1
Mali_e	Spratt, King and Bulat, 2013; Correct Familiar Words per Minute; Grade 1
Kenya4_e	RTI International, 2015; Kiswahili Reading Comprehension
Kenya4_i	RTI International, 2015; Kiswahili Basic Reading Fluency
India3_b	Dixon, Schagen and Seedhouse, 2011; Schonell Spelling Test
Liberia_b	Piper and Korda, 2011; Phonemic Awareness
South Africa1	Mouton, 1995; English Test

Liberia_a	Piper and Korda, 2011; Letter Naming Fluency (letters per minute)
India3_d	Dixon, Schagen and Seedhouse, 2011; Sound Blending Word Test
Kenya4_a	RTI International, 2015; Kiswahili Letter Sound Fluency
Liberia_c	Piper and Korda, 2011; Familiar Word Fluency (words per minute)
Liberia_f	Piper and Korda, 2011; Reading Comprehension
Liberia_e	Piper and Korda, 2011; Oral Reading Fluency (words per minute)
Liberia_d	Piper and Korda, 2011; Unfamiliar Word Fluency (words per minute)
Kenya4_f	RTI International, 2015; Kiswahili Listening Comprehension

Language arts test scores: grade sub-group analysis

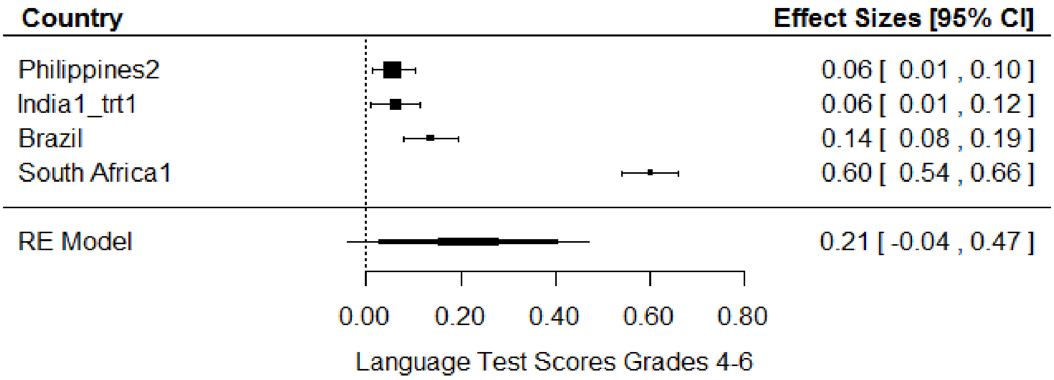
The following two forest plots provide separate analyses for effect sizes for language arts test scores, split by early grades (grades 1-3 - see Figure 6.1g) and later grades (grades 4-6 – see Figure 6.1h). The overall average effect for the grades 1-3 sub-group is 0.23, 95% CI [0.11, 0.35], calculated with robust variance estimation. It is comparable to the overall average for all language test scores reported in Figure 6.1f and this is likely because the majority of effect sizes are for that grade grouping. Homogeneity tests again indicate a large amount of between study variability (I^2 = 97.96%, Q(df = 62) = 3036.0232, p-val < .0001).

Figure 6.1 g: Language Arts Test Scores for Grades 1-3 sub-group (see table 6.1d for key)



The overall average effect for the grades 4-6 sub-group is of a similar magnitude as the full sample and the 1-3 grade sub-group, but it is less precise (SMD=0.21, 95% CI [-0.04, 0.47)]. Homogeneity tests again indicate a large amount of between study variability (I^2 = 98.91%, Q(df = 3) = 242.9698, p-val < .0001). There are only four studies included in this analysis and the sensitivity analysis indicates that removing the 'South Africa1' estimate for Mouton (1995) reduces the overall average effect to 0.08 (95% CI [0.03, 0.19]).

Figure 6.1 h: Language Arts Test Scores for Grades 4-6 sub-group[33]



Maths test scores

Figure 6.1i contains 24 separate effect sizes from 14 different studies as many studies provide multiple effect sizes representing different measures[34]. The overall average effect of pedagogy interventions on maths learning outcomes is positive and statistically significant when calculated with robust standard errors (SMD=0.14, 95% CI [0.08, 0.20]). The effect sizes range from the negative and statistically significant (SMD=-0.09, 95% CI [-0.15, -0.03]) to the very large and statistically significant 0.74 95% CI [0.70, 0.79]. Tests indicate that heterogeneity is again very high (I^2 = 98.42%, Q(df = 23) = 1452.6546, p-val < .0001).

Examining the forest plot for maths test scores indicates that, unlike with language arts outcomes, the largest effect was for the 'machines and activities' treatment arm from the Indian PicTalk year two programme[35] (He, Linden and MacLeod, 2007).

[33] Philippines2: Abeberese, Kumler and Linden, 2011; Reading; 7 month follow-up
India1_Trt1: He, Linden and MacLeod, 2007; Year 2 programme; Machines and Activities; English Test; Brazil: Leme, 2010; Portuguese Proficiency; Eighth Grade
South Africa1: Mouton, 1995; English Test

[34] So, for example, in Figure 6.1i, the entry for Liberia reports on outcomes for the Early Grade Reading Assessment (EGRA) Plus (Piper and Korda, 2011). Each of the eight effect sizes for this programme covers a different measure with examples including number identification, addition and subtraction.

[35] T1: PicTalk machines and activities; T2: machines only; T3: activities only. The PicTalk machine lets children point at pictures with a stylus and hear the word pronounced aloud.

Figure 6.1 i: Maths Test Scores (see table 6.1d for key)

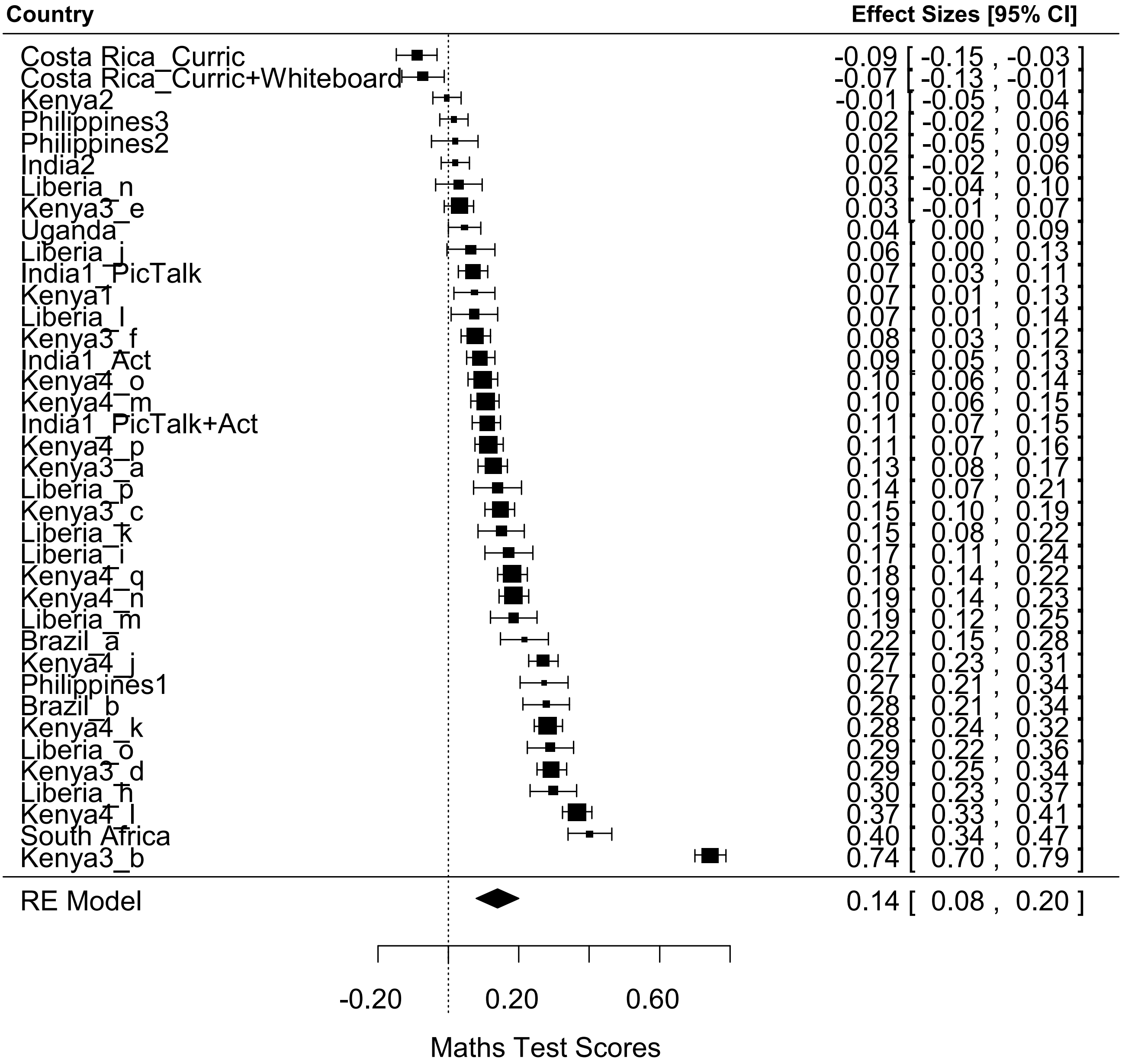


Table 6.1 e: Maths Test Scores: Key (ordered as in forest plot)

Effect label	Description
Costa Rica_Curric	Berlinski and Russo, 2013; New Curriculum Treatment Arm; Maths
Costa Rica_Curric+Whiteboard	Berlinski and Russo, 2013; New Curriculum and Interactive White-board Treatment Arm; Maths
Kenya2	Lucas *et al.,* 2014; Kenya programme; Maths
Philippines3	Abeberese, Kumler and Linden, 2011; Maths
Philippines2	Tan, Lane and Lassibille, 1999; Maths
India2	He, Linden and MacLeod, 2007; Year 1 programme; Maths
Liberia_n	Piper and Korda, 2011; Subtraction 1 (lower difficulty)
Kenya3_e	Piper and Mugenda, 2014; Quantity Descrimination
Uganda	Lucas *et al.*, 2014; Uganda programme; Maths
Liberia_i	Piper and Korda, 2011; Quantity Discrimination (correct per minute)
India1_PicTalk	He, Linden and MacLeod, 2007; Year 2 programme; Machines only; Maths
Kenya1	Jukes and Dubeck, 2015; Numeracy
Liberia_l	Piper and Korda, 2011; Addition 1 (lower difficulty)
Kenya3_f	Piper and Mugenda, 2014; Number Identification
India1_Act	He, Linden and MacLeod, 2007; Year 2 programme; Activities only; Maths
Kenya4_o	RTI International, 2011; Subtraction 2 (higher difficulty)
Kenya4_m	RTI International, 2011; Addition 1 (lower difficulty)
India1_PicTalk+Act	He, Linden and MacLeod, 2007; Year 2 programme; Activities and Machines; Maths
Kenya4_p	RTI International, 2011; Subtraction 2 (higher difficulty)
Kenya3_a	Piper and Mugenda, 2014; Word problems
Liberia_p	Piper and Korda, 2011; Fractions
Kenya3_c	Piper and Mugenda, 2014; Addition Fluency
Liberia_k	Piper and Korda, 2011; Addition 2 (higher difficulty)
Liberia_l	Piper and Korda, 2011; Addition 1 (lower difficulty)
Kenya4_q	RTI International, 2011; Word problems
Kenya4_n	RTI International, 2011; Addition 2 (higher difficulty)
Liberia_m	Piper and Korda, 2011; Subtraction 2 (higher difficulty)
Brazil_a	Leme, 2010; Grade 8; Maths
Kenya4_j	RTI International, 2011; Number Identification
Philippines1	San Antonio *et al.*, 2011; Maths
Brazil_b	Leme, 2010; Grade 4; Maths
Kenya4_k	RTI International, 2011; Quantity Discrimination (correct per minute)
Liberia_o	Piper and Korda, 2011; Multiplication
Kenya3_d	Piper and Mugenda, 2014; Missing Numbers
Liberia_h	Piper and Korda, 2011; Number Identification
Kenya4_l	RTI International, 2011; Missing Numbers
South Africa	Mouton, 1995; Maths
Kenya3_b	Piper and Mugenda, 2014; Subtraction Fluency

Maths test scores: grade sub-group analysis

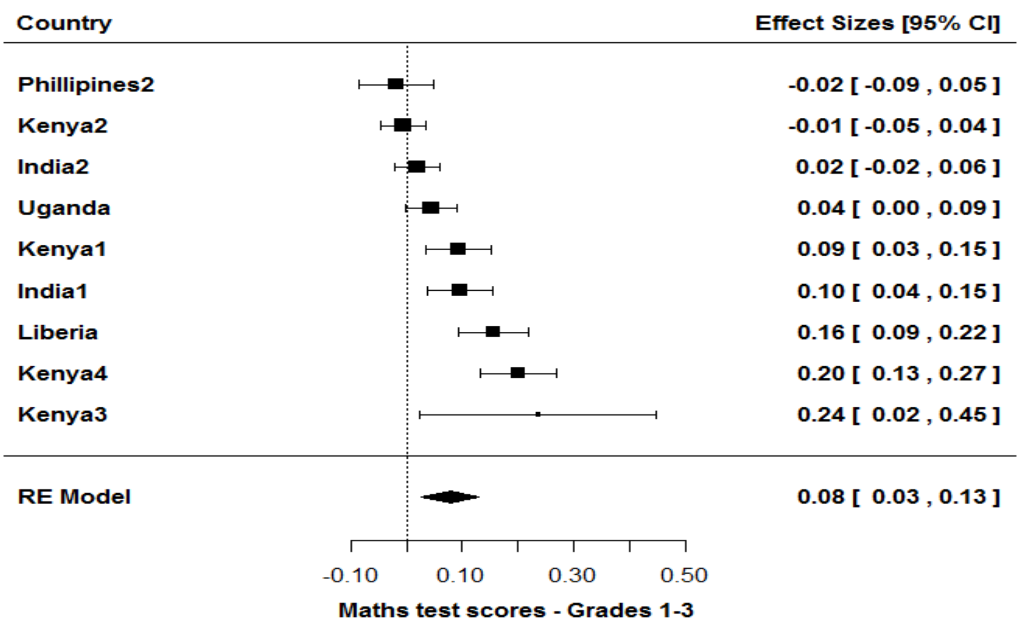
The average effect of pedagogy interventions on maths test scores in grades 1-3 is 0.08, 95% CI [0.03, 0.13]. The effect sizes range from -0.02, 95% CI [-0.09, 0.05] to 0.16, 95% CI [0.09, 0.22]. Tests indicate that heterogeneity is high (I^2 = 86.19%, τ^2 = 0.0716, Q(df = 8) = 50.0119, p = < .0001). The results are not sensitive to the removal of any single study.

For grades 4-6, the overall average effect is 0.21, 95% CI [0.04, 0.37]. Tests indicate there is high heterogeneity (I^2 = 97.05%, τ^2 = 0.0275, Q(df = 3) = 100.7261, p = < .0001). Effect sizes range from -0.02, 95% CI [-0.09, 0.05] to 0.24, 85% CI [0.02, 0.45]. The results are sensitive to the removal of studies from Brazil (Leme, 2010) and South Africa (Mouton, 1995). When either of them is removed the average effect falls to 0.18, 95% CI [-0.04, 0.41] or 0.14, 95% CI [-0.01, 0.29] respectively and the confidence intervals cross the line of no effect.

For grades 7-11, the overall average effect is 0.13, 95% CI [-0.10, 0.35]. Tests indicate heterogeneity is high (I^2 = 92.78%, τ^2 = 0.0375, Q(df = 2) = 38.0804, p = < .0001). The results are sensitive to the removal of the study from Costa-Rica. Without this study there is no heterogeneity and the magnitude of the overall effect is substantially larger and more precise (SMD=0.26, 95% CI [0.16, 0.35]).

Looking across sub-groups, there is an overall positive effect of the pedagogy interventions on maths test scores. It appears effects are larger in magnitude for higher grades.

Figure 6.1 j: Maths Test Scores for Grades 1-3 sub-group[36]



[36] Philippines2: Tan, Lane and Lassibille, 1999; Grade 1; Kenya2: Lucas *et al.,* 2014; Kenya programme; Grades 1 to 3; India2: He, Linden and MacLeod, 2007; Year 1 programme; Grades 3 to 5
Uganda: Lucas *et al.,* 2014; Uganda programme; Grades 1 to 3; Kenya1: Jukes and Dubeck, 2015; Grades 1 to 2; India1: He, Linden and MacLeod, 2007; Year 2 programme; Machines and Activities; Grades 1 to 2; Liberia: Piper and Korda, 2011; Grades 2 to 3; multiple measures combined into a single synthetic effect; Kenya4: RTI International, 2011; Grade 1; multiple measures combined into a single synthetic effect; Kenya3: Piper and Mugenda, 2014; Grades 1 to 2; multiple measures combined into a single synthetic effect

Figure 6.1 k: Maths Test Scores for Grades 4-6 sub-group[37]

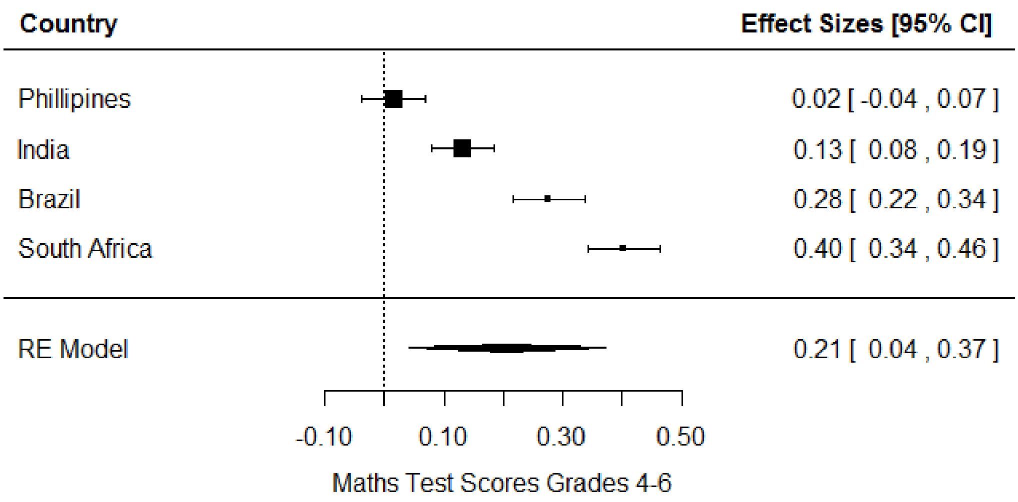
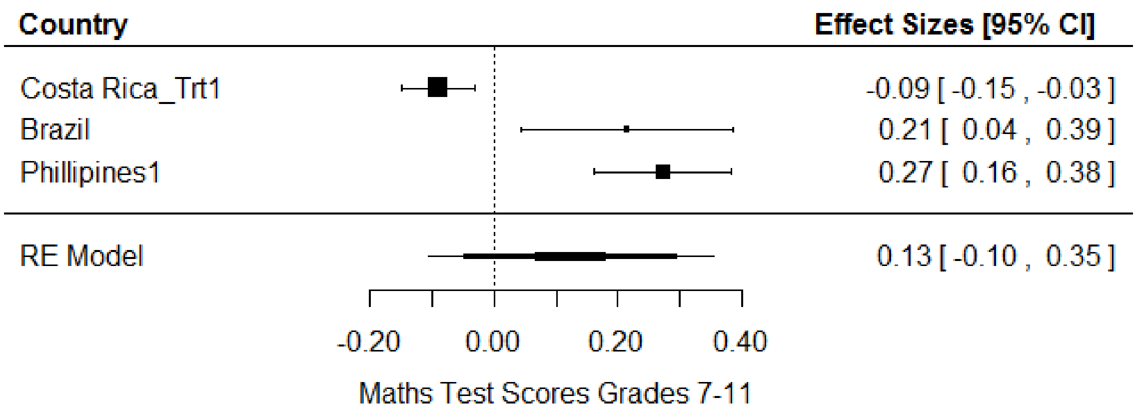


Figure 6.1 l: Maths Test Scores for Grades 7-11 sub-group[38]



6.1.4 Summary of findings and discussion

We identified 21 studies of structured pedagogy programmes across 11 countries in South Asia, East Asia and the Pacific, Latin America and Sub-Saharan Africa. We were able to examine the effects on student attendance, completion, cognition, maths, language arts and composite test scores using meta-analysis. We also aimed to identify barriers and facilitators of the effectiveness of structured pedagogy programmes through qualitative synthesis of the evidence on process and implementation. Table 6.1f and Table 6.1g provide the summary findings from this synthesis, together with details about the context for which they apply. We draw on this analysis when discussing the results.

---

[37] Philippines: Abeberese, Kumler and Linden, 2011; Reading; 7 month follow-up Grade 4
India: He, Linden and MacLeod, 2007; Year 2 programme; Machines and Activities; Grades 3-5
Brazil: Leme, 2010: Grade 4
South Africa: Mouton, 1995; Grade 5
[38] Costa Rica_Trt1: Berlinski and Russo, 2013; curriculum only treatment arm; Grade 7; Brazil: Leme, 2010: Grade 8; Philippines1: San Antonio *et al.*, 2011; Grade 6

The overall average effects relating to school participation range from 0.01, 95% CI [-0.02, 0.03] for attendance to 0.13, 95%CI [-0.02, 0.28] for completion. Improving school participation is not a key objective of most pedagogy interventions so it is unsurprising that these interventions do not seem to have affected attendance rates in programme schools. There was however one intervention that explicitly set out to reduce dropout (Tan, Lane and Lassibille, 1999) and, in this case, a positive effect was observed. There is also some limited evidence that structured pedagogy interventions may increase the likelihood of students finishing the school year. Leme (2010) found that students from both grade 4 and grade 8 were more likely to complete the school year as a result of the intervention, although the confidence intervals of the grade 4 result crossed the line of no effect.

Average effects on learning outcomes range from 0.01, 95% CI [-0.04, 0.07] for cognitive test scores, to 0.06, 95% CI [0.03, 0.08] for composite test scores, 0.14, 95% CI [0.08, 0.20] for maths test scores and 0.23, 95% CI [0.13, 0.34] for language arts test scores. While there is no evidence of an effect on cognitive outcomes, there is relatively strong evidence that structured pedagogy interventions can lead to substantial improvements of learning outcomes for composite, language arts and maths test scores.

However there is a large range in effects. Effect sizes for maths test scores range from -0.02, 95% CI [-0.09, 0.05] to 0.16, 95% CI [0.09, 0.22], while effect sizes for language arts test scores range from -0.14, 95% CI [-0.23, -0.05] to 0.90 95% CI [0.85, 0.94]. This suggests that intervention design, implementation and context can play an important role in determining how effective or not pedagogy interventions are.

Tables 6.1f and 6.1g provide the process and implementation factors which affected these interventions, based on evidence from qualitative studies, process documenta and included impact evaluations. If these findings are examined with reference to Figure 6.1a, the pedagogy programme theory, it is clear that programmes did not always meet all the key assumptions. While sometimes teachers were receptive to programmes (Abeberese, Kumler and Linden, 2011 [SAS]; KAPE, 2004 [SRP]), others opposed proposed changes, with some teacher unions taking strike action (Piper and Korda, 2011 [EGRA]; Piper and Mugenda, 2014 [PRIMR]; Spratt, King, and Bulat, 2013 [RLL]). Programmes frequently suffered from problems in implementation, such as failure to deliver tools and supplies (KAPE, 2004 [SRP]; Lucas *et al.*, 2012 [Uganda and Kenya]; Spratt and Ralangaita, 2013 [RLL]; Piper and Mugenda, 2013 [PRIMR]) or provision of training on time and to a sufficient standard (KAPE, 2004 [SRP]; RTI, 2015 [PRIMR-REP]; Mouton, 1995 [EOP]; Spratt, King and Bulat, 2011 [RLL]). For example, in the case of the Read-Learn-Lead (RLL) intervention in Mali, schools did not receive the planned materials and teacher training. They also note that the intervention started later than planned due to an overly ambitious plan for implementation, that teacher turnover caused problems and that classes were not held as often as intended.

There was some evidence that teachers were not knowledgeable or experienced enough to fully understand their training (Berlinski and Russo, 2013 [Costa Rica]; Dixon, Schagen and Seedhouse, 2011 [SP]; He, Linden and MacLeod, 2007 [PicTalk Yr2]; He, Linden and MacLeod, 2009 [Shishuvachan]; Mouton, 1995 [EOP]) and did not always implement lessons as intended or as often as planned (KAPE 2004 [SRP]; Piper and Mugenda, 2014 [PRIMR]). For example, RLL classes were not always taught, as intended, in local languages, something that was also problematic for the Reading to Learn (RtL) intervention in Kenya. Lucas *et al.* (2014) report that, though RtL classes were meant to be provided in Swahili, they were often taught in English. In the case of the Philippines SAS programme, teachers tended to keep books in school rather than let children take them home (Abeberese, *et al.,* 2011).

There were also reported problems related to a failure to fully take into account key contextual factors such as the limited resources of education systems (Dixon, Schagen and Seedhouse, 2011 [SP]; Tan, Lane, and Lassibille, 1999 [DIP]; Lucas *et al.,* 2014 [RtL]; Nonoyama-Tarumi and Bredenberg, 2009 [SRP]; Piper & Mugenda, 2014 [PRIMR]); Mouton, 1995 [EOP]), or high rates of enrolment (Lucas *et al.,* 2014 [RtL]; Nonoyama-Tarumi and Bredenberg, 2009 [SRP]; Piper & Mugenda, 2014 [PRIMR]; Spratt, King and Bulat, 2013 [RLL]). For example, in South Africa, the English and Operacy Programme (EOP) training was not implemented as planned. Conditions in schools and their limited resources made it very difficult to implement the small group resource-dependent exercises as planned (Mouton, 1995). There were also reports that some teachers struggled to grasp the new material and approaches being taught.

These implementation and contextual issues may have reduced the effectiveness of some programmes. However, several of the interventions where problems were observed also report relatively large effects on maths and language arts test scores. For example, the Early Grade Reading Assessment (EGRA) programme in Liberia produced some relatively large effects on test scores, but study authors Piper and Korda (2011) report that teachers lacked the necessary experience and capacity to implement the approach correctly and PRIMR was disrupted by teacher strikes. Similarly, the RLL's implementation was delayed, there were issues with teacher training and the programme was implemented in low-resource settings (Spratt, King and Bulat, 2013). It appears programmes can succeed, in spite of implementation failures and challenging contexts. Overall, the evidence does not indicate that any combination or single factor in terms of intervention design, implementation or context was decisive in determining interventions' success. Table 6.1f and Table 6.1g together with the pedagogy section in Appendix J provide a more detailed account of the evidence on these issues. To sum up, there is relatively strong evidence for the beneficial effects of structured pedagogy interventions on maths and language arts outcomes in most contexts. There is however considerable heterogeneity in effects, which we are not able to explain.

Table 6.1 f: Key descriptive findings: process and implementation

Descriptive finding	Citation and context
Materials and equipment were distributed as expected and were well-maintained	Abeberese, Kumler and Linden, 2011 (Philippines, SAS)
Materials and equipment were delayed and/ or not of the desired standard	KAPE, 2004 (Cambodia, SRP) Lucas *et al.,* 2012 (Uganda and Kenya, RtL) Spratt and Ralaingita, 2013 (Mali, RLL)
Teachers attended training sessions	Abeberese, Kumler and Linden, 2011 (Philippines, SAS) Brooker *et al*., 2013 (Kenya, HALI) He, Linden and MacLeod, 2007 (India, PicTalk Yr2)
Issues with the quality of the teacher training may have prevented teachers from delivering new content appropriately	KAPE, 2004 (Cambodia, SRP) Mouton, 1995 (South Africa, EOP) Spratt, King and Bulat, 2013 (Mali, RLL) RTI, 2015 (Kenya, PRIMR-REP)
Problems with implementation meant that some interventions did not begin on time	Piper & Mugenda, 2013 (Kenya, PRIMR) Spratt, King and Bulat, 2011 (Mali, RLL)
Teachers prevented students from taking books home	Abeberese, Kumler and Linden, 2011 (Philippines, SAS) KAPE, 2004 (Cambodia, SRP)
Teachers welcomed new programmes	Abeberese, Kumler and Linden, 2011 (Philippines, SAS) KAPE, 2004 (Cambodia, SRP)
Teachers opposed some of the changes that a programme wanted to make	Piper and Mugenda, 2013 (Kenya, PRIMR)
Teachers valued text messages with instructional tips and motivation	Jukes and Dubeck 2015 (Kenya, HALI)
There were insufficient resources provided for implementation staff supporting the teachers	KAPE, 2004 (Cambodia, SRP) Piper & Mugenda, 2013 (Kenya, PRIMR) RTI, 2015 (Kenya, PRIMR-REP)
Teachers lacked the necessary experience and capacity to implement new instructional approaches	Abeberese, Kumler and Linden, 2011 (Philippines, SAS) Mouton, 1995 (South Africa, EOP) Piper and Korda, 2011 (Liberia, EGRA)

Descriptive finding	Citation and context
Teachers were concerned about the time and additional work required by programmes	Jukes and Dubeck, 2015 (Kenya, HALI) RTI, 2011 (Kenya, PRIMR-REP) San Antonio *et al.*, 2011 (Philippines)
High teacher turnover was sometimes a problem	Piper & Mugenda, 2014 (Kenya, PRIMR) Spratt, King and Bulat, 2013 (Mali, RLL)
Actual lesson time was often less than that scheduled by programmes	KAPE 2004 (Cambodia, SRP) Piper and Mugenda, 2014 (Kenya, PRIMR)
The national government may have reduced the supply of school inputs to schools involved in a programme	Spratt, King and Bulat, 2013 (Mali, RLL)
Teachers implemented programme activities and used instructional aids in class	Abeberese, Kumler and Linden, 2011 (Philippines, SAS) Jukes and Dubeck, 2015 (Kenya, HALI) Lucas *et al.*, 2014 (Uganda and Kenya, RtL) Spratt, King and Bulat, 2013 (Mali, RLL)
Over a half of the teachers did not implement the new methods properly	Mouton, 1995 (South Africa, EOP)
Teachers felt the programme improved their professional competence	San Antonio *et al.*, 2011 (Philippines) Spratt, King and Bulat, 2013 (Mali, RLL)
The programme improved teacher and learner attitudes	Kerwin and Thornton, 2015 (Uganda, NULP)
The programme promoted participatory local language classes over learning-by-rote of English	Kerwin and Thornton, 2015 (Uganda, NULP)

Table 6.1 g: Key descriptive findings: context

Descriptive finding	Citation and context
Programme implementation was disrupted by political events and flooding	Abeberese, Kumler and Linden, 2011 (Philippines, SAS) Piper and Mugenda, 2014 (Kenya, PRIMR) Spratt, King and Bulat, 2013 (Mali, RLL)
Context of high growth in enrolment put pressure on schools' resources and capacity	Lucas *et al.,* 2014 (Uganda and Kenya, RtL) Nonoyama-Tarumi and Bredenberg, 2009 (Cambodia, SRP) Piper & Mugenda, 2014 (Kenya, PRIMR) Spratt, King and Bulat, 2013 (Mali, RLL)
Intervention design stipulated that content should be taught in students' local language to promote learning	Lucas *et al.,* 2014 (Kenya, RtL) Kerwin and Thornton, 2011 (Uganda, NULP) RTI, 2015 (Kenya, PRIMR-REP)
Students were often taught and tested in a language other than their local language	Jukes and Dubeck, 2015 (Kenya, HALI) Lucas *et al.,* 2014 (Kenya, RtL) Spratt, King and Bulat, 2013 (Mali, RLL)
Intervention implementation disrupted by teacher strikes	Piper and Korda, 2011 (Liberia, EGRA) Piper and Mugenda, 2014 (Kenya, PRIMR) Spratt, King, and Bulat, 2013 (Mali, RLL)
Low levels of teacher knowledge and experience. Traditional instructional practices prevalent	Berlinski and Russo, 2013 (Costa Rica) Dixon, Schagen and Seedhouse, 2011 (India, SP) He, Linden and MacLeod, 2007 (India, PicTalk Yr2) He, Linden and MacLeod, 2009 (India, Shishuvachan) Mouton, 1995 (South Africa, EOP)
Many schools suffer from resource constraints, with limited availability of material and large classes	Abeberese, Kumler and Linden, 2011 (Philippines, SAS) KAPE, 2004 (Cambodia, SRP) Mouton, 1995 (South Africa, EOP) Piper and Mugenda, 2013 (Kenya, PRIMR)
Programmes implemented in low resource settings	Dixon, Schagen and Seedhouse, 2011 (India, SP) Tan, Lane, and Lassibille, 1999 (Philippines, DIP) Lucas *et al.*, 2014 (Kenya, RtL) Piper and Mugenda, 2013 (Kenya, PRIMR) Mouton, 1995 (South Africa, EOP) Nonoyama-Tarumi and Bredenberg, 2009 (Cambodia, SRP)

6.2 Computer Assisted Learning (CAL)

Computer assisted learning programmes (CAL) use computers, either in the form of laptops or computer labs, to aid or support children's learning. In some cases, they are delivered as an integrated package together with new content and instructional approaches, alongside training for teachers in delivering this material. In other cases, the main focus is simply on providing children with access to computers.

6.2.1 Description of included studies

We identified 18 studies reported in 25 different papers that evaluated the effect of CAL programmes on education outcomes. These studies assessed 16 unique programmes and were published between 2003 and 2014. Table 6.2.b provides the summary of characteristics of included studies, as described in more detail in this section.

Population

Twelve of the studies focused on the primary school level, three at the secondary school level (Cristia *et al.,* 2013; Imbrogno, 2014[39]), and three at both levels (Barrera-Osorio *et al.,* 2009; Sharma, 2014; Berlinski *et al.* 2013). Fourteen studies evaluated programmes implemented in public schools, one in private for-profit schools (Lai *et al.,* 2011) and two in NGO-run schools[40]. For one study (Berlinski *et al.,* 2013) it was unclear in which type of school the programme was implemented. Although the studies covered a range of different grades from grade 1 to grade 11, the majority of the studies covered grades 3, 5, 6 and 7.

Setting

The programmes evaluated by the studies cover a range of settings in East Asia and the Pacific, South Asia, and Latin America and the Caribbean. Four of the studies were conducted in China (Mo *et al.,* 2014; Lai *et al.,* 2013., Yang *et al.,* 2013[41]; Lai *et al.,* 2011, ; Lai *et al.,* 2013); two in India (Linden *et al*., 2008, Banerjee *et al*., 2007), four in Peru (Cristia *et al*., 2012; David and Quispe 2013; Cristia *et al.,* 2013; Humpage, 2013); one in Colombia (Barrera-Osorio *et al.* 2009); one in Ecuador, (Carillo *et al*., 2010); one in Uruguay (De Melo *et al.* N.d); one in Mexico (Imbrogno, 2014); one in Chile (Imbrogno, 2014); and one in Nepal (Sharma, 2014).

Six studies took place in a rural setting (Mo *et al.*,2014; Banerjee *et al*., 2007; Yang *et al.,* 2013; Lai *et al*., 2013; Cristia *et al.,* 2012; Humpage, 2013). Three studies took place in an urban setting (Cristia *et al*., 2013; Lai *et al*., 2011;David and Quispe 2013; De Melo *et al.* N.d;) and four studies included both urban and rural locations in their sample (Linden *et al.* 2008[42]; Berlinski *et al.* 2013; Sharma, 2014). For the final four studies it was unclear whether they took place in rural or urban settings (Barrera-Osorio *et al*., 2009, Carillo *et al*., 2010; Imbrogno, 2014).

Interventions

The interventions grouped into this category include different combinations of some key intervention components, as illustrated in table 6.2a. All 16 programmes included the key feature of CAL programmes, namely the use of computer technology to facilitate learning through interactive instruction.

In/ out of school hours: Learning took place during schools hours ('in-school') for eight of the programmes (Mo *et al*., 2014; Linden *et al*., 2008; Cristia *et al.,* 2013; Carillo *et al.,* 2010; Barrera-Osorio *et al*., 2009 Berlinski *et al.,* 2013; Imbrogno, 2014), outside of official school hours ('out-of-school') in four programmes (Lai *et al.,* 2011 Yang *et al.,* 2013; Lai *et al.*,2013; Linden *et al.,* 2008), and both in-school and out-of-school for six programmes (Cristia *et al*., 2012; David and Quispe 2013; De Melo *et al.,* (n.d); Banerjee *et al.,* 2007; Sharma, 2014; Humpage, 2013).

[39] Imbrogno (2014) represents two studies: The MCT Programme in Mexico and the MCT programme in Chile.
[40] Linden, two treatment arms considered two studies.
[41] This paper reports on three experiments included in the review. These are treated as three different studies.
[42] Linden, two treatment arms considered two studies.

Subject focus: Thirteen programmes focused on (a) specific subject area(s), nine targeting maths (Lai *et al.,* 2011; Lai *et al.,* 2013, Yang *et al.,* 2013; Mo *et al.*, 2014; Linden *et al.*, 2008; Berlinski *et al.,* 2013; Banerjee *et al.*, 2007; Imbrogno, 2014), two language arts (Barrera-Osorio *et al.*, 2009; Yang *et al.*, 2013), and two combining language arts and maths (Carillo *et al.*, 2010; Sharma, 2014). Five programmes did not target a specific subject area but rather provided CAL as a general learning strategy (Cristia *et al.,* 2012; David and Quispe 2013; De Melo *et al.*, n.d; Cristia *et al.,* 2013; Humpage, 2013).

Seven programmes also provided a new curriculum in either maths or language arts (Lai *et al.*, 2011; Lai *et al.,* 2013; Mo *et al.*, 2014; Linden *et al.*, 2008; Carillo *et al.*, 2010; Berlinski *et al.*, 2013; Imbrogno, 2014). Three programmes provided a customised curriculum for each child (Carillo *et al.,* 2010; Imbrogno, 2014). Nine programmes made use of the schools' existing curricula, with CAL activities being integrated into lessons.[43] (Cristia *et al.,* 2012; David and Quispe 2013; De Melo *et al.* n.d., Linden *et al.*, 2008; Cristia *et al.,* 2013; Barrera-Osorio *et al.,* 2009,; Banerjee *et al.*, 2007; Sharma, 2014; Humpage, 2013).

Hardware: Thirteen programmes provided either desktop computers or laptops as part of the intervention.[44]. The Vadodara CAL programme in India took advantage of a previously implemented reform by the Government of Gujarat, which provided computers to public primary schools in the city of Vadodara (Banerjee *et al.*, 2007). The MCT interventions in Chile and Mexico relied on computers available in the programme schools (Imbrogno, 2014). Only the intervention in Costa Rica provided additional student materials in addition to computers (Berlinski *et al.,* 2013).

Software: Thirteen programmes provided specific learning software as part of the CAL intervention. The three programmes that did not do so provided computers equipped with basic computer software (Cristia *et al.*, 2013; Barrera-Osorio *et al.*, 2009; Berlinski *et al.,* 2013).

Teacher training: Teachers delivered the CAL to students in all but two programmes, where lessons were delivered by trained external teachers.[45] (Banerjee *et al.,* 2007; Humpage, 2013). Out of the 14 programmes delivered by teachers, 11 programmes provided some sort of teacher training to guide teachers in delivering content using CAL. Only the Gyan Shala programme, both in-school and out-of-school, and the MCT programmes in Mexico and Chile, did not provide any training for teachers. The length of the training for teachers varied from a one-off two-day training to on-going training over a 20-month period (Barrera-Osorio *et al.,* 2009), with the majority of training courses lasting a number of days. One study also included teacher mentoring as part of the intervention (Cristia *et al.* 2013).

Nine of the programmes also provided teacher resources to guide the teachers (Lai *et al.*, 2011b,; Lai *et al.*, 2013, Yang *et al.*, 2013; Mo *et al.*., 2014; Carillo *et al.,* 2010; Berlinksi *et al.*, 2013: Barrera-Osorio *et al.,* 2009; Imbrogno, 2014). For the MCT programmes in Chile and Mexico, these resources consisted of student performance reports generated by the maths software provided (Imbrogno, 2014).

[43] For the Vadodara Programme in India, evaluated by Banerjee *et al.* (2007), it is reported that CAL was not necessarily incorporated in the curriculum, but students played games emphasising competencies in maths.

[44] It should be noted that the Pedagogical Support Pilot Programme (PSPP) in Peru did also not provide hardware in itself. However, since this was an extension of the existing OLPC programme, which provided students and teachers with laptops, the provision of hardware should be considered part of this programme as well. The same logic has been applied to summarising the PSPP's other components (e.g. software).

[45] Under The Pedagogical Support Pilot Programme, which was an extension of the OLPC programme, trained teachers delivered training and workshops to teachers, parents and students. Under the OLPC programme, however, teachers were to deliver the intervention to students.

Additional components: Three CAL programmes in China included monitoring of the programme and teachers by external volunteers (Lai *et al.*, 2011b; Lai *et al.,* 2013; Yang *et al.,* 2013). The Pedagogical Support Pilot Programe (PSPP) in Peru also incorporated teacher monitoring in the form of observation of teachers by external trainers (Humpage, 2013). All the CAL studies in China provided a stipend to the teacher-supervisors who monitored all CAL sessions to ensure they were provided according to the protocol (Lai *et al.,* 2011b; Lai *et al,.* 2013, Yang *et al.*, 2013; Mo *et al.*, 2014). Furthermore, two of the four CAL studies in China provided technical support and free computer repair and maintenance for the entire semester (Lai *et al.,* 2011b; Lai *et al.*, 2013,). Finally, two programmes incorporated parent engagement as part of the programme: The Más Tecnología programme, aimed to ''engage parents in the various activities and stages of the project'' (Carillo *et al.,* 2010), and the Pedagogical Support Pilot Programme (PSPP) included parent workshops (Humpage, 2013). The latter also provided special training and awareness raising about the use of laptops as an educational tool for students.

Finally, seven programmes were designed by the research team (Mo *et al.*, 2014; Lai *et al.,* 2013; Lai *et al.,* 2011b; Yang *et al.,* 2013; Berlinski *et al.,* 2013; Imbrogno, 2014[46]), six by a government (Barrerra-Osorio *et al.*, 2009,; Cristia *et al.,* 2013; De Melo *et al.,* (n.d); Cristia *et al.,* 2012; David and Quispe 2013; Sharma, 2014; Humpage, 2013), and three by an NGO (Banerjee *et al.,* 2005, 2007, Linden *et al.,* 2009; Linden, 2008[47]). The Más Tecnología programme (Carillo *et al.,* 2010) was implemented by the Municipality of Guayaquil but managed by an NGO (E-ducate), who delivered the infrastructure to schools.

Duration: For the studies that report details of the duration of the CAL sessions the duration varied significantly. From two sessions of 40 minutes a week for the CAL programmes in China (Yang *et al.,* 2013, Lai *et al.,* 2011b; Lai *et al.,* 2013, Mo *et al.,* 2014) to two times a week for the OLPC programme in Peru (Sharma, 2014); one hour of daily computer practice for the Gyan Shala programmes (Linden *et al.,* 2008); and two days of a typical school week for the MCT programmes in Mexico and Chile (Imbrogno, 2014).

Four programmes were either national or, at least, large-scale (Cristia *et al.,* 2013; De Melo *et al.,* n.d.; Cristia *et al.*, 2012; David and Quispe 2013). The Más Tecnología and Vadodara programmes were implemented in 155 schools respectively (Banerjee *et al.,* 2007; Carillo *et al.,* 2010). The rest of the programmes were pilots, implemented in a smaller number of schools (n=4-36) (Yang *et al.* 2013; Lai *et al.,* 2011b; Lai *et al.,* 2013, Mo *et al.,* 2014; Linden *et al.,* 2008; Humpage, 2013; Sharma, 2014; Imbrogno, 2014).

Outcomes

All studies apart from one (Cristia *et al.,* 2013) reported on learning outcomes, including maths. Eleven of these studies also measured effects on language arts outcomes (Carillo *et al.*, 2010; Banerjee *et al.,* 2007; Lai *et al.,* n.d; De Melo *et al.*, n.d; Barrera-Osorio *et al.,* 2009; Linden *et al.,* 2008; Lai *et al.*, 2011b; Cristia *et al.*, 2012; Humpage, 2013; Sharma, 2014) and six report effects on composite scores (Linden *et al.*, 2008; Barrera-Osorio *et al.,* 2009, Cristia *et al.,* 2012; Banerjee *et al.,* 2007; Sharma, 2014). Cristia *et al.* (2013) reported on enrolment, drop-out and repetition rates, while Cristia *et al.* (2012), evaluating the OLPC programme in Peru reported enrolment and attendance rates in addition to learning outcomes. Finally, Sharma (2014) and Barrera-Osorio *et al.* (2009) reported dropout and repetition, and attendance rates respectively, in addition to learning outcomes.

---

[46] Although it was not explicitly reported, it seems that the MCT interventions in Chile and Mexico were implemented by the researchers as part of the study.
[47] Linden (2008) represents two programmes: both the in-school and out-of-school Gyan Shala programmes.

Table 6.2 a: CAL Intervention design features

	Cristia *et al.* (2013)	Carillo *et al.* (2010)	Barrera-Osorio *et al.* (2009)	Berlinski *et al.* (2013)	Banerjee *et al.* (2007)	Yang *et al.* (2013)	Lai *et al.* (2011b)	Lai *et al.* (2013)	Mo *et al.* (2014)	Cristia *et al.* (2012)	David and Quispe (2013)	De Melo *et al.* (N.d)	Sharma (2014)	Humpage (2013)	Linden *et al.* (2008)	Linden *et al.* (2008)	Imbrogno (2014)	Imbrogno (2014)
Programme	Huascaran Programme	Más Tecnología Programme	Computadores Para Educatar	Costa Rica CAL Field Study	Vadodara (India) CAL Programme	CAL Programmes China				One Laptop Per Child (OLPC) (Peru, Uruguay, Nepal)				OLPC Pedagogical Support Pilot Programme (PSPP) Peru	Gyan Shala CAL Progamme (In-School)	Gyan Shala CAL Progamme (Out-of-School)	Math Cognitive Tutor (MCT) Mexico	Math Cognitive Tutor (MCT) Chile
Main Design Features																		
Mode of teaching: In-school	✓	✓	✓	✓	✓				✓	✓	✓	✓	✓	✓	✓		✓	✓
Mode of teaching: Out-of-School					✓	✓	✓	✓		✓	✓	✓	✓	✓		✓		
Curriculum: New		✓		✓			✓	✓	✓						✓		✓	✓
Curriculum: Existing (with incorporation of new ICT activities)	✓		✓		✓					✓	✓	✓	✓	✓		✓		
Curriculum: General (not targeting to a specific subject)	✓									✓	✓	✓	✓	✓				
Curriculum: Language		✓	✓			✓							✓					
Curriculum: Math		✓		✓	✓		✓	✓	✓				✓		✓	✓	✓	✓

Materials: Hardware (e.g. desktop computers, laptops) (incl. basic software)	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
Materials: Specific learning software		✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Materials: Other student materials				✓														
Teacher development: Training	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓				
Teacher development: Resources (e.g. guide, handbook)		✓	✓	✓		✓	✓	✓	✓								✓	✓
Teacher development: Teacher mentoring	✓																	
Implementation: Trained teacher delivered intervention					✓									✓				
Implementation: Programme/teacher monitoring						✓	✓	✓						✓				
Other: Teacher stipend						✓	✓	✓	✓									
Other: Technical support							✓	✓										
Other: Parent engagement		✓												✓				
Other: Student training/engagement														✓				

There is an even distribution of different follow-up periods across the studies. Six studies had follow-up periods of less than 12 months (Yang *et al.,* 2013; Lai *et al.* 2011b; Lai *et al.* 2013; Berlinski *et al.* 2013; Imbrogno *et al.* 2013), with the shortest follow up period (3- 4 months) for Lai *et al.*'s (2011b) Migrant School CAL Programme in China. Six studies had follow-up periods 12 to 23 months (Linden *et al.* 2008; Carillo *et al.* 2010; Mo *et al.* 2014; Cristia *et al.* 2012; Humpage, 2013). While another six studies had a follow-up period of 24 months or more (Cristia *et al.* 2013; Barrera-Osorio *et al.* 2009; Banerjee *et al.* 2007; de Melo *et al.* n.d.; David and Quispe 2013; Sharma, 2014), with the longest follow-up period (54 months) for David and Quispe 2013's evaluation of the One Laptop per Child (OLPC) programme in Peru.

Study Design

Fifteen studies were cluster RCTs (Linden *et al.,* 2008; Barrera- Osorio *et al.,* 2009; Carillo *et al.,* 2010, Mo *et al.* 2014, Banerjee *et al.*, 2007, Yang *et al.,* 2013, Lai *et al.*, 2011b, Lai *et al.*, 2013, Berlinski *et al.,* 2013, Cristia *et al.,* 2012; Imbrogno, 2014; Humpage, 2013; Cristia *et al.*, 2013). Three studies use controlled before-after design (David and Quispe 2013; De Melo *et al.*, n.d.; Sharma, 2014).

Comparisons

Most of the included studies compare the effect of an intervention to business as usual (that is, a comparison group with no intervention). One study used a pipeline design (Sharma, 2014).

Table 6.2 b: Characteristics of included studies

Included study	Setting	Description of the intervention	Included outcomes	Follow- up	Study design	Sample Size
Linden *et al.*, 2008	India (rural & urban) Primary School Age: not reported Grade: 2 and 3	Gyan Shala CAL programme (In- school) was designed to complement the maths curriculum. Each child is allocated a particular computer in the common classroom. As part of the in- school treatment arm, students attended the Gyan Shala schools for the normal three-hour period, but worked on the computer-based worksheets instead of participating in the structured Gyan Shala classroom curriculum. Teachers are trained prior to the beginning of the academic year	Learning	12 months	Cluster RCT	Schools: 23 Students: 1640
Cristia *et al.*, 2013	Peru (urban) Secondary School Age: not reported Grade: 1-7 (not 6). Grade 7 only for enrolment outcomes	Huascaran programme Schools selected into the programme received hardware, software, teacher training and they were prioritized to receive internet access. In addition, the programme funded "innovation room coordinators", individuals trained in IT and pedagogy, responsible for ensuring the effective use of computer labs in all subject areas.	Enrolment; Drop-out; Completion	24 months	Cluster RCT	Schools: 6749
Barrera-Osorio *et al.*, 2009	Colombia (Not clear) Primary and Secondary Schools Age: 12.05 (mean for treatment group)	The Computadores Para Educar (CPE): The programme refurbishes computers donated by private organization, installs them in public schools, and runs a programme that teaches teachers to use computers in specific subjects.	Attendance; Learning	24 months	Cluster RCT	Schools: 97 Students: 5201

Carillo *et al*., 2010	Ecuador (not clear) Primary School Age:10.183 (mean for treatment group) Grade: 5	Más Tecnología Programme: Schools in the programme receive basic infrastructure for computer labs and four computers per school. The computers contain software specifically designed to facilitate learning of language arts and maths. The software personalizes the curriculum of each student based on the results of an initial assessment, and students are expected to use the software at least three hours per week. Because the APCI platform is individualized and does not require teachers, it enables students to continue learning outside of the classroom. In addition, teacher training is provided for school principals and at least two teachers in each school. Teachers are able to track the academic progress of each student using the software.	Learning	18 months	Cluster RCT	Students: 724
Mo *et al.*, 2014	China (rural) Primary School Age: 9.74 (mean for treatment group) Grade: 3 and 5	Shaanxi CAL programme: The intervention involved computer assisted maths remedial tutoring sessions designed to complement the regular in-class maths curriculum for the school year 2011-2012. The treatment group had two 40-minute CAL sessions per week as regular classes in school under the supervision of teacher-supervisors trained by the research team. The CAL software provided remedial tutoring material (both animated reviews and remedial questions) in maths for the third and fifth grade students in keeping with the national uniform maths curriculum.	Learning	12 months	Cluster RCT	Schools: 72 Students 4757

Banerjee *et al*., 2007	India (rural) Primary Schools Age: not coded go back to study Grade: 4	The Pratham Computer Assisted Learning (CAL) Programme takes advantage of a policy put in place by the government of Gujarat that delivered four computers to each of the 100 municipal government-run primary schools in the city of Vadodara. Instructors from the local community received a five-day computer training. These instructors provided children with two hours of shared computer time per week; one hour during class time and one hour either immediately before or after school. During that time, the children played a variety of educational computer games, which emphasized basic competencies in the official mathematics curriculum.	Learning		Cluster RCT	Students: 5523
Yang *et al.,* 2013	China (rural) Primary schools Age: 9-11 Grade: 3	Language of Wider Communication (LWC) Programme: The intervention provides computer assisted Mandarin language arts sessions to minority students. The students in the treatment group received two CAL sessions per week of 40-minute during lunch break or after school. Teachers were trained by the researchers to supervise the students during the sessions.	Learning	7 months	Cluster RCT	Students: 1717
Lai *et al.,* 2011b	China (urban) Primary School Age: 8.53 (mean for treatment group) Grade: 3	Migrant School Computer Assisted Learning Program: The intervention consisted of providing computer-assisted maths remedial tutoring sessions to grade 3 students in migrant schools in Bejing. Under the supervision of one trained teacher-supervisor, the students in the treatment group had two 40-minute CAL sessions per week during the lunch break or after school. The content	Learning	3-4 months	Cluster RCT	Students: 2157

		(instructional videos and games) of each session emphasized basic competencies in the uniform national maths curriculum. In short, the material was remedial in nature; was based on the material that was in student textbooks; and was material taught the same week. The students were not supposed to consult the other teams or the teacher-supervisor.				
De Melo *et al.*, N.d	Uruguay (urban) Primary School Age: not reported Grade: 3 (at baseline)	Plan Ceibal: Public primary school students and teachers are provided with a laptop and internet access. Students may take the laptops home.	Learning	Programme rolled out over a period of 24 months	Controlled Before-After	Students: 2080
Lai *et al.*, 2013	China (rural) Primary School Age: not reported Grade: not reported	Shaanxi CAL programme (boarding schools): The intervention involved computer-assisted maths remedial tutoring sessions, which were designed to complement the regular in-class maths curriculum. Under the supervision of two trained teacher-supervisors, students received two 40-minute CAL sessions per week after school. The content (instructional videos and games) emphasised basic competencies in the uniform national maths curriculum. The students were not supposed to consult with other teams or the teacher-supervisor.	Learning	4- 5 months	Cluster RCT	Students: 2613

Cristia *et al.*, 2012	Peru (rural) Primary School Age: 10.809 (mean for treatment at follow-up) Grade: 6-11	One Laptop per Child (OLPC) programme: Students were provided with laptops with a set of applications pre-installed that can be classified into five groups: i) Standard (write, browser, paint, calculator and chat,); ii) Games (educational, including Memorize, Tetris, Sudoku and a variety of puzzles); iii) Music (to create, edit and play music); iv) Programming (three programming environments) and v) Other (including sound and video recording and specific sections of Wikipedia). Teachers were intended to receive training designed to teach them how to use the computers and applications.	Attendance Enrolment; Learning	19-20 months approx	Cluster RCT	Students: 4098
David and Quispe, 2013	Peru (urban) Primary School Age: not reported Grade: 5-6	One Laptop per Child (OLPC) programme: Students were provided with laptops with a set of applications pre-installed that can be classified into five groups: i) Standard (write, browser, paint, calculator and chat,); ii) Games (educational, including Memorize, Tetris, Sudoku and a variety of puzzles); iii) Music (to create, edit and play music); iv) Programming (three programming environments) and v) Other (including sound and video recording and specific sections of Wikipedia). Teachers were intended to receive training designed to teach them how to use the computers and applications.	Learning	4-5 years approx.	Controlled Before-After	Students: 1029

Berlinski *et al.,* 2013	Costa Rica (urban and peri-urban) Secondary School Age:13.1 (average age for 7th grade students in sample schools) Grade: 7	Pedagogical Interventions in Mathematics Teaching – CAL: Two treatment arms (1) new curriculum design and a computer lab; (2) new curriculum design and a laptop for every child in the classroom. The class was structured around a single pedagogical design, independent of technology. The intervention included the design of new pedagogical material, which emphasised exploration, verification, and communication of mathematical facts (rather than lecture-style teaching). In order to achieve this and support teachers and students in this transition, the researchers created a teachers' manual and a students' workbook for each treatment arm. For the three technology interventions, the technology was introduced through a set of applets created on open-source dynamic mathematics software that teachers in Costa Rica were already familiar with (GeoGebra). Teacher training explained how to use the teachers' manual and students' workbook, as well as familiarising teachers with the relevant technology. It also encouraged teachers to adapt their teaching approach. Teachers also received a laptop. Before teacher training started, a virtual classroom was designed and installed to support the work of teachers. Once schools had been notified of their participation in the intervention, meetings were organised with schools to explain how it would be implemented. School principals received a manual containing an intervention description, schedule and information on logistics for implementation.	Learning	Approx. four months	Cluster RCT	Students: 1822

Linden *et al.*, 2008	India (rural & urban) Primary School Age: not reported Grade: 2 and 3	Gyan Shala CAL programme (Out-of-school): As part of the out of school treatment arm, students would arrive either before or after school depending on the shift of their class. When one class was going through its normal three-hour daily schedule, the children from other class took turns working on the CAL package. In this way, the programme supplemented rather than replaced the core Gyan Shala curriculum. Teachers were trained prior to the beginning of the academic year.	Learning	12 months	Cluster RCT	students: 1114
Sharma, 2014	Nepal (Both Urban and rural) Primary and Secondary School Age: Not reported Grade: 2, 3 and 6	One Laptop per Child (OLPC) programme: Each student and teacher was provided with a laptop equipped with course-specific teaching and learning materials based on the national curriculum. Maths and English language classes were expected to be taught using laptop-based teaching and learning approaches at least two times per week. Students were also allowed to take the laptops home. Teacher training lasted 10 days and focused on how to teach different subjects using laptop-based teaching-learning materials.	Learning; Dropout; Completion	26 months	Controlled Before-After	Observations: up to 4757

Imbrogno, 2014	Mexico (unclear if rural, urban, or peri-urban) Secondary school Age: not reported Grade: 7	Math Cognitive Tutor (MCT) Programme: The MCT is an interactive software tool that provides students and teachers with a maths curriculum and performance reports that are personally adaptive for each student. The 'Bridge to Algebra' curriculum consists of 14 units, 57 sections and 552 skills, meaning that students can demonstrate proficiency in many different components separately and gradually build up more complex problem-solving skills. Students are expected use the software about two days a week in (both in the classroom and computer lab) lieu of traditional maths time.	Learning	Approx. 6 months	Cluster RCT	Students: 634
Imbrogno, 2014	Chile (unclear if rural, urban, or peri-urban) Secondary school Age: not reported Grade: 7	Math Cognitive Tutor (MCT) Programme - The MCT is an interactive software tool that provides students and teachers with a maths curriculum and performance reports that are personally adaptive for each student. The 'Bridge to Algebra' curriculum consists of 14 units, 57 sections and 552 skills, meaning that students can demonstrate proficiency in many different components separately and gradually build up more complex problem-solving skills. Students are expected use the software about two days a week in (both in the classroom and computer lab) lieu of traditional maths time. Teachers are trained to use the MCT system.	Learning	Approx. 6 months	Cluster RCT	Schools: 310

Humpage, 2013	Peru (rural) Primary School Age: 7.274 (mean age in treatment group) Grade: 2,4, 6	The OLPC Pedagogical Support Pilot Programme (PSPP) - The PSPP is a participatory training programme aimed at strengthening Peru's OLPC programme by focusing on teachers' ability to use and integrate the XO laptops into their lessons and by creating awareness amongst parents and students about the benefits of the laptops as an educational tool. PSPP therefore included extensive teacher training, workshops and monitoring, as well as parent and student workshops by trained teachers in OLPC schools.	Learning	18 months	Cluster RCT	Students: Up to 588

6.2.2 Synthesis of findings

The results of our synthesis are presented in two sections. First, we present the findings of the meta-analysis on the effects of CAL on primary and secondary outcomes, and any results available for sub-groups. Second, we present the findings of the qualitative synthesis of intervention and implementation features associated with relative success and failure in improving educational outcomes, integrating this with a discussion of the results on effects.

Effects of CAL interventions on primary and secondary outcomes

This section reports the results of the meta-analysis of the effects of CAL, addressing research questions 1a and 1b. We have structured the presentation of results according to the causal chain, starting with access outcomes (enrolment, attendance, completion, dropout), followed by learning outcomes (composite test scores, language arts test scores and maths test scores).

All effect sizes are expressed as standardised mean difference (SMD), interpreted as the magnitude of the number of standard deviation changes in the outcome for the intervention group as compared to students in comparison schools. SMD scores are interpreted as the number of standard deviation changes in the outcome.

None of the studies reported on all outcomes, and for several outcomes only one study contributes effect size data. We were able to conduct meta-analysis for enrolment, attendance, completion, dropout, composite test scores, language arts and maths. The number of comparisons with effect sizes range from two for all access outcomes, to 19 for maths.

Several studies report effects for different sub-groups and we were able to conduct meta-analysis for gender and grade for composite scores, language arts and maths. For maths, we were also able to do sub-group analysis by baseline ability level. We have reported the results for all these analyses, but we have presented the forest plots in the main body of the report only if the results are substantially different from the main sample. Finally, all of the sub-group analyses should be interpreted with caution as they are based on few studies and we cannot rule out systematic reporting biases in the presentation of results by sub-groups.

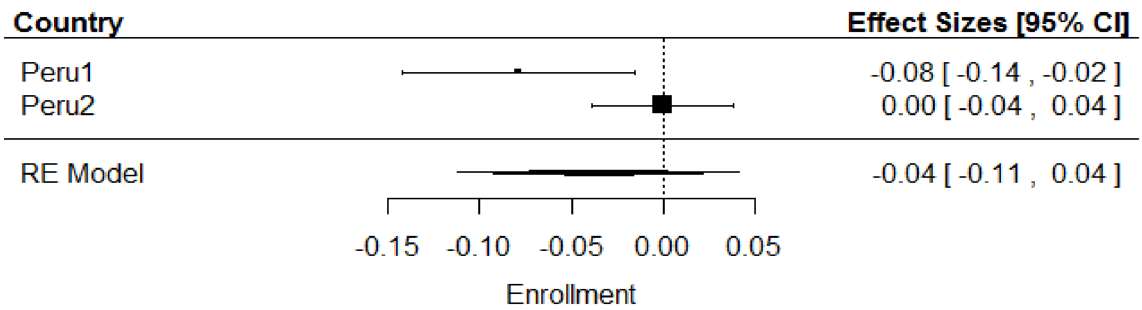
The studies use different follow up periods, from 4 months to 60 months, with the majority of studies measuring follow up at 24 months or less. Two studies report on the same outcome measures at two follow up periods (Banerjee et al., 2007; Sharma et al., 2013). The first follow up was most similar to the other studies in both studies so we included these in the meta-analysis. As for sub-groups, we calculated the SMD for the other follow up periods and comment on these when they are different from the first follow up (results are reported in the table in Appendix H).

A number of the studies include different treatment arms. Berlinksi *et al.*'s study in Cost Rica includes four different treatment arms: (1) new curriculum design; (2) new curriculum design and an interactive whiteboard; (3) new curriculum design and a computer lab; (4) new curriculum design and a laptop for every child in the classroom, all compared to the same comparison group. We included treatment arms 3 and 4 in our analysis of CAL, with arms 1 and 2 classified as pedagogy interventions and reviewed in that section. While these treatment arms rely on the same comparison group we had a sufficient number of studies to use robust variance estimation for the meta-analysis and therefore included both. For Banerjee *et al.*'s study in India only one of the three treatment arms is a CAL intervention, the Vadodara CAL Intervention. We included this treatment arm in the meta-analysis.

Enrolment

Two studies, both from Peru (Cristia *et al.,* 2012; Cristia *et al.*, 2013) but evaluating different programmes, assess the effect of CAL on enrolment rates. The overall average effect is -0.04, 95% CI [-0.11, 0.04]), calculated under a random-effect model. The assessment of homogeneity suggest there is a large amount of variability between the two studies (I^2 = 77.06%, τ^2 = 0.0024, Q(df = 1) = 4.36, p-val = 0.0368). Figure 6.2a presents the forest plot with the results of the individual studies and the pooled point estimate.

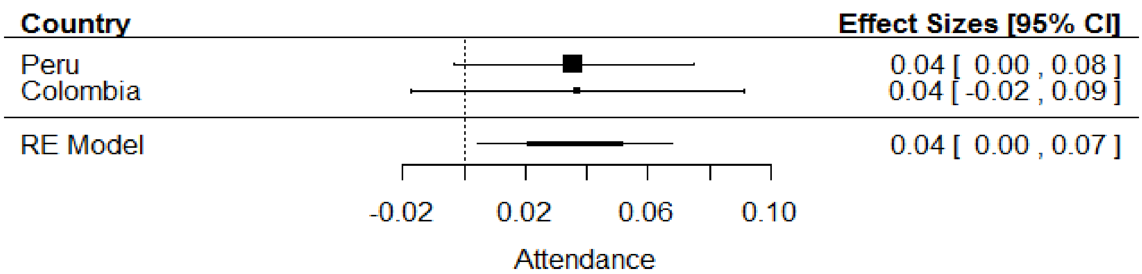
Figure 6.2 a: Enrolment[48]



Attendance

Two studies, from Peru (Cristia *et al.*, 2012; Barrera-Osorio *et al.*, 2009), assess the effect of CAL on student attendance. The overall average effect is 0.04, 95% CI [0.00, 0.07]), calculated under a fixed-effect model. The assessment of homogeneity suggest that the only source of variation is within-study sampling error (I^2 = 0.00%, τ^2 = 0, Q(df = 1) = 0.00, p = 0.9702). Figure 6.2b presents the forest plot with the results of the individual studies and the pooled point estimate, as expected the confidence intervals overlap for these studies.

Figure 6.2 b: Attendance

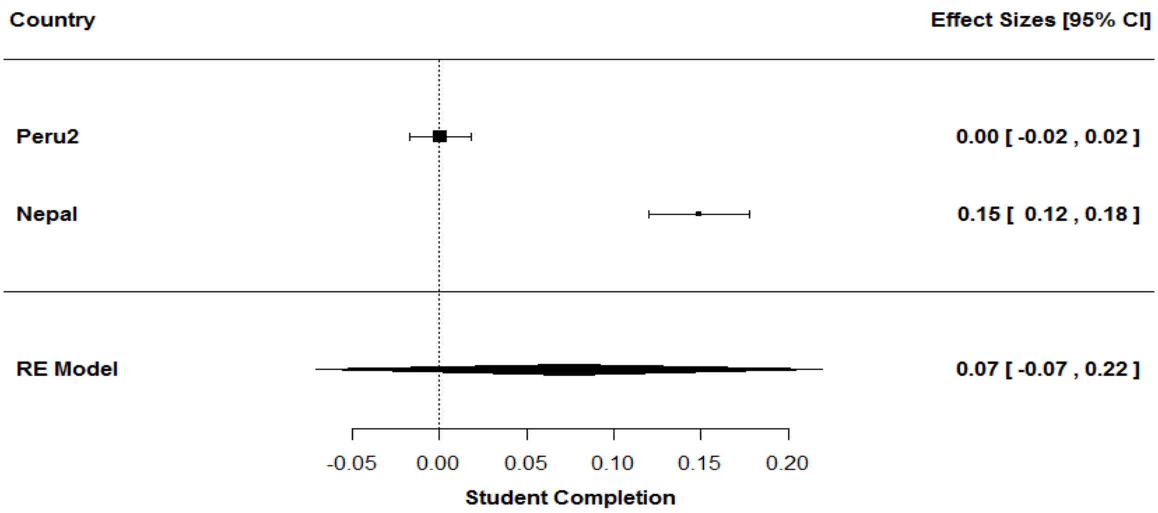


Completion

Two studies, one from Peru (Cristia *et al.* 2013) and one from Nepal (Sharma 2014), assessed the effect of CAL on student completion. The overall average effect is 0.07, 95% CI [-0.07, 0.22], calculated under a random-effect model. The assessment of homogeneity suggest large amount of between-studies variability (I^2 = 98.66%, τ^2 = 0.0108, Q(df = 1) = 74.5943, p-val= < .0001). Figure 6.2c presents the forest plot with the results of the individual studies and the pooled point estimate. As expected, the confidence intervals of the included studies do not overlap.

[48] Peru1 refers to Cristia *et al.* 2012; Peru 2 refers to Cristia *et al.* 2013

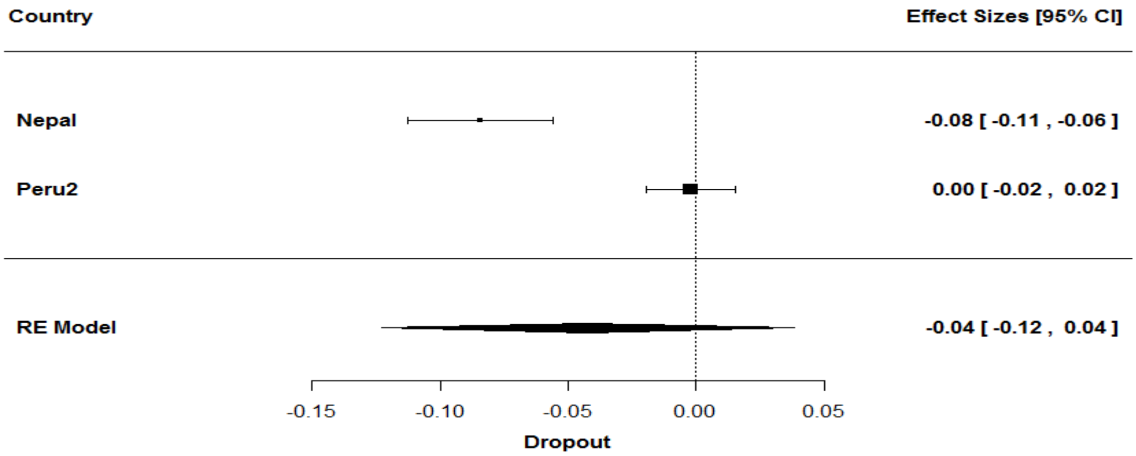
Figure 6.2 c: Completion[49]



Dropout

The same two studies that were included in the analysis above also report effects on drop-out rates. The overall average effect is -0.04, 95% CI [-0.12, 0.04], calculated under a random-effect model. The assessment of homogeneity suggest large amount of between-studies variability (I^2 = 95.77%, τ^2 = 0.0032, Q(df = 1) = 23.6240, p-val < .0001). Figure 6.2d presents the forest plot, and again the confidence intervals of these two effects do not overlap.

Figure 6.2 d: Dropout[50]



Composite test scores

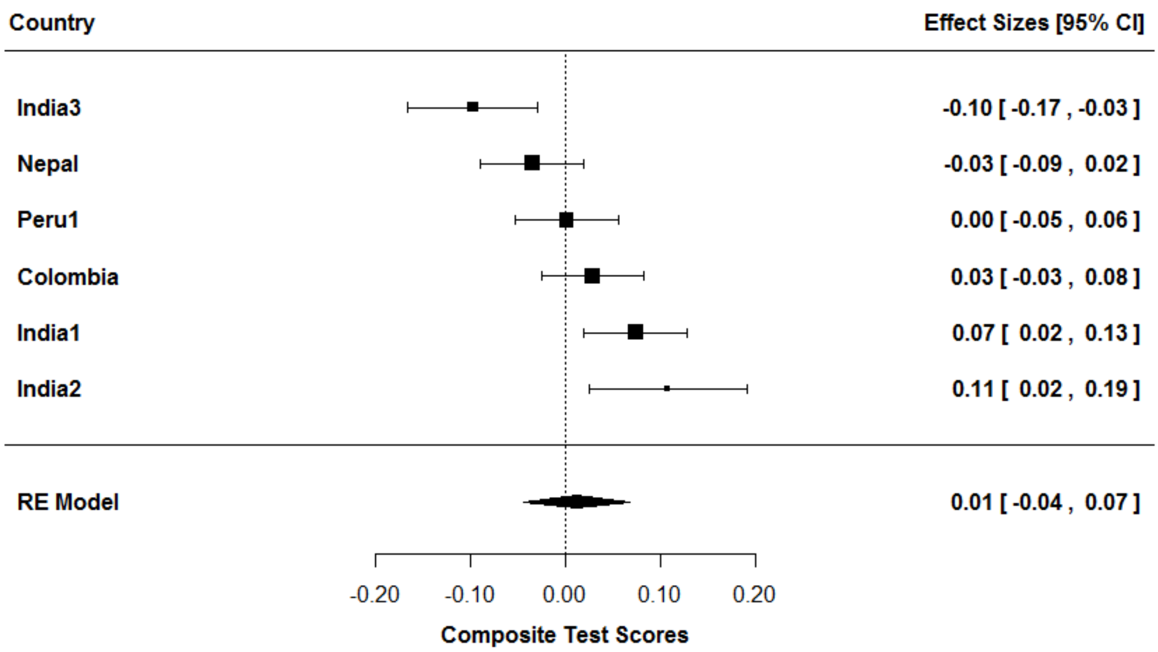
The overall average effect of CAL on learning outcomes as measured by a composite score is 0.01, 95% CI [-0.04, 0.07], calculated under a random effects model. The homogeneity tests (I^2 = 81.13%, τ^2 = 0.0040, Q(df = 5) = 23.3072, p-val = 0.0003) indicate that the effects did not arise from the same population. The forest plot in Figure 6.2e shows several studies without overlapping confidence intervals, supporting the presence of heterogeneity.

[57]Per2 refers to Crista et al.2013

[50] Peru2 refers to Cristia et al. 2013

The effect sizes range from -0.10, 95% CI [-0.17, -0.03] reported for an in-school CAL programme in India (Linden *et al.,* 2008), to 0.11, 95% CI [0.02, 0.19] reported for another study from India, conducted by the same author team a year later (Linden *et al.,* 2008), but with a slight variation in the treatment. Removing the former study from the analysis produces a slight increase in the average pooled effect 0.05, 95% CI [0.00, 0.09] (see Appendix H for results of all sensitivity analyses).

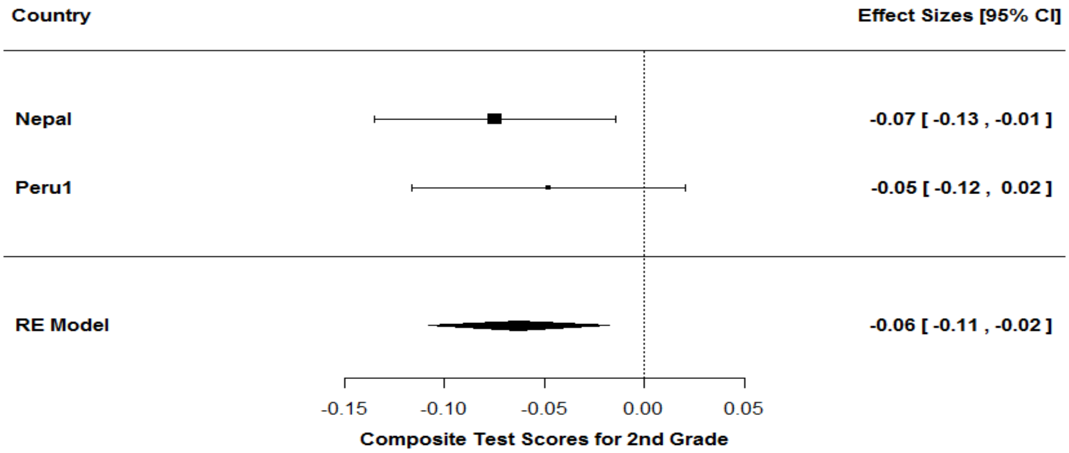
Figure 6.2 e: Composite Test Scores[51]



Two studies report effects for boys and girls separately (Cristia *et al.,* 2012; Barrera-Osorio *et al.,* 2009). The results do not change substantively compared to the main analysis and the average effect for both boys and girls is 0.02 (95% CI [-0.02, 0.06] and 95% CI [-0.02, 0.07]). However the large amount of between study variability observed for the main sample disappears and the assessment of homogeneity suggest suggests no and small amount of between-study variability (Boys: $I^2 = 0.00\%$, $\tau^2 = 0$, Q(df = 1) = 0.4941, p-val = 0.4821; Girls: $I^2 = 18.47\%$, $\tau 2 = 0.0002$, Q(df = 1) = 1.2265, p-val = 0.2681).

[51] India3 refers to Linden et al. 2008 (Gyan Shala in-school programme); Peru1 refers to Cristia et al. 2012; India1 refers to Banerjee et al. 2008; India2 refers to Linden et al.2009 (Gyan Shala out-of-school programme).

Figure 6.2 f: Composite Test Scores 2nd grade[52]



Several studies also report effects for sub-samples of 2nd, 3rd and 6th grade students. The average pooled effect for the 2nd grade sample differs substantially from that of the main sample, suggesting a negative effect on younger children in Nepal and Peru (SMD= -0.06, 95% CI [-0.11, -0.02], I^2 = 0.00%, τ^2 = 0.0004 , Q(df = 1) = 0.3254, p-val = 0.5684). Figure 6.2d provides the forest plot with both the average and individual study effects. The overall effect for the 3rd grade sample is the same as that of the main sample (SMD=0.01, 95% CI [-0.04, 0.06], I^2 = 28.63%, τ^2 = 0, Q(df = 1) = 1.4011, p-val = 0.2365). The average pooled effect for 6th grade students increases only slightly as compared to the main sample (SMD=0.03, 95% CI [-0.07, 0.12]). There is substantial amount of between-study variability for the 6th grade sample (I^2 = 78.33%, τ^2 = 0.0051, Q(df = 2) = 9.6159, p-val = 0.0082) and the result is sensitive to the removal of the study from Colombia. When we remove this study, the point estimate becomes negative, although still small in magnitude (SMD= -0.02, 95% CI [-0.20, -0.01]).

Several studies report estimates for additional sub-groups which we were unable to include in any meta-analysis. The SMD for all these sub-groups are reported in full in Appendix H and they are largely not substantively different from the results reported above. However, Barrera-Osorio *et al.* (2009) report an effect that is much larger in magnitude for 9th grade (SMD = 0.31, 95% CI [0.22, 0.41]) as compared to the full sample, as well as an effect that is much smaller in magnitude for 8th grade (SMD = -0.25, 95% CI [-0.34, -0.15]). Banerjee *et al.* (2007) conduct sub-group analysis by ability level, and report effects that are slightly larger in magnitude than that of the full sample for the bottom third (SMD=0.12, 95% CI [0.06, 0.19] and middle third (SMD=0.10, 95% CI [0.04, 0.17]) of students. Sharma (2014), who also conduct sub-group analysis by ability level report effects that are substantially smaller in magnitude than that of the full sample for both the middle and top quintiles (Middle: SMD =-0.17, 95% CI [-0.27, -0.07]; Top: SMD = -0.11, 95% CI [-0.20, -0.01]).

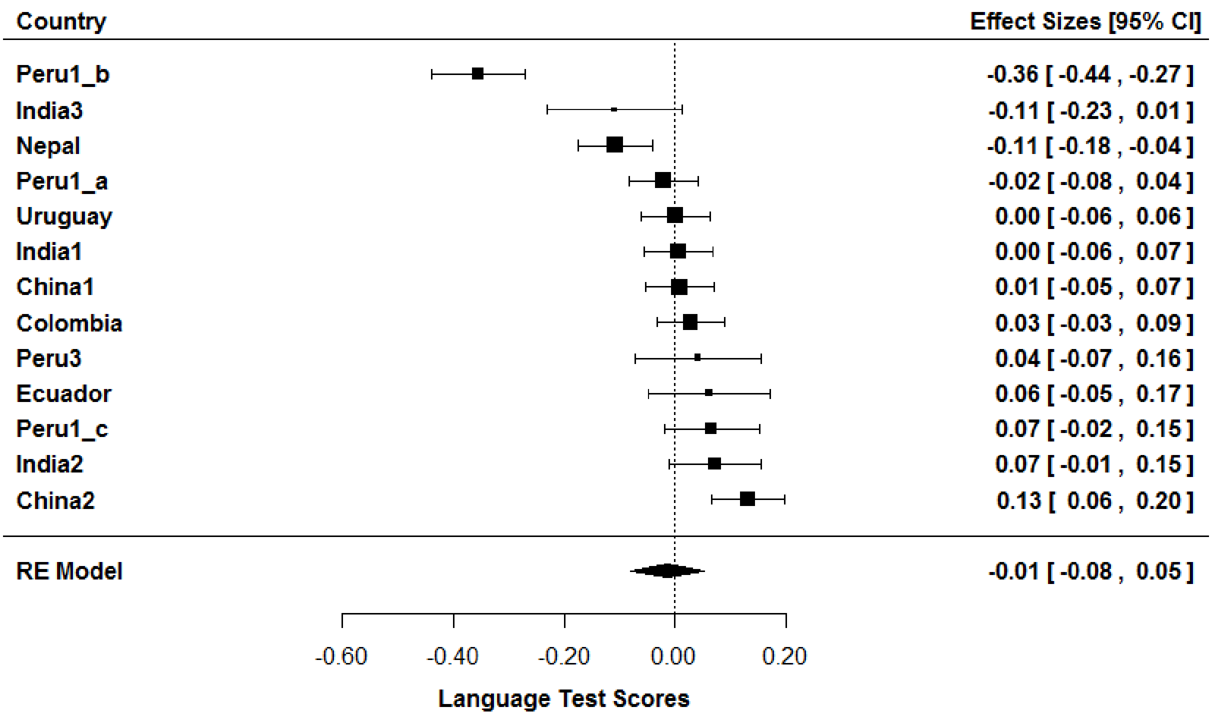
Language arts

Thirteen studies assess the effect of CAL on a language arts test score and the overall average effect is -0.01, 95% CI [-0.08, 0.05] as calculated under a random effects model. There is a large amount of between study variability as indicated by the homogeneity tests (I^2 = 90.08%, τ^2 = 0.0130, Q(df = 12) = 102.1851, p-value < .0001). Figure 6.2g provides the forest plots with the average effect size which supports the presence of large heterogeneity in impact across contexts.

[52] Peru1 refers to Cristia et al.2012

The effects on language arts range from -0.36, 95% CI [-0.44, -0.27] for the One Laptop per Child programme in Peru (Cristia *et al.,* 2012) to 0.13, 95% CI [0.06, 0.20] for the Language of Wider Communication programme in China (Lai *et al.* n.d). Sensitivity analysis suggest the overall result of a small negative effect is robust to the removal of each study, except for the removal of the estimate for the 5th grade sub-sample from the Peru study (see Appendix H for results of all sensitivity analyses). In this case, the overall effect remains small but positive (SMD=0.02, 95% CI [-0.02, 0.05]).

Figure 6.2 g: Language Arts Test Scores[53]



Three studies report effects for boys and girls separately (Cristia *et al.*, 2012; Barrera-Osorio *et al.,* 2009; Sharma). The results do not change substantively compared to the main analysis (Boys: SMD= -0.02, 95% CI [-0.09, 0.05], I^2 = 75.53%, τ^2 = 0.0029, Q(df = 2) = 8.2714, p-val = 0.0160; Girls: SMD= -0.03, 95% CI [-0.12, 0.05], I^2 = 83.47%, τ^2= 0.0049 , Q(df = 2) = 12.0391, p-val = 0.0024). Four studies also report effects on language arts for various sub-groups of 2nd, 3rd, 4th, 5th, 6th, and 8th grade students. The overall average effect of the 3rd grade sample does not differ substantially from the main sample (SMD = 0.01, 95% CI [-0.07, 0.08], I^2 = 54.11%, τ^2 = 0.0016, Q(df = 1) = 2.1790, p-val = 0.1399). The pooled effects for the 4th and 6th grade sample are slightly larger in magnitude than that of the main sample, being 0.03 (SMD=0.03, 95% CI, [-0.16, 0.22] and SMD=0.03, 95% CI [-0.05, 0.11] but remain very small.

The average effects of the 2nd, 5th, 6th and 8th grade samples are negative and larger than that of the full sample. The average effects of the 2nd and 5th grade samples are almost equal (2nd grade: SMD =-0.09, 95% CI [-0.14, -0.04], I^2 = 0.00%, τ^2 = 0.0000, Q(df = 2) = 1.4404, p-val = 0.4867; 5th grade: SMD= -0.08, 95% CI [-0.28, 0.12], I^2 = 86.09%, τ^2 =

[53] Peru1_b refers to Quispe et al. 2013 (Grade 5); India3 refers to Linden et al. 2008 (Gyan Shala in-school programme); Peru1_a refers to Christia et al. 2012 (main sample); India1 refers to Banerjee et al. 2008 (Main sample); China1 refers to Lai et al. 2011b (160 – main sample); Peru3 refers to Humpage 2013 (main sample); Peru1_c refers to Quispe et al. 2013 (Grade 6); India2 refers to Linden et al. 2009 (Gyan Shala out-of-school programme) (main sample); China2 refers to Yang et al. 2013 (105- main sample).

0.0188, Q(df = 1) = 7.1894, p-val = 0.0073). The pooled effect of the 8^{th} grade is even larger and negative at -0.15 (95% CI [-0.32, 0.02], I^2 = 78.81%, τ^2 = 0.0120, Q(df = 1) = 4.7202, p-val = 0.0298), although based only on two studies.

Finally, Sharma (2014) conducts sub-group analysis by ability level, and reports substantially larger negative effects for students in the bottom and middle quintiles (bottom quintile: SMD= -0.28, 95% CI [-0.40, -0.16]; middle quintile: SMD= -0.23, 95% CI [-0.35, -0.11]).

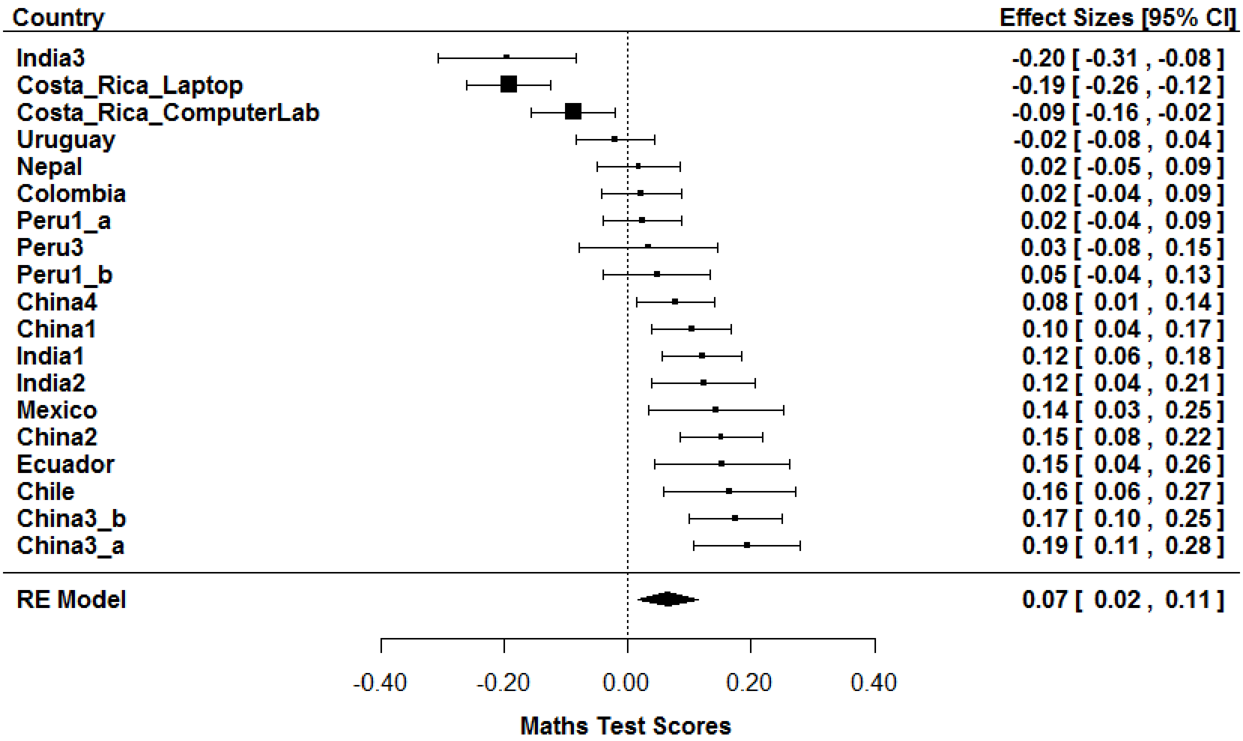
Maths

We included nineteen effect sizes[54] measuring the effect of CAL on maths and the overall average effect is 0.07, 95% CI [0.02, 0.11] as calculated under a random effects model, with robust variance estimation. There is a large amount of between study variability as indicated by the homogeneity tests (I^2=87.23, Q(df = 18) = 140.8256, p-val < .0001). Figure 6.2i supports the presence of heterogeneity.

The effects range from -0.20, 95% CI [-0.31, -0.08] for the Gyan Shala in-school programme in India (Linden *et al.* 2008) to 0.19, 95% CI [0.11, 0.28] reported for the Shaanxi CAL programme in boarding schools (Lai *et al.* 2013). Three of the studies show negative effects with confidence intervals not crossing the line of no effect. There is a group of four studies clustered around zero; and ten of the studies show slightly larger, statistically significant effect sizes. Figure 6.2i presents the forest plot with the results of the individual studies and the pooled point estimates. There are both positive and negative outliers; the confidence intervals between the two groups of studies on either side of the line of no effect do not overlap.

[54] These effect sizes come from 17 different studies evaluating 18 different programmes. Berlinski et al. (2013) evaluate two different treatments as compared to the same control group. Imbrogno (2014) evaluates two programmes: The Math Cognitive Tutor programme in Mexico and the Math Cognitive Tutor programme in Chile. Cristia et al. (2012) and Quispe (2013) evaluate the same programme with an overlapping sample. This creates dependencies between effect sizes. We have used a robust standard error multivariate meta-analysis model to account for this.

Figure 6.2 i: Maths Test Scores[55]



Six studies report effects on maths separately for boys and girls (Linden *et al.*, 2008; Berlinski *et al.,* 2013; Barrera-Osorio *et al.,* 2009; Linden *et al.*, 2008; Sharma, 2014). The overall effect is -0.02, 95% CI [-0.14, 0.09] for boys and -0.06, 95% CI [-0.17, 0.05] for girls, calculated under a random effects model. The large amount of between study variability observed for the main sample remains (Boys sub-group: I^2 = 91.85%, τ^2 = 0.0178, Q(df = 5) = 40.8743, p-val < .0001; Girls sub-group: I^2 = 91.24%, τ^2 = 0.0166 , Q(df = 5) = 40.5315, p-val < .0001).

Figures 6.2j and 6.2k present the forest plots with the results of the analyses for girls and boys respectively. The results are sensitive to the inclusion of the in-school CAL programme in India (Linden *et al.,* 2008), which report relatively large negative effects as compared with the other studies, and the negative effect disappears when this study is removed (see Appendix H for results of all sensitivity analyses).

[55] India3 refers to Linden et al. 2009 (Gyan Shala in-school programme); Costa_Rica_laptop refers to Berlinksi et al. 2013 (full sample, laptop treatment arm); Costa_Rica_ComputerLab refers to Berlinski et al. 2013 (full sample, computer lab treatment arm); Peru1_a refers to Cristia et al. 2012 (Main sample); Peru3 refers to Humpage 2013; Peru1_b refers to Quispe et al. 2013 (Grade 6); China4 refers to Mo et al. 2014; China1 refers to Lai et al. 2011b (160); India1 refers to Banerjee et al. 2008; India2 refers to Linden et al. 2009 (Gyan Shala out-of-school programme); China2 refers to Yang et al. 2013; China3_b refers to Lai et al. 2013 (Grade 6 at second follow-up); China3_a refers to Lai et al. 2013 (Grade 4 at second follow-up).

Figure 6.2 j: Maths Girls[56]

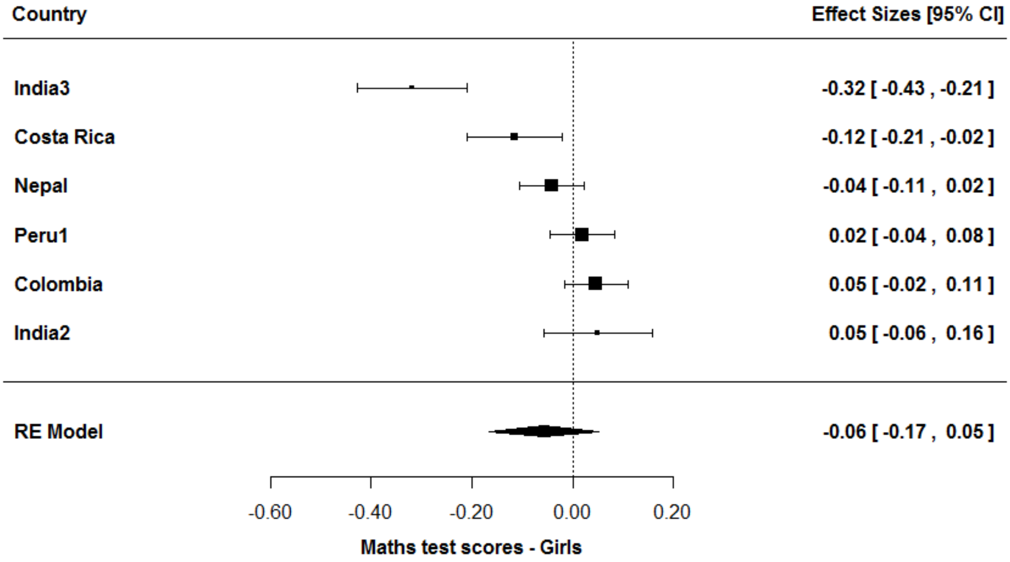
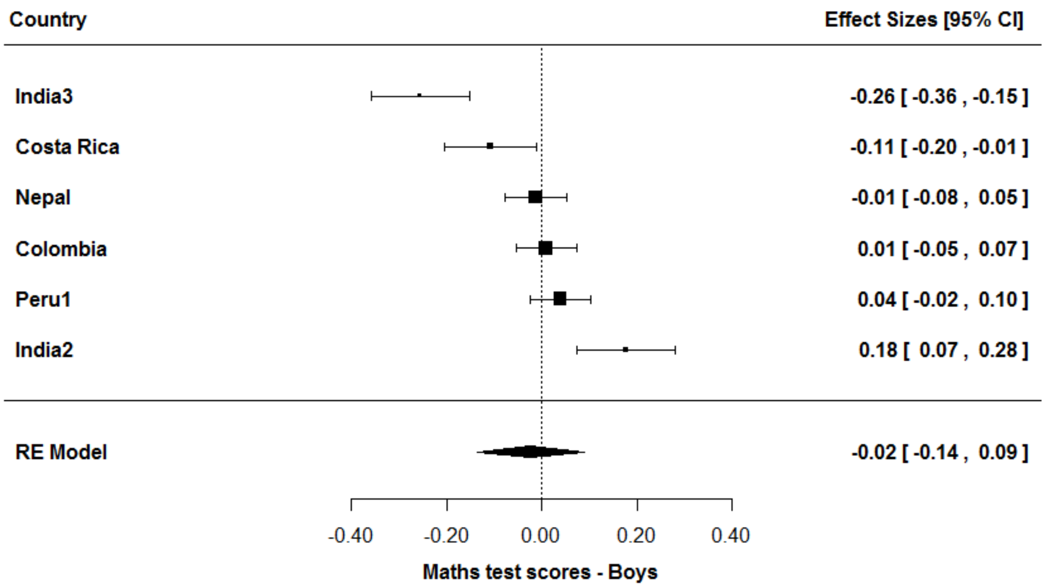


Figure 6.2 k: Maths Boys[57]



Several studies report results by grade sub-groups, allowing for separate meta-analysis for maths test scores for 3rd, 4th, 5th, 6th and 8th grade students. The average effects for the 3rd, 4th and 6th grade samples are equal or similar to the main sample, although they are less precise (3rd Grade: SMD = 0.09, 95% CI [0.02, 0.17]; 4th Grade: SMD= 0.08, 95% CI [-0.05, 0.21]; 6th Grade: SMD = 0.07, 95% CI [-0.04, 0.19]). The pooled effect for Grade 5 is slightly smaller than that of the main sample (SMD = 0.03, 95% CI [-0.04, 0.10].The homogeneity assessment suggest a moderate amount of between-study variability for the four included studies assessing maths scores of grade 5 students (I^2 = 65.71%, τ^2 = 0.0033, Q(df = 3) = 8.4006, p-value = 0.0384).

[56] India3 refers to Linden et al. 2009 (Gyan Shala in-school programme); Peru1 refers to Cristia et al. 2012; India2 refers to Linden et al. 2009 (Gyan Shala out-of-school programme).
[57] India3 refers to Linden et al. 2009 (Gyan Shala in-school programme); Peru1 refers to Cristia et al. 2012; India2 refers to Linden et al. 2009 (Gyan Shala out-of-school programme).

The overall average effect for Grade 8 differs substantially from the main sample in that it is negative (SMD= -0.05, 95% CI [-0.16, 0.05]). However, there are only two studies included for this analysis so it is not representative of the whole sample of CAL studies that report on maths test scores.

Five studies (Linden *et al.,* 2008[58]; Banerjee *et al.,* 2007; Imbrogno, 2014[59]) provide results separately for students grouped by their test scores at baseline. The overall average effects are slightly larger in magnitude for the middle and top terciles than for the bottom tercile (Bottom tercile: SMD = 0.06, 95% CI [-0.17, 0.30]; Middle tercile: SMD= 0.09, 95% CI [-0.14, 0.31]; Top tercile: SMD =0.09, 95% CI [-0.07, 0.25]). Several of the studies measure the effect of CAL on maths test scores for other sub-groups, but we were unable to combine these in any meta-analysis. Table A6.2c in appendix H include the SMD for all of these outcomes.

6.2.3 Summary of findings and discussion

We identified eighteen studies of the effect of sixteen unique CAL programmes implemented in a range of settings in East Asia and the Pacific, South Asia, and Latin America and the Caribbean. We were able to examine effects on enrolment, attendance, completion, dropout, composite scores, maths and language arts using meta-analysis.

The overall average effect range from -0.01 SMD, 95% CI [-0.08, 0.05] for language arts test scores to 0.07 SMD 95% CI [0.02, 0.11] for maths test scores. We identified few studies that evaluated the effects of CAL interventions on access outcomes. The sub-group analyses of test scores suggest some variation in effects, but they are based on a smaller sub-set of studies and it is not clear that these are representative of existing studies. There is also a large amount of between study variability at the outcome level. For example, effect sizes range from -0.41, 95% CI [-0.59, -0.22] for maths scores of bottom tercile students in the Gyan Shala in-school programme in India (Linden *et al.,* 2008) to 0.27 for both maths scores of bottom tercile students in the Gyan Shala out-of-school programme in India (95% CI [0.12, 0.43]) (Linden *et al.,* 2008) and maths scores for middle tercile students in the Math Cognitive Tutor programme in Mexico (95% CI [0.07, 0.46]) (Imbrogno 2014).

The relatively large negative effects observed in some cases is particularly concerning. There are several factors that may help explain these results. Evidence from the qualitative synthesis suggest that programmes in Chile, Colombia, Peru, Uruguay, Mexico and Nepal (Barrera-Osorio *et al.*, 2009; Cristia *et al.*, 2012; Cristia *et al.,* 2013; David and Quispe 2013; De Melo *et al.*, n.d; ; Imbrogno 2014; Sharma, 2014) all faced process and implementation issues. Firstly, several programmes faced technological issues, including insufficient, damaged and dysfunctional equipment, lack of internet access and software not being compatible with hardware (Peru (Huascaran and OLPC programmes); Nepal; Mexico; Chile). Secondly, in both Peru and Nepal it was reported that teachers did not receive sufficient training in delivering the CAL programmes. Finally, findings suggest a lack of integration of the CAL technology into existing learning approaches, with the use of laptops and relevant software reported to either be minimal or unrelated to the curriculum (Chile; Colombia, Peru (Huascaran and OLPC programmes); Nepal; Mexico; Uruguay).

[58] Linden et al. 2008 evaluate both the in- school and the out-of-school programmes in India
[59] Imbrogno 2014 evaluates the Math Cognitive Tutor Programmes in India and Chile

Moreover, Linden *et al.*'s study of two different versions of the Gyan Shala programme in India suggest that whether CAL sessions are instead of or in addition to existing classes may play a role. In this study the in-school version of the programme substituted one of the normal three hours of school curriculum with a CAL session, while the out-of-school version supplemented the normal three-hour curriculum with a CAL session after the normal three hour school day. In the case of the in-school programme the results suggest substantial negative effects on composite, language arts and maths scores, while when CAL was offered as an additional session students were observed to benefit quite substantially. While we did not observe negative effects for all CAL programmes substituting normal school hours with CAL, all programmes where the effect estimates suggest a harmful effect are programmes where CAL was used instead of other approaches. The main issue may not be whether CAL is additional or not, but if the CAL lessons are of a lower quality than the lessons they are replacing this may have a harmful effect on children's learning.

Based on the studies included in our review it is not clear that the overall effect of CAL on children's learning is beneficial. While all average effects are small, for several outcomes they are negative and include negative effects that are quite large in magnitude. Qualitative evidence suggest technological issues, lack of sufficient training for teachers, low use of laptops and a lack of integration of CAL into existing learning approaches may explain some of these results.

Table 6.2 c: Descriptive Findings: process and implementation

Descriptive Findings: Process and Implementation	Context	Citation/ info type
Overall implementation compliance reported as high (e.g. teachers followed instructions, took training, materials were delivered etc.)	Costa Rica, China; Peru (PSPP Programme)	Berlinski and Busso, 2013; Lai *et al.* n.d.; Lai *et al.* 2012; Lai *et al.*, 2013; Mo *et al.* n.d; Mo *et al.* 2014; Humpage, 2013 (all IEs)
Some programmes faced technological issues, including insufficient, damaged and dysfunctional equipment, lack of internet access and software not being compatible with hardware limited	Peru (Huascaran and OLPC programmes); Nepal; Mexico; Chile	Examples of Best Practices in Peru, n.d (Project document); Villanueva-Mansilla, 2012 (Qualitative study); David and Quispe, 2013; Sharma 2014; Imbrogno 2014 (all IEs)
Regional and local government faced a lack of capacity for programme management	Peru (OLPC)	Ministerio de Economía y Finanzas, n.d. (Process evaluation).
Teachers did not receive sufficient training in delivering CAL and there appears to have been a lack of integration of the technology into existing teaching approaches	Peru (Huascaran and OLPC programmes); Nepal; Uruguay	David and Quispe 2014; De melo *et al.* n.d; Cristia *et al.* 2012; Sharma, 2014 (all IEs); Villanueva-Mansilla, 2012 (Qualitative study); Examples of best practices in Peru, n.d. (Project document)
The majority of teachers took part in and passed the training	Costa Rica	Berlinski and Busso, 2013 (IE)
Eighty per cent of parents attended the parent workshops	Peru	Humpage 2013 (Impact Evaluation)
Forty-two per cent of surveyed teachers report use of materials and laptops in non-programme grades	Nepal	Sharma 2014 (Impact Evaluation)
Laptops were used by non-receiving members of the household	Nepal	Sharma 2014 (Impact Evaluation)

Table 6.2 d: Descriptive Findings: Intermediate Outcomes: Computer Use

Descriptive Findings **Intermediate outcomes: computer/ technology use**	**Context**	**Citation/ Info Type**
Use of computers was low among both students and teachers	Colombia	Barrera-Osorio *et al.*, 2009; David and Quispe 2013 (IEs)
Laptops were not used as frequently at home as intended	Peru (OLPC and PSPP programmes)	Cristia *et al.* 2012; David and Quispe 2013; Humpage, 2013 (All IEs)
Students and teachers report high rates of computer usage, but around half of the use appear unrelated to the curriculum	Peru (OLPC), Uruguay	Cristia *et al.,* 2012; de Melo *et al.*, n.d; Ferrando *et al.*, 2011 (IEs)
Students report high rates of use of technology, prescribed software and class materials	Costa Rica	Berlinski and Busso, 2013 (Impact Evaluation)
Positive impact on teachers' computer skills and teachers' laptop use for lessons seen in the third week after programme faded after two years	Peru (PSPP programme)	Humpage 2013 (Impact Evaluation)
Training did not significantly reduce teacher-reported barriers to using the XO laptops	Peru (PSPP programme)	Humpage 2013 (Impact Evaluation)
Programme did not increase laptop use among teachers and students for curricular areas, but did increase use of academic applications	Peru (PSPP programme)	Humpage 2013 (Impact Evaluation)
Students and teachers find educational software useful and easy to use but there was insufficient use both in and outside classroom	Nepal; Mexico; Chile	Sharma 2014; Imbrogno 2014 (Impact Evaluations)
Treatment students report to have better computer skills than control students	Nepal	Sharma 2014 (Impact Evaluation)
The number of programme sections mastered correlates to improved test scores	Chile; Mexico	Imbrogno 2014 (Impact Evaluation)
Infrastructure and implementation school ratings correlate to student MCT usage	Chile; Mexico	Imbrogno 2014 (Impact Evaluation)

Table 6.2 e: Descriptive Findings: Intermediate Outcomes: Motivational Outcomes and classroom activity

Descriptive Findings **Intermediate outcomes:** **motivational outcomes and classroom activity**	**Context**	**Citation/ Info Type**
Students in some CAL programmes report increased levels of motivation and self-efficacy	Uruguay, China	Ferrando *et al.*, 2011; Lai *et al.*, n.d; Lai *et al.*, 2012; Lai *et al.,* 2013; Lai *et al.*, 2011a (IEs)
Students in a CAL programme in Costa Rica report a more active learning environment	Costa Rica	Berlinski and Busso, 2013 (IE)
Seventy-nine per cent of teachers report increase in workload after programme	Nepal	Sharma 2014 (Impact Evaluation)
Students in CAL programme in Chinese boarding schools did not report higher levels of study motivation or self-efficacy	China	Lai *et al.*, 2013 (IE)
The programme did not increase teachers' or students' enthusiasm for the laptops	Peru (PSPP programme)	Humpage 2013 (Impact Evaluation)

6.3 Remedial education

6.3.1 Description of included studies

We included four studies that evaluated the effect of programmes that provide tailored assistance to a group of students. These refer to four unique programmes. The term study is used to refer to a unique output from an author team, which is at times reported in several papers. The characteristics of these studies are described in more detail in the following section.

Setting

The included studies covered two programmes in South America and two in South Asia. Of the two programmes in South America one was implemented in Chile, The Servicio País en Educación, and one in Mexico, an unnamed maths programme (Cabezas *et al.,* 2011; Gutierrez *et al.* 2014). Both the The Balsakhi Programme and the STRIPES programmes were implemented in India (Banerjee *et al.*, 2007; Lakshminarayana *et al.,* 2013). Two programmes were implemented in urban areas, The Balsakhi Programme and the maths programme (Banerjee *et al.,* 2007 and Gutierrez *et al.* 2014) while the STRIPES programme took place in rural areas (Lakshminarayana *et al.*, 2013). The Servicio País en Educación was implemented across several counties and it is likely to be implemented in rural, urban and peri-urban contexts (Cabezas *et al.*, 2011).

Population Characteristics

The majority of the included programmes were targeted at the primary school level only, the Balsakhi Programme, Servicio País en Educación and STRIPES (Banerjee et al,. 2007, Cabezas *et al.*, 2011; Gutierrez *et al.* 2014; Lakshminarayana *et al.*, 2013). One programme was implemented at the secondary school level targeting grade 9 students. The maths programme the Servicio País en Educación targeted grade 4 (Cabezas *et al.*, 2011), and the Balsakhi Programme was targeted at grade 3 and 4 in both cities (Vadodara and Mumbai) (Banerjee *et al.,* 2007). The grade is not reported for the STRIPES programme. Two programmes were implemented in public schools (the maths programme and Servicio País en Educación), one of which was also implemented in private schools (the Servicio País en Educación) (Gutierrez *et al.* 2014 and Cabezas *et al.,* 2011). The type of school that the Balsakhi Programme was implemented in is not clear, however the author makes note that the charity often works in conjunction with municipal schools and it is therefore likely that it was implemented in public schools (Banerjee et al,. 2007).

Interventions

All four programmes provided tailored tutoring with three providing tutoring in smaller groups of children than their usual class (Banerjee *et al.,* 2007; Cabezas *et al.,* 2011; Gutierrez & Rodrigo, 2014). Some programmes were designed for a particular subject, others were designed to target a particular demographic of student and all had a tailored pedagogical approach.

The tutors in the programmes were either volunteers (Cabezas *et al.,* 2011; Gutierrez *et al.* 2014; Lakshminarayana *et al.,* 2013) or hired from the local community (Banerjee *et al.,* 2007). For examples, the Balsakhi Programme hired women from the local community who had completed secondary school. Three programmes provided training to their tutors. The Balsakhi programme provided a two week training session before the school year and tutors were supported throughout their time tutoring (Banerjee et al,. 2007). In the STRIPES programme nominated Community Volunteers were trained by a local NGO, the Naandi Education Research Group (Lakshminarayana *et al.,* 2013). The training provided as part of the maths programme is a four-hour session including access to the maths syllabus and a previous final exam. Tutors were told to look at children's note books, ask them questions and adjust teaching to the group's needs (Gutierrez *et al.* 2014). All programmes focused on core skills, such as literacy and numeracy.

The STRIPES programme also had a community outreach element where implementers informed parents in the community about the programme. Parents of students were also encouraged to enter into verbal contracts with the implementers to ensure that their children attended the additional classes.

Comparisons

All four of the included studies compared the effect of the intervention to business as usual (comparison with no intervention) (Cabezas *et al.,* 2011, Gutierrez et al 2014, Banerjee *et al.,* 2007 and Lakshminarayana *et al.,* 2013). Both Banerjee *et al.* (2007) and Gutierrez *et al.* (2014) use intra-school comparison groups, with the latter also including a comparison group of different schools. In STRIPES both participating and non-participating villages were eligible for another programme, Champion. Both treatment and control groups were randomly selected from these eligible villages.*Outcomes*

The included studies assessed effects on learning outcomes. Three studies report on maths learning outcomes (Banerjee *et al.,* 2007; Gutierrez et al 2014; Lakshminarayana *et al.*, 2013), two report on a composite score (Banerjee *et al.*, 2007; Lakshminarayana *et al.,* 2013) and three report on language arts test scores (Banerjee *et al.,* 2007; Cabezas *et al.*, 2011; Lakshminarayana *et al.,* 2013).

Follow Up

The follow up period for two of the included studies are relatively short. The first (Gutierrez et al 2014) had a follow up period of approximately two months and the second (Cabezas et al 2011) had a follow up period of three months. The Balsakhi Programme was implemented over a period of one year, the follow up time is therefore between 12 and 24 months (Banerjee et al,. 2007) and the STRIPES programme over a provides a 21 months follow up period (Lakshminarayana *et al.*, 2013).

Study Design

Two of the studies were randomised controlled trials (Cabezas *et al.,* 2011, Banerjee *et al.,* 2007). One study was a controlled before and after design with baseline and endline data collection. (Gutierrez et al 2014). The STRIPES programme was a Cluster-RCT (Lakshminarayana *et al.,* 2013).

Table 6.3 a: characteristics of included studies for remedial education

Study ID	Setting	Intervention summary	Included outcomes	Follow- Up*	Study design	Sample Size
Cabezas *et al.*, 2011	Chile. Public & private primary schools. Age: not reported Grade: 4	Students were separated into groups of 5 or 6 with tutors and had group reading sessions using age appropriate traditional stories and informative texts. Students were provided with ten 90 minute sessions. Tutors were volunteers.	Language arts	6 months	RCT	85 Schools
Gutierrez *et al.*, 2014	Mexico (urban). Secondary School. Age: not reported Grade: 9	The maths programme provided two tutors for a group of up to 20 students. They met 2 days a week for 2 hours after school. The average class size was 13.4 children and the programme was focused on marginalised schools.	Maths	2 months	CBA (quasi-experiment with baseline and endline data collection)	60 Schools
Lakshminarayana *et al.*, 2013	India, (rural), Primary School. Grade: not reported Age: not reported	Provided after-school instruction in government primary schools focusing on remedial maths and language arts skills. Parents in the local community were able to select a community volunteer who was trained by the implementing NGO. An outreach programme informed parents about the programme and ensured their cooperation. The lessons reinforced the school curriculum and were tailored to the students' learning levels. Learning materials were also provided.	Maths, language arts and composite scores	21 months	Cluster RCT	3359 students
Banerjee *et al.*, 2007	India (urban) Primary School. Age: not reported Grade: 2 – 3 (Mumbai), 3 -4 (Vadodara).	The Balsakhi Programme provided a female teacher that was hired from the local community. The programme targets under-performing students and includes two hours with the teacher during the school day in group sizes between 15- 20. The curriculum focused on numeracy and literacy skills.	Maths, language arts and Composite score with subgroup analysis by grade	12- 24 months	RCT	Year 1: 14,972 Year 2: 23,160

* Follow-up was calculated as the time between the start of the intervention and data collection, in months.

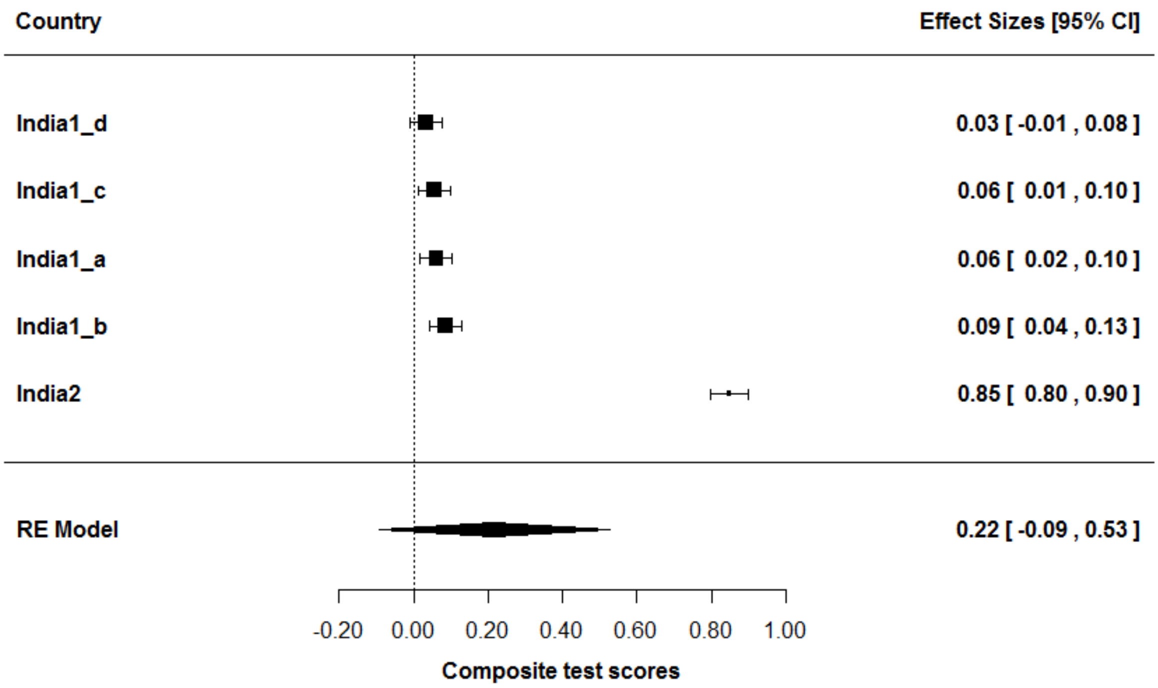
6.3.2 Synthesis of findings

This section reports the results of the meta-analysis of the effects of remedial education interventions, addressing question 1a of the review. All four studies provided data for the meta-analysis, but none of the studies reported on all outcomes. The number of comparisons with effect sizes range from two for composite test scores, three for language arts and maths test scores. For Banerjee *et al.'*s study in India we included four effect sizes each for the composite scores, language arts and maths test scores as the study includes four independent treatment groups. All effect sizes are expressed as standardised mean difference (SMD), interpreted as the magnitude of the number of standard deviation changes in the outcome for the intervention group as compared to students in non-SBM schools. SMD scores are interpreted as the number of standard deviation changes in the outcome.

Composite

The meta-analysis contains five samples from two remedial education studies which took place in India (Banerjee *et al.,* 2007; Lakshminarayana *et al.*, 2013). The overall average effect was 0.22, 95%CI [-0.09, 0.53], but with a large amount of heterogeneity (I^2 = 99%, τ2 = 0.12, Q(4df) = 812.12, p= < 0.0001). Lakshminarayana *et al.* (2013) is a clear outlier, with a large positive effect and without confidence intervals overlapping with the other studies. As expected, the average effect is reduced when removing this study (SMD=0.06, 95% CI [0.04, 0.08]).

Figure 6.3 a: Composite test scores[60]

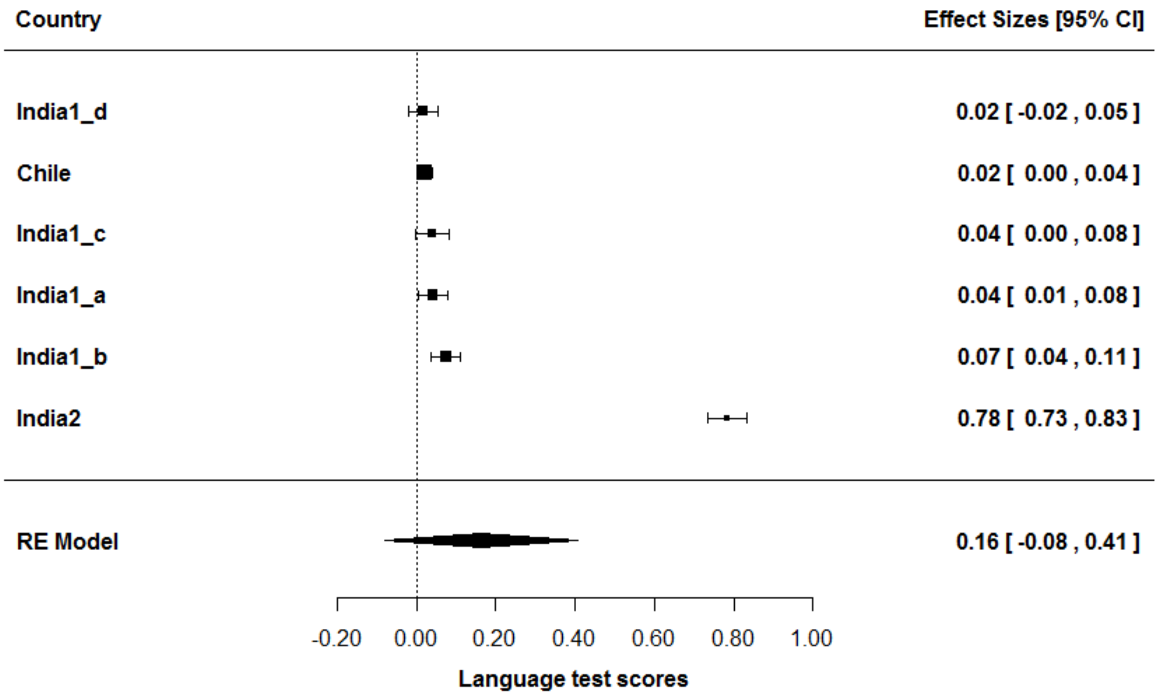


[60] The following four studies measuring composite test scores refer to different samples from Banerjee et al (2007)studies: India1_a = Vadodara, year2; India1_b = Vadodara, year1; India1_c = Mumbai, year 2; India1_d = Mumbai, year1. India 2 refers to Lakshminarayana et al (2013).

Language arts

The overall average effect of remedial education as measured in three studies (with 4 independent samples from one study) is 0.16, 95% CI [-0.08, 0.41]. As with composite test scores, results from the tests of homogeneity suggest a large amount of heterogeneity (I^2 = 99%, τ2 = 0.19, Q(df) = 809.42, p < 0.0001). The study of the STRIPES programme (Lakshminarayana *et al.*, 2013) is a clear outlier, with a large positive effect and without confidence intervals overlapping with the other studies. Sensitivity analysis also shows that the results are sensitive to the removal of this study, reducing the overall effect considerably (Lakshminarayana *et al.*, 2013). Note that Gutierrez *et al.*'s study (2014) of remedial education in Chile reported numerous measures of language arts such text production, reading comprehension and use of language and we combined the measures to create one synthetic effect size for inclusion in the meta-analysis.

Figure 6.3 b: Language arts test scores[61]

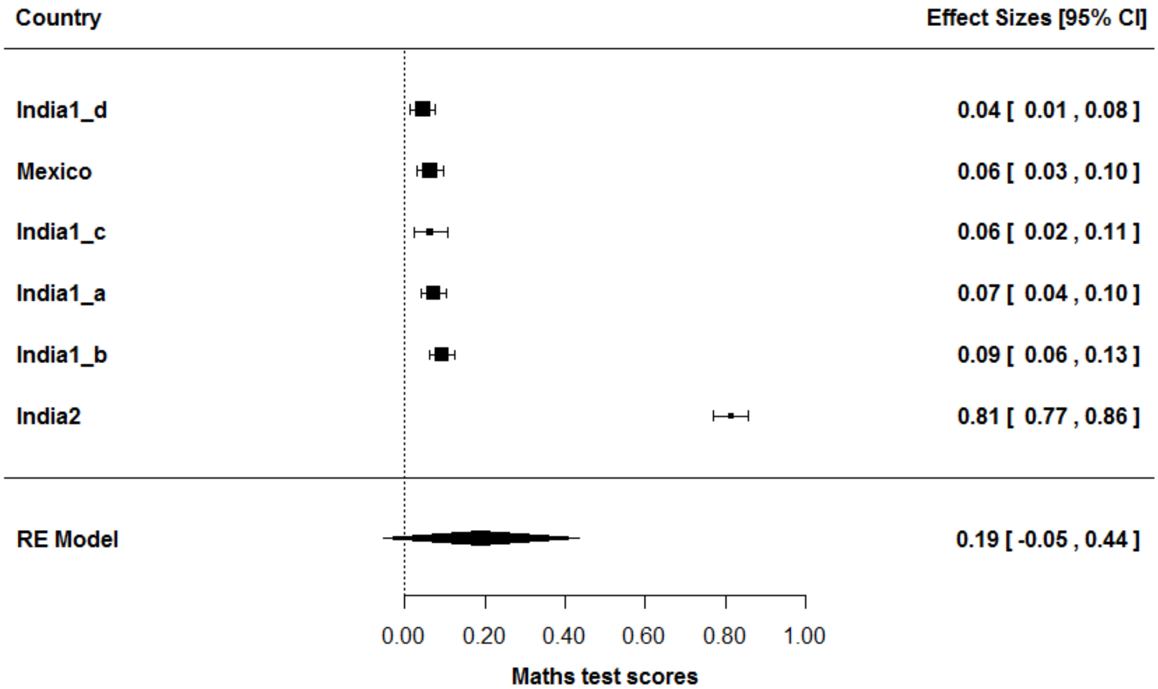


Maths

We included three studies in our main meta-analysis for maths test scores (Gutierrez *et al.,* 2014; Banerjee *et al.,* 2007; Lakshminarayana *et al.*, 2013). The overall average effect of remedial education on maths test scores is 0.19, 95 % CI [-0.05, 0.44]. As above there is a large amount of heterogeneity (I^2 = 99.66%, τ2 = 0.92, Q(5df) = 1017.38, p < 0.0001). Again, the study of the STRIPES programme (Lakshminarayana *et al.,* 2013) is a clear outlier, with a large positive effect without confidence intervals overlapping with the other studies. The overall result is sensitive to the removal of this study, reducing the overall effect to 0.07 SMD (95% CI [0.51, 0.08]).

[61] The following four studies in the language test scores forest plot refer to different samples from Banerjee et al (2007)studies: India1_a = Vadodara, year1; India1_b = Vadodara, year2; India1_c = Mumbai, year 1; India1_d = Mumbai, year2. India 2 refers to Lakshminarayana et al (2013). Chile is Cabezas et al., 2011.

Figure 6.3 c: Maths test scores[62]



6.3.3 Summary of findings and discussion

We included four remedial education studies in our meta-analyses of effects on results for maths, language arts and composite test scores. All the meta-analyses showed an overall positive effect, with the largest effect observed for composite test scores (0.22 SMD, 95%CI [-0.09, 0.53]), with slightly smaller overall effects for language arts (0.16 SMD [-0.08, 0.41]) and maths (0.19 SMD [-0.05, 0.44]). However, there was also a high degree of heterogeneity for all estimates and in each case, the overall effects were sensitive to the removal of Lakshminarayana et al (2013) which had much larger effects for all outcomes than the other included studies.

With such a small sample of studies it is difficult to identify the reason for this heterogeneity and the comparatively large effects observed for the STRIPES programme. However, it appears to have been a more intensive programme than those assessed in the other studies and it was also the only programme that involved a community outreach component. Implementers informed parents in the community about the programme and parents were encouraged to enter into verbal contracts with the implementers to ensure their children attended the additional classes. While only based on a few studies our analysis suggest children attending remedial education programmes benefit from an improvement in test scores on average, with potential for improvements that are relatively large in magnitude, as observed for the STRIPES programme in India. More research is needed to identify the programme components that may produce such large effects and whether these effects can be replicated in different contexts.

[62] The following four studies in the maths test scores forest plot refer to different samples from Banerjee et al (2007) studies: India1_a = Vadodara, year1; India1_b = Vadodara, year2; India1_c = Mumbai, year 1; India1_d = Mumbai, year2. India 2 refers to Lakshminarayana et al (2013). Mexico is Gutierrez et al (2014).

6.4 Extra time in school

Extra time programmes aim to provide a longer school day with increased learning time for students. Typically, these programmes abolish ‘shift’ schooling whereby two separate cohorts attend the same school in a given day, one in the morning and one in the afternoon, and expand existing infrastructure so that all children can attend a full school day.

6.4.1 Description of included studies

We included three studies, reported in four different papers that evaluated the impact of increasing time spent in school. Two of the studies (Bellei, 2009; Valenzuela, 2005) report on the same programme – the Full Day Schooling Programme in Chile. However, they use different datasets, looking at different time periods and different schools, so we included them as two separate studies. The other study in this group looks at a programme in Ethiopia (Orkin, 2013). We have used the term ‘study’ to refer to a unique evaluation of a programme.

Populations

The two studies on the Chilean Full Day Schooling Programme evaluated student outcomes for different samples of students, with Valenzuela assessing fourth grade primary school students, while Bellei assessed ninth and tenth grade secondary school students. The Chilean Full Day Schooling Programme was a national programme implemented in both public and private schools, in urban, peri-urban or rural settings. One of the two studies on this programme covered primary school only (Valenzuela, 2005) while the other covered secondary school only (Bellei, 2009). The Ethiopian programme targeted both rural and urban primary schools.

Interventions

Both extra time interventions examined here aimed to provide students with more time in school actively learning. They did this by moving from ‘shift teaching’ where two shifts of students would attend a school, one in the morning and one in the afternoon, to a full school day. Extra resources were allocated to schools to allow them to expand existing infrastructure and employ teachers to meet the new demand. In the case of the Chilean Full Day Schooling Programme, an increase in the per-student voucher paid to schools by the government facilitated this change (Bellei, 2009; Valenzuela, 2005). For the Ethiopian Longer School Day policy, schools were allocated extra resources to ensure that schools were able to maintain the same per-student resource level as previously (Orkin, 2013). Both interventions aimed to increase the length of the school day by around 30 per cent.

Comparisons, outcomes and study designs

All three studies compared schools that had switched to the new full-day schooling programmes with schools that had not yet changed from the shift system. All three studies provided separate effect sizes for language arts test scores and maths test scores. All three studies were cluster-quasi-RCTs and employed difference-in-difference analysis whereby the average change over time in the outcome variable for the treatment group is compared to the average change over time for the control group. They all had fairly lengthy follow-up periods, with Bellei (2009) following up approximately 72 months after the intervention, Valenzuela (2005) following up after 60 approximately months and Orkin (2013) following up after around 84 months.

6.4.2 *Effects of extra time interventions on learning outcomes*

This section reports the results of the meta-analysis of the effects of extra time in school, addressing question 1a of the review. We identified three studies that evaluated the effects of extra time interventions on our outcomes of interest. Unfortunately, Bellei (2009) did not provide sufficient data for us to calculate effect sizes. As a result we were only able to include the results from the remaining two studies in the meta-analyses for language arts and maths test scores. In cases where more than one measure of test scores were presented (for example, both reading and writing test results), where appropriate we combined the measures to create one synthetic effect size.

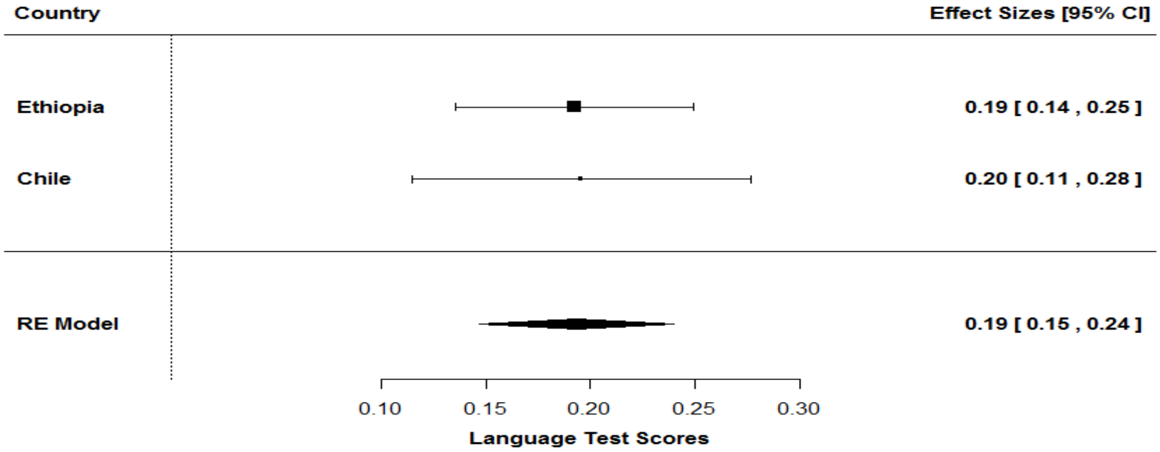
Valenzuela (2009) reports learning outcomes for publicly funded and run schools, which we have included in the meta-analysis. The study also reports results for private schools, further splitting results into those that are partly funded by school fees and those that are not. Orkin (2013) reports outcomes broken down by gender subgroups. Outcome measures not included in the meta-analyses are reported narratively if they provide substantively different findings (all estimates not included in the meta-analysis are reported in technical Appendix H, section 6.6). All effect sizes are expressed as standardised mean difference (SMD), interpreted as the magnitude of the number of standard deviation changes in the outcome for the intervention group as compared to students in non-extra time schools. SMD scores are interpreted as the number of standard deviation changes in the outcome.

Language arts test scores

Figure 6.4a presents the forest plot with the results of the individual studies and the overall average estimate on learning as measured by language arts test scores. The overall average effect of these initiatives on language arts is 0.19, with a 95 per cent confidence interval (CI) [0.15, 0.24], calculated under a random effects model. The assessment of homogeneity suggest the only source of variation is within-study sampling error (I^2 = 0.00%, τ^2 = 0.00, (df = 1) = 0.0039, p-val = 0.9503).

Orkin (2013) provides language arts test scores broken down by gender. The study suggests the effect for boys (SMD = 0.18, 95% CI [0.11, 0.24]) is a little larger than that reported for girls (SMD = 0.14, 95% CI [0.07, 0.20]). For the Full Day Schooling programme in Chile (Valenzuela, 2009), the estimates for private schools that are partly funded by school fees (0.36, 95% CI [0.27, 0.45]) and those that are not (0.34, 95% CI [0.24, 0.43]) are both notably higher than that for publicly funded and run schools (included in the meta-analysis).

Figure 6.3 d: Language Arts test scores[63]

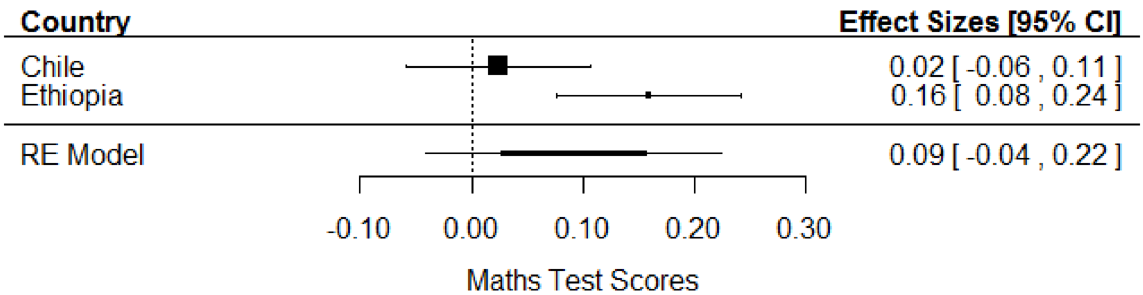


Maths test scores

The overall weighted average effect of extra time interventions on maths test was 0.09, 95% CI (-0.04, 0.22), calculated under a random effects model. Figure 6.4b suggests variability between the two studies and this is supported by the homogeneity tests (I^2 = 80.33%, τ^2 = 0.0074, Q(df = 1) = 5.0836, p-val = 0.0242).

Unlike with language arts test scores, there is no substantive difference between the maths test scores for boys and girls. For the Full Day Schooling programme in Chile, Valenzuela, (2009) also reports language arts test results for schools that are partly funded by school fees (0.24, 95% CI [0.14, 0.34]) and those that are not (0.15, 95% CI [0.6, 0.26]), finding that the effects for maths test scores larger in magnitude when compared to the effect reported in the meta-analysis for publicly funded and run schools.

Figure 6.3 e: Maths test scores[64]



6.4.3 Summary of findings

The effects of increasing the time children spend in school by extending the school day appear promising. The effect on language arts is larger in magnitude than that observed for most other interventions (SMD= 0.21, 95% CI [0.12, 0.30]). The average effect on maths is smaller in magnitude and less precise 0.09, 95% CI [-0.04, 0.22]). However, the findings are based on only two studies and the results should therefore be interpreted with caution. Future studies should assess whether these effects can be replicated in different contexts.

[63] Ethiopia: Orkin, 2013; Chile: Valenzuela, 2005
[64] Ethiopia: Orkin, 2013; Chile: Valenzuela, 2005

6.5 Providing materials

Interventions providing materials can assist teachers, facilitate learning and improve educational quality. Such interventions include any intervention providing ‘traditional hardware’ material such as books, chalkboards or other classroom equipment. For instance, the School Assistance Programme (SAP) funded by the Dutch non-profit organisation International Christelijk Steunfonds (ICS), provided English, Maths and Science text books to primary school children in Kenya (Glewwe *et al.,* 2009).

6.5.1 How may Providing Materials effect education outcomes?

The lack of sufficient and appropriate school materials is thought to significantly undermine the performance of education systems. Programmes providing schools with materials such as blackboards, textbooks and notebooks aim to improve education outcomes by addressing supply side determinants of educational quality (Farrell & Heyneman, 1989; Glewwe *et al.*, 2008; Hunt, 2008). Such programmes may improve outcomes in a number of ways. Increased availability of learning materials can help children engage with the curriculum and promote self study. They can also improve the quality of teaching by assisting teachers in delivering their lessons (Krishnaratne *et al*., 2013). Finally, additional materials may increase schooling expectations among students and parents, and therefore their motivations to enrol or stay in education, or attend classes more often (Hunt, 2008).

However, for materials to improve outcomes a number of steps and assumptions need to hold. First materials need to be procured and distributed to schools, which assumes there is sufficient local infrastructure. Headmasters and teachers then need to distribute and use materials to support lessons. But this assumes they are receptive to the new materials, allocate sufficient time to their use and incorporate them into lessons. Finally, for materials to be used by students and improve outcomes they need to be at the appropriate learning-level and mother tongue of the students (Tan *et al.*, 1999; Glewwe *et al.,* 2008). Figure 6.5a summarise a basic programme theory for programmes providing materials.

.

Figure 6.5 a: Providing materials programme theory

Inputs	Activities	Outputs	Intermediate Outcomes	Final Outcomes
(1) Pedagogical materials (e.g. textbooks, teachers guides, desks, blackboards, pens, notebooks, computers) (2) Staff to distribute materials and provide appropriate guidance on their use (3) Financial resources	(1) Purchase and distribute materials; (2) Promote use among students and teachers	(1) Increase in availability of learning materials per student (2) Increase in use of learning materials by teachers and students	(1) Improved engagement of students with learning materials and curriculum (2) Improved teacher performance	(1) Improved learning outcomes (e.g. test scores) (2) Improved school participation
	Assumptions: (1) Infrastructure to support procurement and distribution	*Assumptions:* (1) Intervention is adequately implemented	*Assumptions:* (1) Headmasters and teachers are receptive of new materials and allocate sufficient time to their use, incorporate them into the curriculum and use them appropriately. (2) Quality of education plays a role in parent & student decision-making	*Assumptions:* (1) Sustained and long-term use of the materials (2) The materials are appropriate to the level and mother tongue of the students

6.5.2 Description of included studies

We identified four studies reported in six different papers that evaluated the effect of the provision of learning materials on education outcomes in L&MICs. These papers referred to four unique programmes and were published between 2000 and 2014. The characteristics of these studies are described in detail in Table 6.6a in the following section.

Population

All studies look at the outcomes of these programmes at the primary school level. While one study looked at public schools (Das *et al.,* 2013), and another at government-assisted community schools (Sabarwal *et al.,* 2014), the school type of the other two studies was unclear (Glewwe *et al.,* 2004; Glewwe *et al.,* 2009). The four studies covered a range of grades from grade two through to eight.

Three programmes were targeted at underserved populations. Both programmes in Kenya were implemented by ICS International in Busia and Teso districts, which have below-average income levels and average KCPE scoresroughly at the median for Kenya as a whole. One of these programmes targeted schools deemed the neediest (Glewwe *et al.,* 2009). The other programme targeted schools that had roughly similar mean scores as the district as a whole since previous ICS programmes had already assisted the worse off schools (Glewwe *et al.*, 2004). The programme in Sierra Leone targeted government, government-assisted and community schools that had not previously been targeted by other programmes, and which had a 3:1 student-textbook ratio or higher.

Setting

The programmes evaluated by the studies covered both South Asia and Sub-Saharan Africa. Two of the evaluated programs were located in Kenya (Glewwe *et al.*, 2004; Glewwe *et al.,* 2009), one in Sierra Leone (Sabarwal *et al.,* 2014), and one in India (Das *et al.,* 2013). Three of the four studies took place in a rural setting (Glewwe *et al.,* 2004; Glewwe *et al.*, 2009; Das *et al.*, 2013). For one study, this was not clear (Sabarwal *et al.*, 2014).

Intervention

The interventions provided either learning materials directly, or a grant for the school to purchase learning materials. The programme in Sierra Leone provided a set of core textbooks for every child in the treatment schools (Sabarwal *et al.,* 2014). One of the Kenyan programmes distributed mainly flip charts to the selected schools (Glewwe *et al.*, 2004). The programme in India provided a per-pupil school grant to be spent on materials used directly by pupils.[65] (Das *et al.,* 2013). Finally, the second Kenyan programme had two treatment arms: one providing textbooks, the other providing grants to purchase educational materials including textbooks (Glewwe *et al.*, 2009). Both of these treatment arms are included in the same analysis, the two treatments can therefore be regarded as one intervention including both components.

Comparisons

Three studies compared the effect of an intervention to business as usual (a control group with no intervention) (Glewwe *et al.*, 2004; Das *et al.,* 2013; Sabarwal *et al.*, 2014). One study implemented a pipeline design where two different treatment arms were rolled out over time. The first treatment group (provision of textbooks) were first compared to the remaining 3 groups that had not yet received assistance. The researchers then compared the first

[65] The grant was typically spent on notebooks, writing materials, workbooks, stationery (Das et al., 2013)

(provision of books) and second treatment group (provision of grant to purchase learning materials), and later also the 3rd treatment group to the remaining control schools. We include the effects for using the business as usual comparison in our analysis.

Outcomes

The included studies reported on a variety of education outcomes. All four studies evaluated learning through tests in maths and language arts. Although all studies report outcomes for both subjects separately, one study additionally provides a composite score of both subjects (Das *et al.,* 2013). Another study provides a composite score that includes maths, language arts and science (Glewwe *et al.,* 2009). Three of four studies use English as their measure of language arts. English is an official language in all three countries where these studies were implemented (Glewwe *et al.,* 2004; Glewwe et al, 2009; Sawarbal *et al.,* 2014).

Two studies assessed languages other than English. Glewwe *et al.* (2004) measured Kiswahili, an official language in Kenya where the study was implemented, in addition to maths and English language. Das *et al.* (2013) measured Telugu, the official language in Andhra Pradesh, addition to math. Two studies provide an additional sub-group analysis for learning outcomes by grade (Glewwe *et al*. (2004; Glewwe *et al.* (2009). Sabarwal *et al.* (2014) also reported on enrolment and attendance rates, including by gender sub-groups and Glewwe *et al.* (2009) evaluated drop-out and completion rates.

Three of four programmes in this sub-category had similar lengths of follow up. Sabarwal *et al.* (2014), Glewwe *et al.* (2009) and Das *et al.* (2013) respectively measured outcomes approximately 19[66], 17 [67]and 21[68] months after the start of their respective programmes. Glewwe *et al.* (2009) was rolled out over a period of approximately four years before the end line assessment took place, so that the schools were exposed to the programme for a period of between one and four years. [69]

Study Design

All included studies were cluster randomised control trials where assignment to the intervention took place at the school level.

[66] The programme started in the beginning of the 2008-2009 Academic year. Sierra Leone’s academic year starts in September (http://www.epdc.org/sites/default/files/documents/EPDC%20NEP_Sierra%20Leone.pdf). The end line survey was conducted in December 2009 (Sabarwal *et al.* 2014).
[67] Flip charts were distributed in February 1997. Endline data was collected in July 1998 (*Glewwe et al.*, 2007, 2009).
[68]) The Intervention was implemented in August 2005. The end line survey was conducted in March-April 2007 (Das *et al.* 2013).
[69] *“In Early, 1996, 25 schools were randomly selected from these 100 to receive textbooks in early 1996 (Grades 3-7 received English text books, and grades 3, 5 and 7 received maths textbooks). In early 1997, maths textbooks were given to grades 4 and 6, and agriculture textbooks to grade 8. In 1997, another 25 of the 100 schools were selected to receive grants equal to US$2.65 per student, or on average US$727 per school. ..More schools received similar grants in early 1998 and 2000"* (Glewwe *et al*., 2007:5-7). The end line data was collected in 2000.

Table 6.5 a: Characteristics of included studies: Providing materials

Included studies	Setting	Intervention summary	Included outcomes	Follow- Up	Study design	Sample Size
Glewwe *et al*., 2004	Kenya (rural) Primary school Grade: 6-8 Age: not reported	Flipchart study. Science, health, and maths flip charts appropriate for grades 5-8 were provided to schools in addition to a teachers' guide for science and a wall map for Geography. Each chart covered twelve different aspect of the topic. The charts are brought into the classroom when they are relevant to the day's lesson, meaning that they can be used for more than one grade.	Learning	18 months	Cluster RCT	Up to 25,060 students
Sabarwal *et al.,* 2014	Sierra Leone (urban, rural) Primary school Grade: 4 & 5 Age: not reported	Government textbook distribution programme. In 2008, the Government of Sierra Leone distributed textbooks to schools on the basis of student enrolment rates in order to provide a set of core textbooks for every child in programme schools. The actual transportation of books from central warehouses to schools was undertaken by local service provides who were competitively selected by the government.	Enrolment; Attendance; Learning	Approx. 15 months	Cluster RCT	Up to 325 schools

Included studies	Setting	Intervention summary	Included outcomes	Follow- Up	Study design	Sample Size
Glewwe *et al.*, 2009	Kenya (rural) Primary school Grade:3-8 Age: 15 (average)	The Schools Assistance Programme (SAP). A hundred schools were randomly divided in four treatment groups of twenty-five schools. Group 1 received textbooks from early 1996. These textbooks covered various subjects (English, Math, Agriculture) and were provided to different grades ranging from 3-8. Pupils in grades 3-5 could not take textbooks home, but pupils in grade 6-8 were put in pairs to share textbooks and were supposed to be able to take the textbook home on alternate days. Group 2 received grants equal to to US$2.65 student or US$727 per school. Group 3 and 4 received similar grants in early 1998 and 2000 respectively.	Drop-out; Completion; Learning	12 months; 24 months	Cluster RCT	Up to 10, 080 students
Das *et al.*, 2013	India (rural) Primary school Grade: 2-5 Age: not reported	The Andhra Pradesh (AP) School Block Grant Experiment. Schools received a grant worth around US$3 per pupil to be spent only on materials used by students directly including notebooks, writing materials, workbooks, stationery, etc.). Although schools had the freedom to decide how to spend the grant, they received guidance in doing so from the implementing organisation, who also distributed the materials. In the first programme year the grant was announced two months into the school year, meaning that it was unanticipated by schools and households. In the second year of the programme the grant was anticipated.	Learning	19-20 months	Cluster RCT	Up to 13926 students

6.5.3 Synthesis of findings

Here we present the findings addressing our main research questions. First, we present the findings of the meta-analysis on the effects of providing materials interventions on education outcomes. Second, we present a discussion of the findings incorporating evidence from our descriptive qualitative synthesis of intervention design, implementation and context factors that may influence the effectiveness (Questions 2a and 2b).

Effects of providing materials interventions on enrolment, attendance, dropout rates, completion and learning outcomes

None of the studies reported on all outcomes, and for several outcomes only one study contributes effect size data. We were able to conduct meta-analysis for language arts, maths and composite test scores. Individual studies also report on enrolment, attendance, drop-out, completion, teacher attendance and teacher performance. We report the SMDs for these outcomes in a narrative at the end of the results section. All outcomes available in the included studies, but not included in the meta-analysis are provided in table A6.6 in appendix H.

All effect sizes are expressed as standardised mean difference (SMD), interpreted as the magnitude of the number of standard deviation changes in the outcome for the intervention group as compared to students in non-intervention schools. SMD scores are interpreted as the number of standard deviation changes in the outcome.

Teacher attendance and performance

Sabarwal *et al.* (2014) is the only study assessing teacher outcomes, reporting an effect of 0.09, 95% CI [-0.02, 0.2] for teacher attendance and 0.19, 95% CI [0.08, 0.3] for teacher performance.

Access outcomes

Sabarwal et al (2014) also measure effects on enrolment and attendance. The authors report effects of -0.06, 95% CI [-0.21, 0.95] for enrolment (main sample); 0.02, 95% CI [-0.15, 0.18] for grade 4 attendance; and 0.09, 95% CI [-0.09, 0.27] for grade 5 attendance. The authors of this study also presented results by gender for each grade for attendance. Notably, the effect for Grade 5 girls was much larger in magnitude than for all other groups (SMD=0.24, 95% CI [0.05, 0.41]).

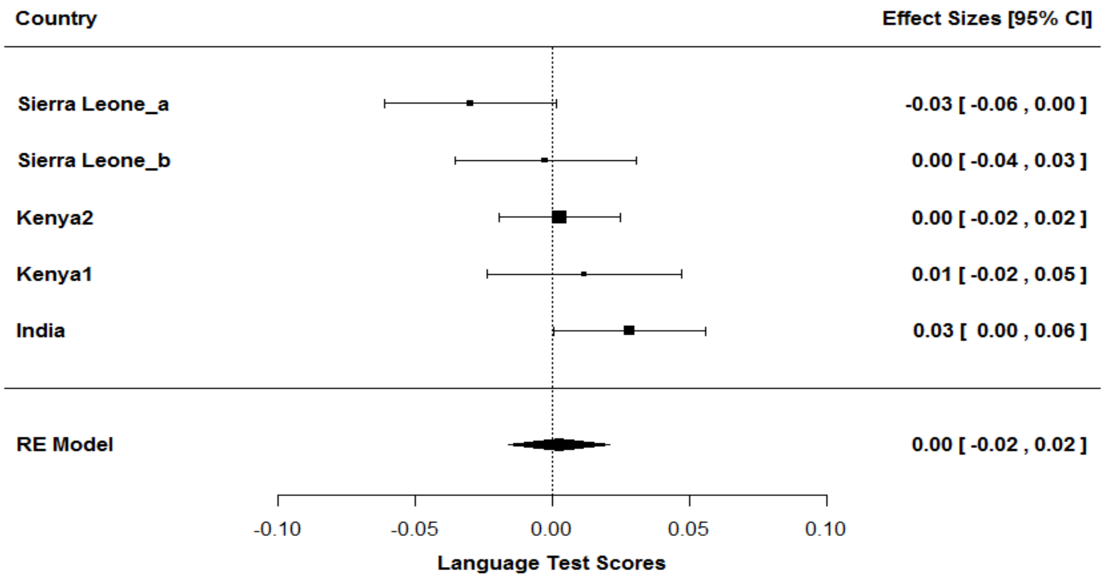
Glewwe et al (2009) assess the effect of Kenyan SAP programme on drop-out and completion rates across year one and year two of the programme. The effects are very small across all measures (range: SMD=-0.03, 95% CI [-0.05, 0.00] and 2 (SMD=-0.03, 95% CI [-0.05, 0.00]) of the study.

Language arts

We included results from five independent samples across the four included studies which reported on language arts test scores.[70] The overall average effect of materials on language arts scores was zero (95% CI [-0.02, 0.02). The assessment of homogeneity suggests a moderate amount of between studies variation (I^2 = 49.10%, τ^2 = 0.0002, Q(df = 4) = 7.74, p = 0.10). The effect sizes range from -0.03, 95% CI [-0.06, 0.00] in Sierra Leone (Sabarwal *et al.,* 2014) to 0.03, 95% CI [0.00, 0.06] in India (Das *et al.,* 2013). Results of the sensitivity analysis show that the effect is not sensitive to the removal of any of the studies (Appendix H provides full details of all sensitivity analyses).

[70] There are five samples over 4 studies in this forest plot as we treated the Grade 4 and 5 samples from Sarbawal et al (2014) separately.

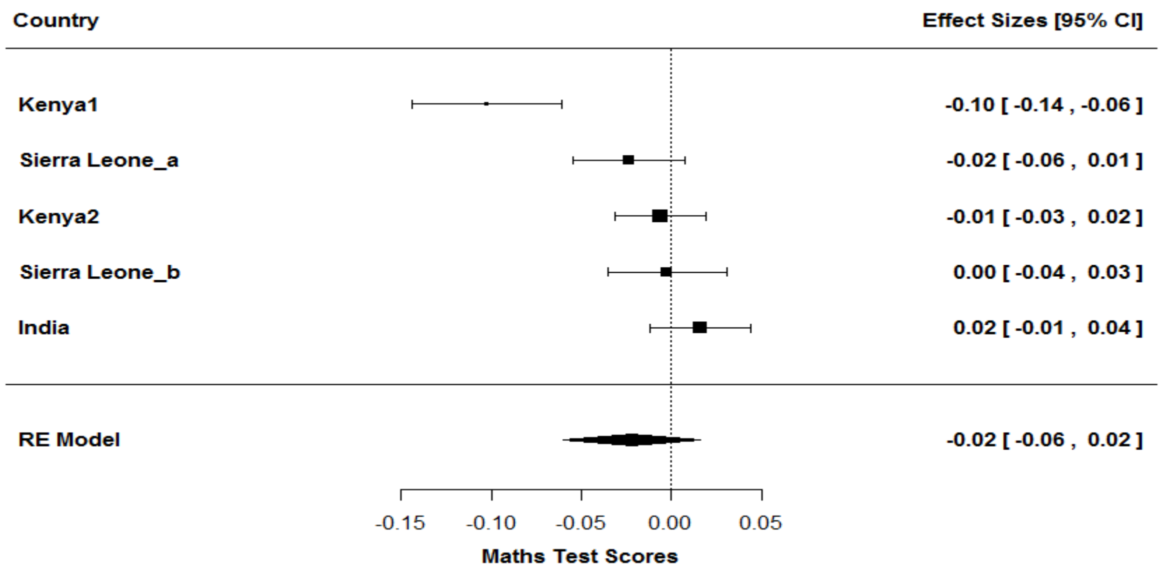
Figure 6.5 b: Language Arts Test Scores[71]



Maths

The overall average effect of providing learning materials on maths scores is -0.02 (95% CI [-0.06, 0.02]). The test for homogeneity showed a large amount of variation between studies (I^2 = 86.78%; τ^2 = 0.0013, Q(df = 4) = 6.06, p = 0.1945). The effects range from -0.10, 95% CI [-0.14, -0.06] in Kenya (Glewwe *et al.*2009) to 0.02, 95% CI [-0.01, 0.04] in India (Das *et al.* 2013). As can be seen from Figure 6.6a, the confidence intervals of most of the most studies overlap, apart from one study from Kenya which reports an effect larger in magnitude than the other studies. The overall effect is also sensitive to the removal of The Year 1 data from Glewwe et al (2009) giving an overall effect of -0.00 (CI [-0.02, 0.01]. The overall effect is not sensitive to the removal of any one study (see Appendix H).

Figure 6.5 c: Maths Test Scores[72]



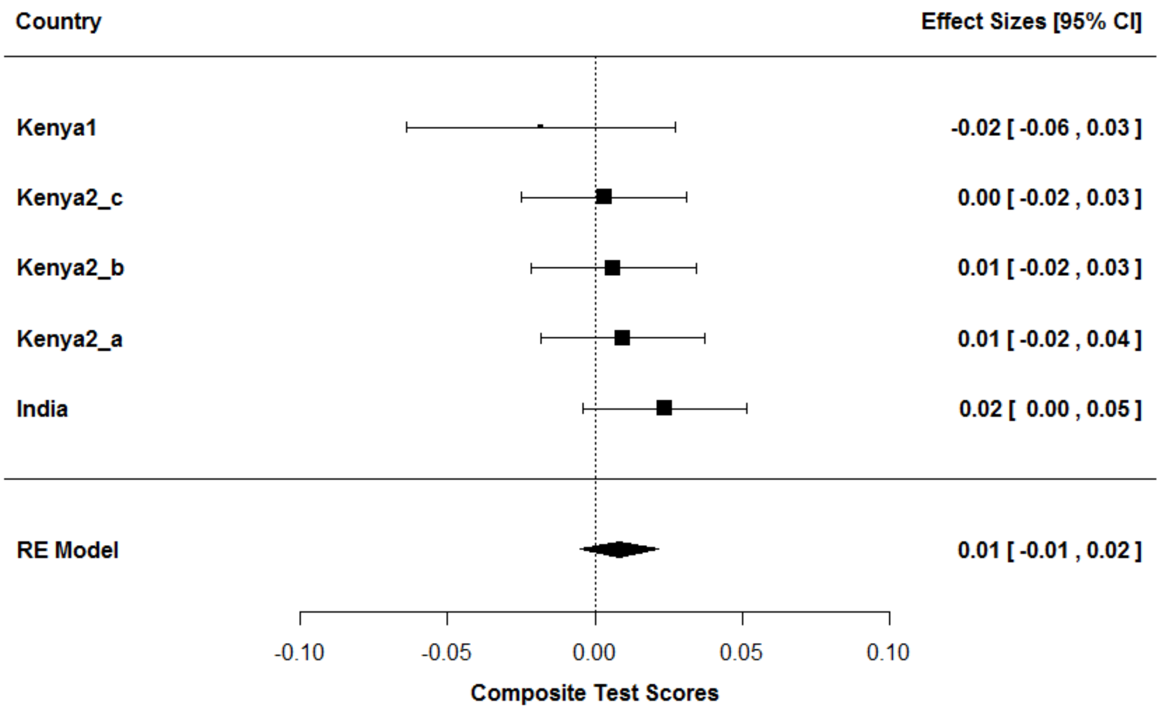
[71] Sierra Leone_a refers to Sabarwal *et al.* 2009 (Grade 5 sample); Sierra Leone_b refers to Sabarwal *et al.* 2009 (Grade 6 sample); Kenya2 refers to Glewwe *et al.* 2004; Kenya1 refers to Glewwe *et al.* 2009;

[72] Kenya1 refers to Glewwe *et al.* 2009; Sierra Leone_a refers to Sabarwal *et al.* 2009 (Grade 5 sample); Kenya2 refers to Glewwe *et al.* 2004; Sierra Leone_b refers to Sabarwal *et al.* 2009 (Grade 6 sample)

Composite scores

The overall average effect of providing learning materials on composite test scores is 0.01, 95% CI [-0.01, 0.02]. The homogeneity test indicated no between studies variability (I^2 = 00.00%, $\tau^2 = 0$, Q(df = 4) = 2.64, p = 0.6204). The effects range from -0.02, 95% CI [-0.06, 0.03] in Kenya (Glewwe *et al.,* 2009) to 0.02, 95% CI [0.00, 0.05] in India (Das *et al.,* 2013).The results are robust to the removal of any one study (see appendix H for all sensitivity analyses).

Figure 6.5 d: Composite scores[73]



6.5.4 Summary of findings and discussion

We identified four studies of programmes that provided learning materials across three different countries in South Asia and Africa. We were able to examine effects on maths, language arts and composite test scores using meta-analysis. The overall average effect ranges from -0.02, 95% CI [-0.06, 0.02] for maths, to 0.01, 95% CI [-0.01, 0.02] for composite test scores. While there is a large amount of heterogeneity, there is a similar pattern of small or negative effects across learning outcomes.

Why do programmes that provide materials appear to have limited effect on children's learning outcomes? While there was limited evidence on implementation, the included studies do provide some possible explanations for this lack of effect, summarised in table 6c (full synthesis in Appendix J). Several of the studies report issues with implementation. Many schools in both Sierra Leone and India did not receive the text books that were provided as part of the programme (Sabarwal *et al.,* 2014; Das *et al.,* 2013). There were also suggestions that the increase in supply of materials might have been less than intended as schools spent funds on other items, such as classroom construction (Glewwe *et al.,* 2009) and parents adjusted their contribution in response to the grant programme (Das *et al.,*

[73] Kenya1 refers to Glewwe *et al.* 2009 ; Kenya2_c refers to Glewwe *et al.* 2004 (Grade 8 sample); Kenya2_b refers to Glewwe *et al.* 2004 (Grade 7 sample); Kenya2_a refers to Glewwe *et al.* 2004 (Grade 6 sample);

2013). In Sierra Leone books were often kept in storage and not distributed to students. In both Sierra Leone and Kenya authors report an improvement in teacher's performance and use of text books, although this does not appear to have translated into better outcomes for children. One possible reason for this suggested in one of the studies is that the text books, which were in English (the third language of most students) might have been too difficult for most students (Glewwe *et al.,* 2009).

Overall the results suggest that children in schools receiving materials or grants for the purpose of buying materials do not do any better than children in schools not receiving such materials. This finding is based on few studies and should be interpreted with caution. Many of the other interventions assessed at the school level do suggest beneficial effects of programmes that include the provision of materials together with other components. So one interpretation may be that in addition to there being issues with poor implementation, materials are a necessary but not sufficient condition for children's learning.

Table 6.5 b: Descriptive findings - providing materials

Descriptive findings: Process, implementation and context	Context	Citation (Info type)
Many schools did not receive text books	Sierra Leone, India	Sabarwal *et al.,* 2014; Das *et al.*, 2013 (Impact Evaluations)
Half of the grant was spent on text books, the remainder on classroom construction	Kenya	Glewwe *et al.*, 2009 (Impact Evaluation)
The majority of the grant was spent on stationary, classroom materials and practice materials	India	Das *et al.*, 2013 (Impact Evaluation)
Parents adjusted their contribution to school inputs in response to grant programme for schools to purchase materials	India, Kenya	Das *et al.*, 2013; Glewwe *et al.*, 2009 (Impact Evaluation)
About a third of students still reported they did not have access to textbooks	Kenya	Glewwe *et al.*, 2009 (Impact Evaluation)
While books were reportedly provided to teachers, most books were kept in storage and not distributed to students (Sabarwal *et al.*, 2014).	Sierra Leone	Sabarwal *et al.*, 2014 (Impact Evaluation)
There was an increase in supply of learning materials	India	Das *et al.,* 2013 (Impact Evaluation)
Teachers reported high levels of use and familiarity with the flipcharts	Kenya	Glewwe *et al.,* 2004 (Impact Evaluation)
There was an increase in teachers' use of textbooks, but this did not reach more than sixty-two per cent at its peak	Kenya	Glewwe *et al.,* 2009 (Impact Evaluation)
The programme increased the likelihood of teachers having a lesson plan and the likelihood of teachers teaching	Sierra Leone	Sabarwal *et al.*, 2014 (Impact Evaluation)
The programme had a small impact on teachers encouraging the use of textbooks	Sierra Leone	Sabarwal *et al.,* 2014 (Impact Evaluation)
The text books, which were in English (the third language of most students) might have been too difficult for most students	Kenya	Glewwe *et al.*, 2009 (Impact Evaluation)
Decentralisation process at time of programme lead to confusion on chain of command and roles and responsibilities of different agents in programme implementation	Sierra Leone	Sabarwal *et al.*, 2014 (Impact Evaluation)

6.6 New schools and infrastructure

New schools and infrastructure (NSI) interventions are programmes involving the construction of schools in areas where there were none previously, or improvement or rehabilitation of existing school infrastructure. In this category, we also include interventions to provide access to clean water for drinking and washing, safe waste disposal and separate toilets for girls to remove health related barriers to schooling as well as tackle incidents of harassment and humiliation in school toilets (Birdthistle *et al.*, 2011).

6.6.1 How might new schools and infrastructure programmes affect education?

There are two main channels through which NSI are expected to improve education outcomes. Firstly, by improving access to schooling. Increased availability of schools in areas that previously had none the distance children have to travel is reduced, removing one of the most important barriers to education (Burde & Linden, 2009). It may also provide an opportunity to attract better teachers and reduce teacher absence rates if the community is able to provide an adequate and accessible working environment (Levy *et al.,* 2009). Secondly, by improving the learning environment through better facilities students' learning experiences and expectations of schooling might be improved (Hunt, 2008), potentially leading to higher enrolment and attendance rates and lower drop-out rates. The provision of separate toilet facilities for boys and girls for example might be important to ensure girls retention in school (Colclough *et al.*, 2000).

6.6.2. Description of included studies

We included seven studies reported in ten different papers that evaluated the effect of new schools and/or school infrastructure programmes on educational outcomes in L&MICs. These papers were published between 2002 and 2013 and referred to seven unique programmes. Table 6.6a provides an overview of the characteristics of these studies, as summarised below.

Population

Four of these studies looked at the outcomes of these programmes at the primary school level (Burde and Linden, 2011, 2013; Freeman *et al.*, 2012; Borkum *et al.*, 2013; Dumitrescu *et al.*, 2011), and one at the primary and secondary school levels (Newman *et al.,* 2002). Adukia's (2014) evaluation of the School Sanitation and Hygiene Education Programme (SSHE) covered children in primary and upper primary school in India. For one study, the school level is not clear (Lokshin & Yemtsov, 2004). Three of the studies provided information on the grades in which students were assessed. Adukia (2014) assesses students in grades 1-8, Freeman *et al.* (2012) assesses students in grades 4-8 and Borkum *et al.* (2013) report the mean grade of students in their sample as grade 4.

Setting

The programmes evaluated by the studies cover a range of settings in South Asia, Sub-Saharan Africa, Latin America and the Caribbean, and Europe and Central Asia. Two of these programmes were located in India (Adukia, 2014; Borkum *et al.*, 2013); one in Afghanistan (Burde & Linden, 2013); one in Kenya (Freeman *et al.,* 2012); one in Bolivia (Newman *et al.,* 2002); one in Georgia (Lokshin & Yemtsov, 2004); and one in Niger (Dumitrescu *et al.,* 2011) . Four studies took place primarily in rural settings (Newman *et al.,* 2002; Burde & Linden, 2013; Lokshin & Yemtsov, 2004; Dumitrescu *et al.,* 2011). One study took place in an urban setting (Borkum *et al.*, 2013). Finally, for two studies it was unclear whether they took place in a rural, urban or peri-urban setting (Freeman *et al.*, 2012; Adukia, 2014).

Interventions

The interventions evaluated in the studies included for this category are very diverse, and so we grouped them into three broad sub-categories for the purposes of analysis.

Two studies included the provision of latrines and were classified as hygiene infrastructure interventions (Adukia, 2014; Freeman *et al.*, 2012). Adukia (2014) evaluated widespread school latrine construction implemented as part of the Total Sanitation Campaign (TSC) in India. Freeman *et al.* (2012) also evaluated latrine construction, implemented as part of a broader sanitation intervention also including hand washing, drinking water containers and one-year supply of WaterGuard.

The second category of studies focused on the establishment of new, community-based schools in underserved areas. We identified two studies that evaluated such a program. Firstly, the PACE-A programme in Afghanistan (Burde & Linden, 2013), which included school construction and management once the school was built. The schools were mostly a single, multi-grade class since few children in the project villages had access to education prior to the establishment of the schools.[74] Secondly, the IMAGINE programme in Niger built schools that were based on a prototype that included three classrooms, housing for three female teachers, a preschool, and separate latrines for boys and girls equipped with hand washing stations.

Finally, we identified three studies that reported on the improvement or replacement of school infrastructure, including the provision of a library (Borkum *et al.,* 2013) and rehabilitation of existing infrastructure[75] (Lokshin & Yemtsov, 2004; Newman *et al.,* 2002). The Akshara Library Programme rehabilitated existing libraries in a hub-and-spokes system by providing them with age and language-appropriate material and training librarians who would in turn provide reading-focused educational activities (Borkum *et al.*, 2013). The school rehabilitation projects included the repair of existing infrastructure such as roofs, windows, and floors, replacing inner pipes, installing sanitary and heating equipment and repainting walls (Lokshin and Yemtsov, 2004). The Social Investment Fund (SIF) education projects in Bolivia repaired and replaced desks, blackboards, and playgrounds in existing schools (Newman *et al*., 2002).

Comparisons

The majority of the included studies compared the effect of an intervention to business as usual. The PACE-A programme in Afghanistan was a pipeline design, in that the comparison villages were scheduled to receive new schools the following year after the evaluation took place (Burde & Linden, 2013), so was the only study with a 'pure' (no intervention) control. The Niger study also compared groups according to whether or not they received a new girl-friendly school through the IMAGINE program[76] (Dumitrescu *et al.,* 2011). The comparison

[74] This was a large scale programme implemented in 1,672 communities in 97 districts and 19 provinces in Afghanistan. The programme was implemented by four partner organisations (CRS, CARE, International Rescue Committee, Aga Khan Foundation) and as such it could arguably be classified as a public private partnership programme. However, as the purpose of the programme was to transfer the schools to government control once the government has the capacity to manage the schools we classified this as 'new schools'.

[75] This included rebuilding school buildings at schools already operational.

[76] The IMAGINE program, in theory, consisted of a package of "soft" interventions as well as the construction of girl-friendly school and so the evaluation initially aimed to measure the effect of having the school on top of the additional package of interventions. However, the soft interventions were only partially implemented so, in practice, the study measures the impact of having the girl-friendly schools against (close to) business as usual (Dumitrescu *et al.,* 2011). We have taken this into account elsewhere in our analysis (see appendix J.).

groups in the remaining studies were standard access to schooling. An experiment in Kenya included multiple treatment arms (Freeman *et al.,* 2012). In the first treatment arm, teachers in the chosen schools received training on hygiene promotion, behaviour change and water treatment methods plus hand washing and drinking water containers and a one-time, one-year supply of point-of-use water disinfectant. The second group of schools received these components as well as the provision of latrines. We included the second treatment arm that provided latrines in our meta-analysis.

Outcomes

The majority of the included studies measured two or more education outcomes. Four studies used some measure of learning achievement (Adukia, 2014; Borkum *et al.*, 2013; Burde and Linden, 2013; Dumitrescu *et al.,* 2011). Apart from Adukia (2014), who reported the scores of middle school state-board exams, the other studies evaluated learning through tests in maths and language arts. In addition, Burde and Linden (2013) provided a composite score of the two subjects.

Five studies reported on enrolment rates (Adukia, 2014; Burde & Linden, 2013; Lokshin & Yemtsov, 2004; Newman *et al*., 2002; Dumitrescu *et al.* 2011); six on attendance (Burde and Linden, 2011, 2009; Freeman *et al.,* 2002; Newman *et al.*, 2002; Borkum *et al.*, 2013; Dumitrescu *et al.* 2011; Adukia, 2014 – as measured through attendance of an exam); two on dropout rates (Adukia, 2014; Newman *et al.,* 2002); and two on completion rates (Lokshin & Yemtsov, 2004; Newman *et al.*, 2002).

The follow up period between the start of the intervention and the end-line outcome survey was 17 months for the hygiene promotion experiment in Kenya (Freeman *et al.*, 2012) and 23 months for the IMAGINE programme in Niger (Dumitrescu *et al.* 2011). For the remaining five included studies we only have estimates of the follow up period. These vary between approximately 4 months for the first follow survey in the PACE-A programme in Afghanistan (Burde & Linden, 2013), up to approximately 60 months for the SIF education investment programmes in Bolivia (Newman *et al.*, 2002).

Study Design

Four studies were cluster-randomised control trials where assignment to the intervention took place at the school or village level (Borkum *et al.*, 2012; Burde and Linden, 2011, 2013; Freeman *et al.*, 2012; Dumitrescu *et al.*, 2011). Newman *et al.* (2002) contained two different study designs evaluating the same programme taken from different areas of Bolivia; a cluster RCT undertaken in the Chaco region of the country, and a controlled before and after study for the Resto Rural region that used propensity score matching to equalise group differences between treatments and comparisons. The remaining studies were controlled before and after studies (Adukia, 2014; Lokshin & Yemtsov, 2004). Both of these studies used difference-in-differences analysis to adjust for confounding between groups.

Qualitative studies, process evaluations and project documents

We identified eleven additional documents on the seven included new schools and infrastructure programmes. This included six project documents and five process evaluations. We did not identify any qualitative or mixed methods studies. Two of these project documents cover the education infrastructure rehabilitation projects in Georgia, two covered the PACE-A programme in Afghanistan (one process evaluation and one project document), five covered the SIF education investments in Bolivia (three process evaluations and two project documents), one the Akshara library programme in India (one project document), and one the IMAGINE programme in Niger (process evaluation).

Table 6.6 a: Characteristics of included new schools and infrastructure programmes

Included study	Setting	Description of the intervention	Included outcomes	Length of follow up	Study design	Sample Size
Freeman *et al*. (2012) **Hygiene infrastructure**	Kenya (Unclear as to whether rural, urban or peri-urban) Primary schools Age:13.3 (mean age for the Hygiene Promotion and Water Treatment Intervention); 13.2 (mean age for Hygiene Promotion, Water Treatment and Sanitation Intervention) Grade: 5.5 (mean grade for the Hygiene Promotion and Water Treatment Intervention); 5.9 (mean grade for Hygiene	No programme name: There were two treatments tested under this hygiene promotion experiment: 1. A hygiene promotion and water treatment arm that received a 3-day training of teachers on hygiene promotion, behaviour change and water treatment methods and regular follow-up visits throughout the school year. The programme provided handwashing and drinking water containers and a one-time, one-year supply of WaterGuard, a 1.2 per cent chlorine-based point-of-use water disinfectant. 2. Schools in the second intervention arm received components listed above, and in addition, the provision of latrines. In addition, students in both intervention and control schools were dewormed after the baseline and again in 2008 with a single 400mg dose of albendazole.	Attendance	17 months	Cluster RCT	6036 students (all treatment groups)
Adukia (2014) **Hygiene infrastructure**	India (rural, urban and peri-urban areas); Primary schools, Upper primary schools	School Sanitation and Hygiene Education Programme (SSHE): The Total Sanitation Campaign (TSC) was launched in 1999 by the Indian government and the School Sanitation and Hygiene Education	Enrolment Attendance (attended an exam) Dropout	Approx. up to 36 months (dependi ng on	CBA (quasi-experiment with baseline and endline	121,206 primary schools 17,796 upper-

	Age: not reported Grade: 1-5 and 6-8	programme (SSHE) was one component of this program. The primary goal of the SSHE programme was to increase school latrine coverage in rural areas to 100 per cent to: 1. create a healthier environment through the elimination of open defecation, reducing disease and worm infestation 2. to reduce security risks for girls attending school		when latrines were built)	data collection)	primary schools
Burde *et al.* (2013) **Construction of new schools**	Afghanistan (rural) Primary schools Age: 8.321 (average age in treatment group) Grade: village-based schools that were not age-graded	Partnership for Advancing Community-Based Education in Afghanistan (PACE-A): The programme involved the establishment of community-based schools in rural villages, specifically to increase girls primary participation in education by reducing the distance they needed to travel to school. Programme partners provided teacher and community training, which was streamlined with Ministry of Education teacher training, as well as administrative support and materials. In some cases supplies are provided to make the schools accessible and useable in winter.	Attendance Enrolment Learning	Two follow ups: approx. 4 months and 9 months	Cluster RCT	1477 students (31 villages)
Borkum *et al.* (2013) **Improvement/ replacement of school infrastructure**	India (Urban and possibly peri-urban areas) Primary schools Age: approx. nine (mean age of sample) Grade: 3,4 and 5	The Akshara library program: The programme replaced existing libraries with libraries which were well-equipped with books designed to support the existing school curriculum and staffed with a dedicated librarian. The librarian provided regular reading-focused educational activities to encourage use of the library and facilitate	Attendance Learning	Approx. 18 months	Cluster RCT	20858 students

		interaction with the available books. Akshara libraries were organised according to a hub and spoke system with each hub school attached to several spoke schools in the same geographic area. Hub schools contain physical libraries while spoke schools do not have a physical library but instead are visited regularly by a mobile librarian.				
Newman *et al.* (2002) **Improvement/ replacement of school infrastructure**	Bolivia (rural) Primary schools, Secondary schools Age: not reported Grade: not reported	Social Investment Fund (SIF) Education investments: SIF-financed education projects either repaired existing schools or constructed new ones and provided new desks, blackboards, and playgrounds. New schools were often constructed in the same location as old schools, which were then used for storage or converted into housing for teachers. The wider SIF project consisted of a number of infrastructure and complementary measures in the areas of primary education, basic health care and sanitation.	Enrolment Attendance Completion Dropout	Approx. 60 months	Two samples: - Chaco: Cluster RCT - Resto Rural: CBA (quasi-experiment with baseline and endline data collection)	Chaco: 84 schools Resto Rural: 64 schools

Lokshin & Yemtsov (2004) **Improvement/ replacement of school infrastructure**	Georgia (rural) School level: not reported Age: not reported Grade: not reported	No programme name: The investments typically involved the rehabilitation and repair of existing schools rather than construction of new schools. These focused on improving school buildings: repairing roofs, windows, and floors, replacing inner pipes, installing sanitary and heating equipment, and repainting walls, including weatherisation of infrastructure to make it useable during the winter period.	Enrolment Completion	Approx 12 months for enrolment outcome. Approx 48	CBA (quasi-experiment with baseline and endline data collection)	249 villages
Dumitrescu *et al.* (2011)	Niger (rural) Primary schools Age: nine years Grade: not reported	The programme consisted of the construction of 68 primary schools and implementation of a set of complementary interventions. The schools were based on a prototype that included three classrooms, housing for three female teachers, a preschool, and separate latrines for boys and girls equipped with hand washing stations. In addition, schools were located near a water source. The complementary interventions involved design and dissemination of training modules for teachers, supply of learning materials, the promotion of extracurricular activities, teacher incentives, and a girls' education campaign. This latter component included advocacy days, local action plans, capacity building through school committees, adult literacy and income generating projects.	Learning, Enrolment, Attendance	23 months	Cluster RCT	Up to 13969 students

6.6.3 Synthesis of findings

We have structured this section by the three sub-groups of interventions: hygiene infrastructure interventions, construction of new schools and interventions that involve improvement or construction of new school infrastructure.

Effects of new schools and infrastructure interventions on enrolment, attendance, dropout rates, completion and learning outcomes

This section reports the results of the effects of new schools and infrastructure interventions (addressing question 1a and 1b). As previously described, we identified seven studies that evaluated any type of school infrastructure intervention. Only five of these studies provided enough data to calculate effect sizes; we did not have sufficient data to calculate effect sizes for Newman *et al.* (2002).[77] Due to the limited numbers and diversity of intervention designs, we were only able to conduct meta-analysis for two outcomes (enrolment and attendance) for hygiene infrastructure programmes and new schools. For other outcomes and for all results for infrastructure programmes, we present the individual effect sizes only.

All effect sizes are expressed as standardised mean differences (SMD), interpreted as the magnitude of the number of standard deviation changes in the outcome for the intervention group as compared to students in comparison schools.

Hygiene infrastructure interventions

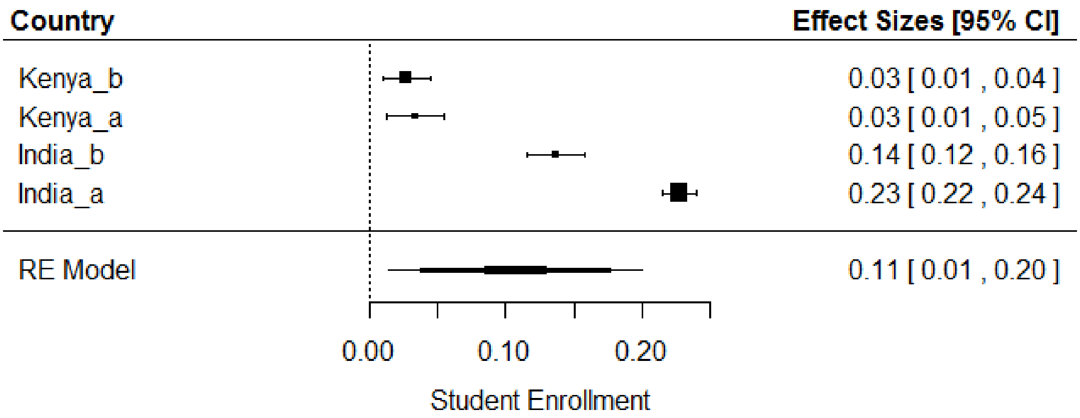
Enrolment

Figure 6.6a presents the forest plot with the results of the individual latrine construction studies and the pooled point estimate on student enrolment in school. The four effect sizes come from two different latrine-building programmes but refer to four different independent samples within these programmes[78]. The overall average effect of interventions to build latrines on enrolment is 0.11 SMD, with a 95% confidence interval (CI) [0.01, 0.20].

The assessment of heterogeneity suggests almost total between study variability (I^2 = 99%, τ^2 = 0.009, Q(3 df) = 452.5, p-value < 0.000). This can also be seen visually in Figure 6.6a; the effect sizes range between 0.03 SMD (95% CI [0.01, 0.04]) in Kenya (Freeman *et al.*, 2012: Kenya_b) and 0.23 SMD (95% CI [0.22, 0.24]) in India (Adukia *et al.*, 2013: India_a). Apart from the two samples from Kenya, the confidence intervals are not overlapping. As expected, the sensitivity analysis indicates that the results are sensitive to the removal of the two observations from Kenya and India_a (see Appendix H for results of all sensitivity analyses). The removal of Kenya_a and Kenya_b increases the point estimate to around 0.13 SMD in both cases with p-values at 0.024 and 0.18 respectively. . Conversely, the removal of India_a reduces the point estimate to 0.07 SMD (95% CI, [0, 0.14]).

[77] We contacted the authors requesting additional data however did not get a response.
[78] Kenya_a refers to the effect on water scarce schools. Kenya_b refers to the effect on schools where water was already available. India_a refers to the effect on primary schools. India_b refers to the effect on upper primary schools.

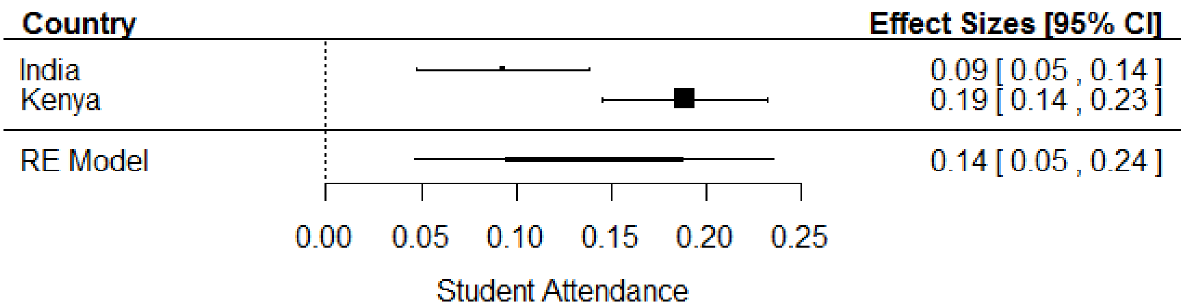
Figure 6.6 a: Student Enrolment



Attendance

Figure 6.6b presents the forest plot with the results of the individual latrine construction studies and the pooled point estimate on student attendance. The overall average effect of latrine construction on attendance is 0.14 SMD, which is statistically significant at the 1 per cent level (95% CI [0.05, 0.24]). The assessment of heterogeneity suggests a large amount of between study heterogeneity with an I^2 of 89 per cent (also τ^2 = 0.004, Q[1 df] = 8.9, p-value = 0.003).

Figure 6.6 b: Student Attendance

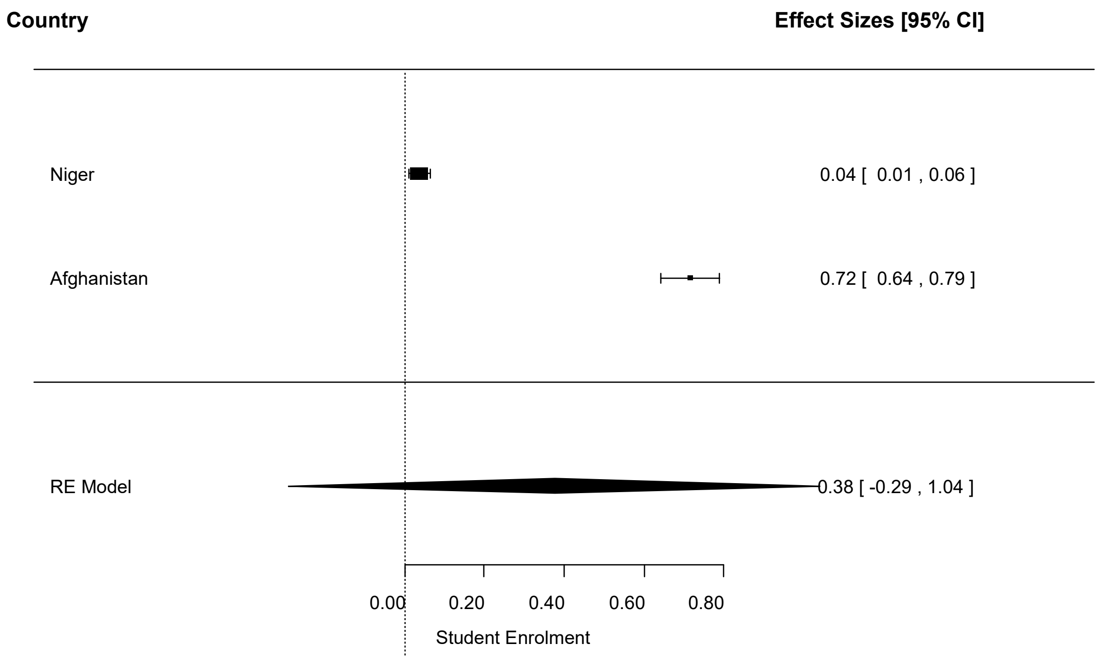


Construction of new schools

Enrolment

The overall average effect of interventions to build new schools on enrolment is positive but not significant (0.38 SMD, with a 95% CI [-0.29, 1.04]). The assessment of heterogeneity suggests large between study variability (I^2 = 99%, τ^2 = 0.2301, Q(3 df) = 284.2, p-value < 0.001). This can also be seen visually in Figure 6.6c – the confidence intervals are not overlapping, indicating a substantial difference between the effects of school construction in Niger and school construction in Afghanistan. The large effects observed in Afghanistan is larger for girls than for boys where we see an effect size of 0.48 SMD (95% CI [0.39, 0.57]) for girls' enrolment and only 0.26 SMD for boys (95% CI [0.17, 0.35]) (Dumitrescu *et al.,* 2011).

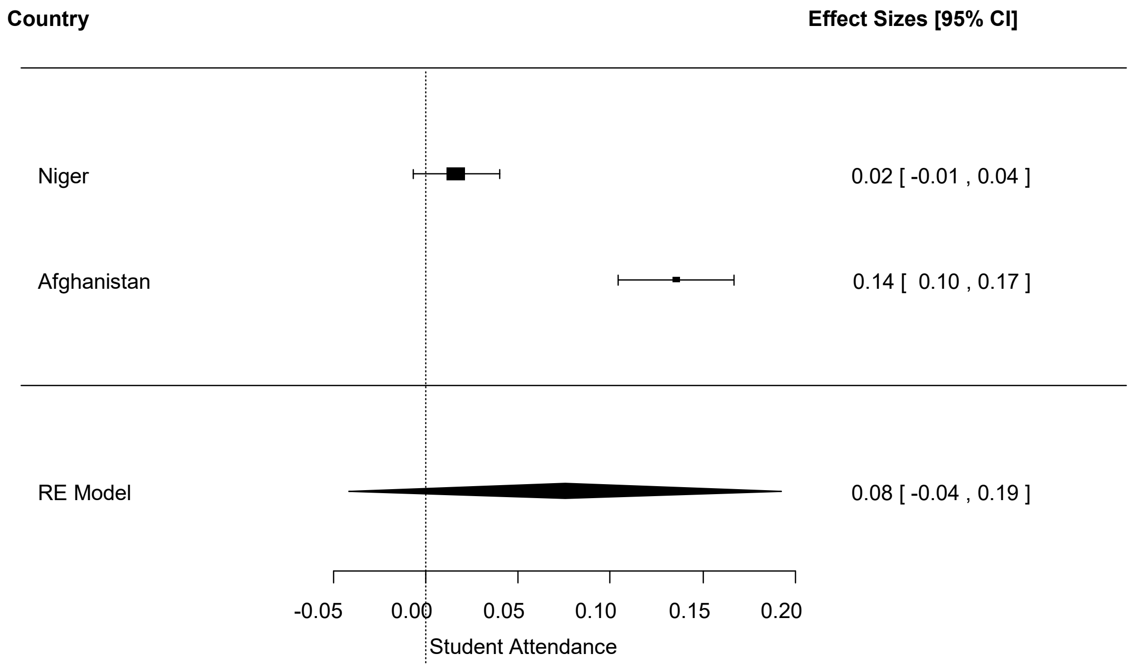
Figure 6.6 c: Student Enrolment



Attendance

The overall average effect of these interventions on attendance is 0.08 SMD, 95% CI [-0.04, 0.19]. The homogeneity tests suggest a large amount of heterogeneity (I^2 = 97%, τ^2 = 0.0069, Q(3 df) = 35.18, p-value = < 0.001). As Figure 6.6d shows, the confidence intervals of the two studies are not overlapping.

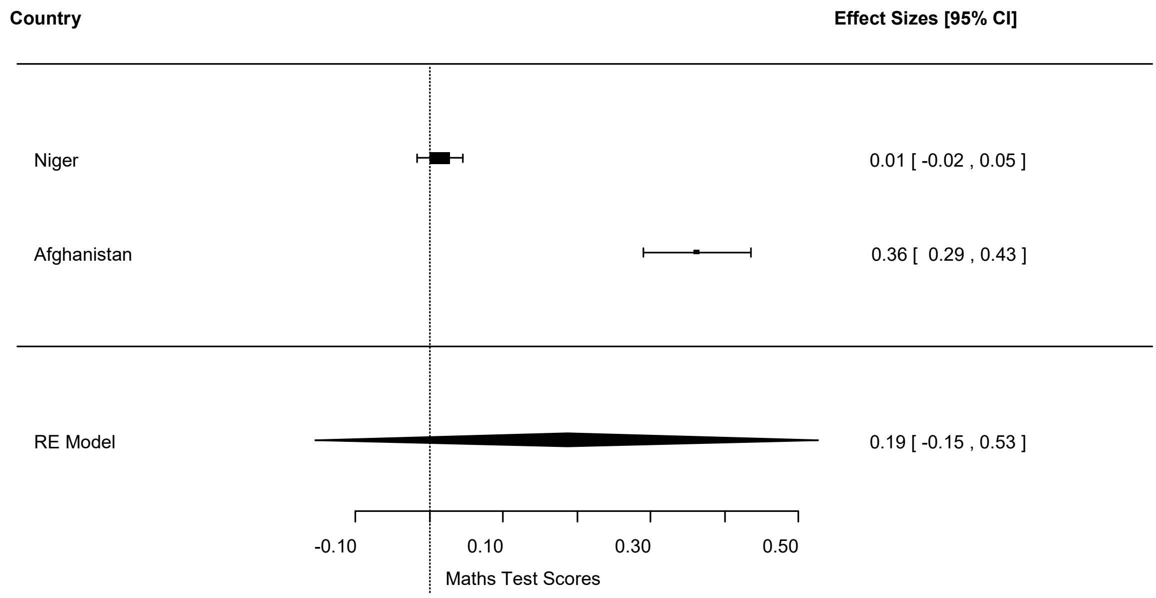
Figure 6.6 d: Student attendance



Maths test scores

As can be seen from Figure 6.6e the overall average effect on students' maths test scores is 0.19 SMD, 95% CI, [-0.15, 0.53]. The assessment of heterogeneity suggests a very high level of between study variability (I^2 = 99%, τ^2 = 0.0597, Q(3 df) = 74.08, p-value < 0.001). Similar to the forest plots above, the confidence intervals are not overlapping

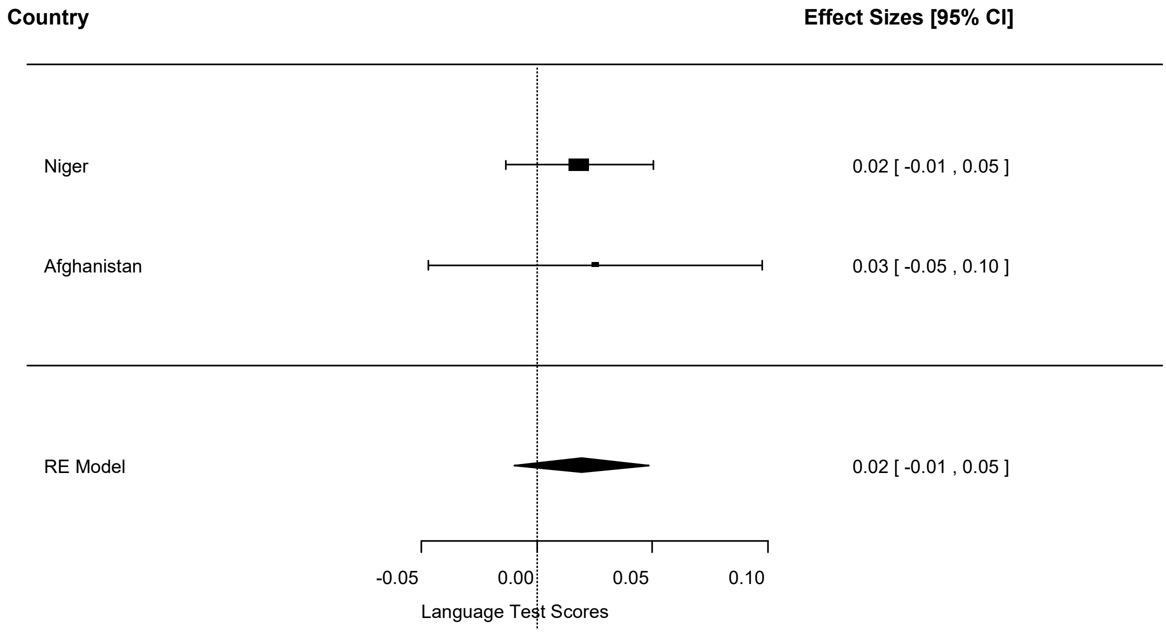
Figure 6.6 e: Maths test scores



Language arts test scores

Figure 6.6f presents the forest plot with the results of the school construction studies and the pooled point estimate on students' language arts test scores. The overall average effect of these interventions is 0.02 SMD, 95% CI [-0.01, 0.05]. The homogeneity tests suggest no heterogeneity (I^2 = 0%, τ^2 = 0, Q(3 df) = 0.03, p-value = 0.8601) and the confidence intervals are overlapping.

Figure 6.6 f: Language arts test scores



Improvement or construction of new school infrastructure

In table 6.6c we report effect sizes for two programmes that improved or replaced school infrastructure. These are the Infrastructure Rehabilitation programmes in Georgia (Lokshin and Yemtsov, 2004) and the Akshara library programme in India (Borkum *et al.* 2013). The Infrastructure Rehabilitation programmes in Georgia measured effects on enrolment (SMD=.10 SMD, 95% CI [-0.19, 0.39]) and completion rates (0.06 SMD, 95% CI [-0.23, 0.35]).The Akshara library programme in India found zero and close to zero effects on both learning outcomes and student attendance (effect size data provided in appendix H).

6.6.4 Summary of findings and discussion

New schools and infrastructure (NSI) interventions are programs involving the construction of schools in areas where there were none previously, or improvement or rehabilitation of existing school infrastructure. The interventions evaluated in the studies included for this category are very diverse, and so we grouped them into three sub-categories for the purposes of analysis. Two studies included the provision of latrines and were classified as hygiene infrastructure interventions. Two studies focused on the establishment of new, community-based schools in underserved areas, and finally three reported on the improvement or replacement of school infrastructure, including the provision of a library and rehabilitation of existing infrastructure.

Overall, the small pool of studies on new schools and infrastructure suggested a beneficial effect on school participation, as measured by enrolment and attendance, but these results should be interpreted with caution as they come from a small number of studies. Construction of new schools may improve enrolment and attendance in some contexts, with large improvements observed in a context of low school participation in Afghanistan, particularly for girls (enrolment: SMD=0.38, 95% CI [-0.29, 1.04]; attendance: SMD=0.08, 95% CI [-0.04, 0.19]), with effects of smaller magnitude observed in Niger. There may also be beneficial effects of construction of new schools on maths and language arts (maths: SMD=0.19, 95% CI [-0.15, 0.53]; language arts: SMD=0.02, 95% CI [-0.01, 0.05]).

The results of the two hygiene infrastructure studies in India and Kenya suggested that these interventions can have a positive effect on school participation (enrolment, 0.11 SMD, 95% CI [0.01, 0.20] and student attendance 0.14 SMD, 95% CI [0.05, 0.24]). The effect of improving existing infrastructure is not clear from the two included studies for which we were able to use data.

6.7 Grouping students by ability

We identified three studies assessing different interventions using student ability to allocate students to classes and groups. Broadly speaking, these interventions group students by ability so that classes can be better tailored to their needs. One study reports on a 'grade retention' intervention whereby students are given additional time to learn the material from a given grade, rather than falling further behind. Two others report on 'tracking' interventions that group students by ability in order to ensure that teaching can be targeted to students' abilities.

6.7.1 Description of included studies

Chen *et al.* (2010) evaluate the effect of a grade retention intervention whereby students that are behind their peers re-sit the grade, thereby giving them more time to learn the relevant material. Duflo *et al.* (2011) and Duflo *et al.* (2015) evaluate tracking interventions designed to teach students by ability and thereby focus classes on students' competency levels. Due to the heterogeneity of the 'grade retention' and 'tracking' interventions, we did not combine them in the synthesis, but simply presented the descriptive and analytical results in the same chapter due to the common feature of ability grouping. We have summarised the descriptive characteristics of the included studies, and then presented the results. Table 6.7a provides an overview of the characteristics of included studies described in more detail in the following section.

Populations

Chen *et al.* (2010) evaluated the introduction of a new government policy allowing schools to retain students in China. The programme was implemented in urban, public primary schools. It targeted students aged 10-16 and in grades 2 to 4, though the majority (73%) were between the ages of 11-12. Approximately 45 per cent of the sample was female. The two tracking intervention were in Kenya (Duflo *et al.,* 2011) and India (Duflo *et al.,* 2015) respectively. The Extra Teacher Programme (ETP) was implemented in primary schools in Western Province in Kenya. Participating students were aged between 7 and 8, and in grades 1 and 2. The Learning Enhancement Programme (LEP) was implemented in rural public primary schools in the Indian states of Haryana. Students were in grades 1-4.

Interventions

Chen *et al.* (2010) report on a Ministry of Education initiative in China that altered the maximum number of students that schools could retain. As a result, schools had the power to decide how many students could be retained every year, though families of retained students still had to pay to cover the additional year of education. Duflo *et al.* (2011) examines an intervention that provided an additional teacher to schools in Kenya, either civil-service teachers hired by the government or contract teachers hired by local parent teacher associations. Schools split their first grade classes in two and students were 'tracked' into each class based on initial achievement. Additional classroom facilities were also provided across schools in order to facilitate the intervention (Duflo *et al.,* 2011). Duflo *et al.* (2015) evaluated the Learning Enhancement Programme (LEP) in India, under which students were briefly assessed on their Hindi skills at the start of the academic year and then a portion of the school day was set aside to group and teach students by ability, regardless of their age or grade.

Comparisons

All three of the included studies compared the effect of the intervention to business as usual (no intervention). However, Duflo *et al.* (2015) also evaluate whether a combination of the LEP with another intervention called the Continuous and Comprehensive Evaluation (CEE) programme was any more or less effective than the LEP alone. In schools receiving the two programmes together, the tracking component of the LEP was combined with ongoing assessment and teacher feedback components of the CCE. The CCE is included elsewhere in this report in its own right as part of the chapter on 'diagnostic feedback'. In the synthesis section later in this chapter we have briefly examined the difference between the effectiveness of the LEP alone and the LEP/CEE combination.

Outcomes

Chen *et al.* (2010) and Duflo *et al.,* (2015) report separate maths and language arts test scores. Duflo *et al.*, (2011) report on maths and language arts outcomes, providing separate analysis for upper primary students (equivalent to secondary school students) and primary students. They also undertake subgroup analysis by gender for primary school students only. Chen *et al.* (2010) followed up approximately 12 months after the start of the programme, while Duflo *et al.* (2011) followed up 18 months and 30 months after programme inception and Duflo *et al.* (2015) following up around 15 months after.

Study Design

Both tracking studies (Duflo *et al.,* 2011 and 2015) were based on cluster randomised controlled trials. Chen *et al.,* (2010) employed a controlled before and after design that employed Propensity Score Matching (PSM) and Difference in Difference (DID) techniques to control for selection bias and confounding.

Table 6.7 a: Characteristics of school participation by ability programmes

Included study	Population	Intervention summary	Included outcomes	Follow up	Study design	Sample size
Chen *et al.*, 2010	Public primary schools in Shaanxi province, north-western China. Students were aged between 10 and 16, and in grades 2 to 4.	**Grade retention:** In 2006 the Chinese Ministry of Education abolished the restriction on the maximum number of children a school could retain. As a result, school administrators were given more autonomy to decide how many students would be retained each year. Retaining students was intended to allow them to relearn material and catch up with their peers.	Separate maths & language arts test scores	Approx. 12 months	Controlled before and after design with Propensity Score Matching (PSM) and Difference-in-difference (DID)	36 schools and 1,649 students
Duflo *et al.*, 2011	Primary schools in Western Province in Kenya. Students were aged between seven and eight, and in grades 1 and 2.	**Tracking:** The Extra Teacher Programme (ETP) allowed schools to hire an additional teacher. Schools were thus able to assign grade one students into two groups, either on a random basis or according to prior test scores. 'Tracking' students in this way allowed classes to be tailored to students' needs, but might also disadvantage low-achieving students while benefiting high-achieving students, thereby exacerbating inequality. The purposive and random assigning strategies allowed these different possibilities to be assessed.	Composite maths and language arts test score, separate maths & language arts test scores. Teacher attendance, teacher performance and student attendance.	18 months, 30 months	Cluster RCT	121 schools, approx. 10,000 students
Duflo *et al.*, 2015	India, rural public primary schools in Haryana. Students were in grades 1-4.	**Tracking:** Under the Learning Enhancement Programme (LEP), students were briefly assessed on their Hindi skills at the start of the academic year and then a portion of the school day was set aside to group and teach students by ability, regardless of their age or grade.	Separate maths & language arts test scores	Approx. 15 months	Cluster RCT	300 Schools, 9,392 students

6.7.2 Effects of interventions grouping students by ability

We identified one study that evaluated the effect of grade retention and two studies evaluating the effect of tracking interventions on our primary outcomes of interest. Though these studies do have some characteristics in common, they are still fundamentally different and therefore we have analysed the evidence for each of these types of intervention separately. We calculated effects as standardised mean difference (SMD), interpreted as the magnitude of the number of standard deviation changes in the outcome for the intervention group as compared to students in schools with no intervention. SMD scores are interpreted as the number of standard deviation changes in the outcome. We have only been able to synthesise outcomes data for the two tracking studies for maths and language arts outcomes. All other outcomes for all studies are reported narratively.

6.7.3 Grade retention findings

The single grade retention study (Chen *et al.,* 2010) provided outcomes data for maths and language arts test scores. The effect size for both reported outcomes indicate a negative and statistically significant effect for grade retention on children's learning. The effect on Chinese test scores was -0.095, 95% CI [-0.17, -0.02], while for maths test scores it was also negative, but relatively larger in magnitude (SMD = -0.19, 95% CI [-0.26, -0.11]).

6.7.4 Tracking intervention findings

Effects of tracking on teacher attendance, teacher performance and learning outcomes

There were two studies that evaluated the effect of tracking interventions. The first is an evaluation of the Extra Teacher Programme (ETP) in Kenya, which reports effects on teacher attendance, performance and student learning outcomes (Duflo *et al.,* 2011). A second study evaluated the effect of Learning Enhancement Programme (LEP) in India on maths and language arts (Duflo *et al.,* 2015). For language arts, the authors provided outcome data for both Hindi writing and reading tests, which we combined in a synthetic effect size. For maths, the authors provide both oral and written maths test scores, which we again combined into a single synthetic effect. The authors also provided all learning outcomes divided into results for each gender sub-group. We meta-analysed outcomes from the two studies for language arts and maths test scores.

The Duflo *et al.* (2011) Kenya Extra Teacher Programme (ETP) study reported language arts and maths outcomes eighteen months and thirty months after the programme was implemented, while the Duflo *et al.* (2015) Indian Learning Enhancement Programme (LEP) study reported outcomes fifteen months after the programme was implemented. In the meta-analyses, we combined outcomes data from the fifteen/eighteen month follow-ups. Outcome measures not included in the meta-analyses are reported narratively if they provide substantively different findings (all estimates not included in the meta-analysis have been reported in the technical Appendix H, section 7.1)

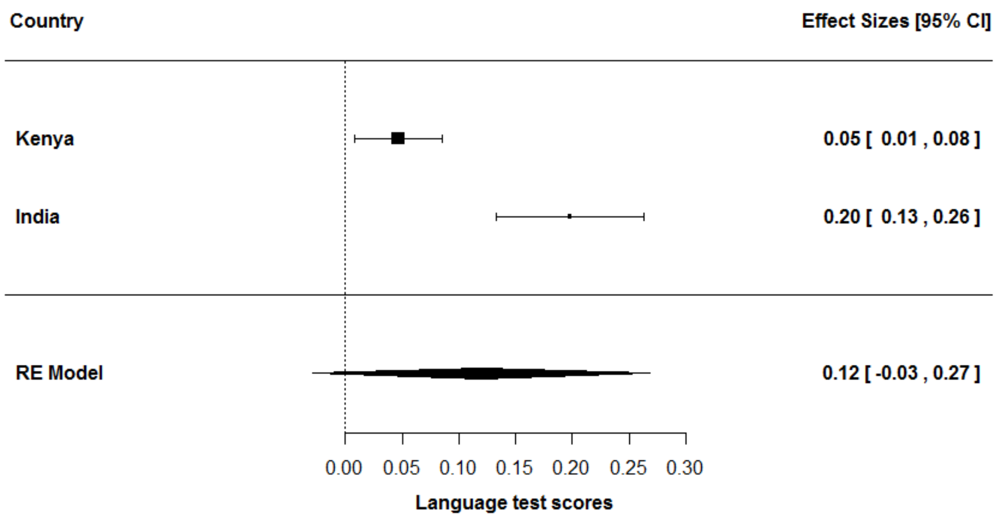
Teacher attendance and teacher performance

The study of the Extra Teacher Programme (ETP) in Kenya find an improvement in teacher attendance of 0.11, 95% CI [0.05, 0.17] and a very small negative effect on teacher performance, as measured by whether the teacher was found in the class teaching (SMD = -0.02, 95% CI [-0.07, 0.01]) (Duflo *et al.,* 2011).

Language arts test scores

The overall average effect of tracking on language arts test scores is 0.12, 95% CI [-0.03, 0.27], calculated under a random-effects model. The assessment of homogeneity indicates the effects do not arise from a common population and suggests that the results should therefore be interpreted with caution (I^2 = 93.55%, τ^2 = 0.0107, Q (df = 1) = 15.5095, p = < 0.001). Figure 6.7a presents the forest plot with the results of the analysis. The effect reported by (Duflo *et al.,* 2011) for the Kenyan ETP is smaller in magnitude than that observed for the Indian LEP and the confidence intervals of the two studies do not overlap (SMD = 0.06, 95% CI [0.01, 0.08]; SMD = 0.20, 95% CI [0.13, 0.26]).

Figure 6.7 a: Language arts test scores[79]



Duflo *et al.* (2011) find that the Kenyan ETP produced a marginally larger effect after thirty months, when compared to the effect after eighteen months that is included in the meta-analysis (SMD = 0.07, 95% CI [0.03, 0.11]). Duflo *et al.* (2015) find that the LEP in India produced a slightly larger effect on girls' language arts test scores (SMD = 0.27, 95% CI [0.22, 0.31]) than it did for boys' test scores (SMD = 0.22, 95% CI [0.18, 0.27]).

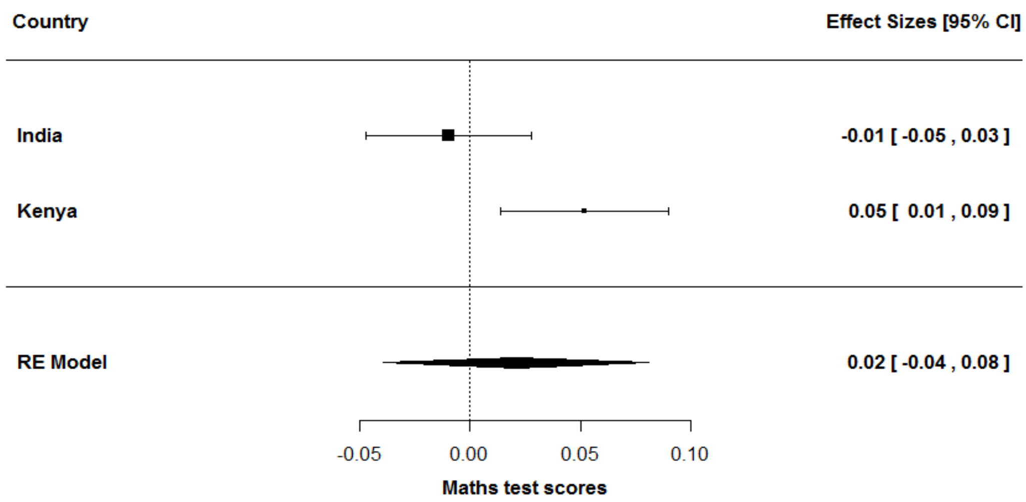
Duflo *et al.* (2015) also evaluate whether a combination of the LEP with another intervention called the Continuous and Comprehensive Evaluation (CEE) was any more or less effective than the LEP alone. In schools receiving the two programmes together, the tracking component of the LEP was combined with ongoing assessment and teacher feedback components of the CCE. The authors find effects are small for the combined CCE and LEP programme on both Hindi reading test scores (SMD = 0.01, 95% CI [-0.03, 0.04]) and written Hindi test scores (SMD = 0.03, 95% CI [0.00, 0.07]).

[79] Kenya: Duflo *et al.,* 2011 (ETP)
India: Duflo *et al.,* 2015 (LEP)

Maths test scores

The overall average effect of tracking interventions on maths test scores is 0.02, 95% CI [-0.04, 0.08], as calculated under a random-effects model. The assessment of homogeneity indicates the effects do not arise from a common population and therefore the results should be interpreted with caution (I^2 = 80.27%, τ^2 = 0.0015, Q (df = 1) = 0.0244, p = < 0.001). Figure 6.7b supports the heterogeneity of effects, although the confidence intervals are overlapping.

Figure 6.7 b: Maths test scores[80]



For the Kenyan ETP, Duflo *et al.* (2015) again find that after a longer follow-up of thirty months, the effect of the programme on maths test scores was marginally larger than the effect after eighteen months (SMD = 0.06, 95% CI [0.02, 0.10]). Duflo *et al.* (2015) find that the effect of the LEP in India were not substantively different for girls than they were for boys.

As before, Duflo *et al.* (2015) also evaluate whether a combination of the LEP with the CEE was any more or less effective than the LEP alone and find effects are not substantively different for the combined CCE and LEP (SMD = 0.02, 95% CI [-0.01, 0.06]; SMD = 0.03, 95% CI [0.00, 0.07]).

6.7.5 Summary of findings and discussion

This section has presented the findings of studies assessing programmes grouping students by ability. We only identified three studies assessing interventions using student ability to organise students into groups based on ability. Chen et al (2010) find that students in schools with a grade retention policy performed worse on tests in Chinese and maths. The evidence regarding tracking interventions is also limited, with only two studies included.

[80] Kenya: Duflo *et al.,* 2011 (ETP); India: Duflo *et al.,* 2015 (LEP)

7. Teacher-level interventions

Teacher interventions are designed to improve teachers' qualifications, skills, knowledge and commitment, and/or to ensure that there are more teachers per pupil and better quality contact-time and teaching. We reviewed the evidence on four types of interventions, teacher incentives, teacher hiring interventions, diagnostic feedback and teacher training.

This chapter provides the findings of our synthesis of studies evaluating the effect of teacher hiring and teacher incentive on learning outcomes and access to schooling. The chapter is organised by intervention. Each sub-section starts with a description of the intervention type and its theory of change, followed by descriptive results and the findings addressing our research questions.

7.1 Teacher Incentives

Interventions providing teacher incentives and promoting accountability seek to improve the working conditions in schools so that teachers are motivated to come to work and improve their performance. Such interventions take many forms, such as providing direct payments to teachers based on their attendance or based on their students' achievement (Glewwe *et al.,* 2008; Cueto *et al.* 2008). For instance, a programme in India offered teachers a cash bonus linked to their pupils' performance in independent tests (Muralidharan & Sundararaman, 2009). The Teacher Incentive programme in Kenya offered primary-school teachers in-kind rewards based on pupils' exam scores rather than cash (Glewwe *et al.*, 2010). Other interventions may use monitoring in order to keep track of teachers performance. Such monitoring may be undertaken by school principals, external assessors, or community members (Guerreo *et al.,* 2012).

7.1.1 How may Teacher Incentive programmes affect education outcomes?

Teachers are one of the most important factors determining students' performance (Miguel & Barsaga, 1997; Mourshed *et al.,* 2010). This applies even more so in resource-poor contexts where the teacher has to compensate for lack of materials, parental support and may often be the only educated person in the child's life (Carron & Chau, 1996, Ngware, Oketch & Ezeh, 2011). However, evidence shows that education systems in L&MICs are significantly affected by teacher absenteeism, low teacher qualifications and low norms for teacher effort (Bennell, 2004; Chaudhury *et al.* 2006, Duflo and Hannah, 2005, (Muralidharan and Sundararaman, 2008). Programmes introducing performance-linked incentives for teachers aim to address these issues.

Figure 7.1a briefly outlines the programme theory and key assumptions for teacher incentive programmes. The basic causal chain of teacher incentive programmes follows the logic that rewarding teacher performance increases teachers' intrinsic motivation and effort, and that this then translates into improved student outcomes. In order for teacher incentives to increase teacher efforts and quality of learning, bonuses need to act as a positive incentive. They need to be administered correctly, both in terms of the size and type of bonus, timeliness of disbursement and structure of the incentive scheme (Cueto, 2008; Gallego, 2008; Contreras *et al.* 2012; Santibañez *et al.* 2007). However, as suggested by the Principal-Agent theory (PA) (Le Grand, 2006), as teachers are generally not supervised during their teaching activity by the school principal, it is difficult to know to what extent teachers improve their teaching practices and to ultimately establish a direct link between teachers' effort and student performance. This creates opportunities for teachers to 'play' the system in a way that allows them to collect the bonus with minimum effort. Several characteristics of teacher incentive programmes may affect the extent to which teachers are incentivized to promote student learning rather than gaming of the system.

Incentives awarded to individual teachers may act as strong motivation for individual teacher performance. However, they may also create competition among teachers that can discourage teacher team work and peer-learning (Glewwe *et al.* 2010, (Umansky, 2005, Levacic, 2009). Group or team incentives, providing incentives to the team regardless of the performance of each teacher, can be used to foster cooperation among teachers rather than harmful competition (Lavy 2003). This cooperation may include more consultation and reflection among teachers on teaching methods and is likely to have a positive impact on their professionalism. However, group-based incentives awarded based on average school performance may promote free-riding by some teachers (Muralidharan and Sunararaman 2011). Depending on the size of the school, peer pressure and social norms may minimise this risk (Kandel and Lazear 1992; Kandori 1992)

Because teacher performance is difficult to observe directly, many teacher incentive programmes rely on observable proxy measures such as teacher attendance and student test scores. Teachers whose bonus is determined by students' performance on a particular test may put most of their efforts into teaching the students how to pass that specific test (e.g, training the students on how to deal with questions they are likely to find in the test) (Muralidharan and Sundararaman, 2011), rather than focusing on improving students' overall understanding of the subject matter. By doing this, they may achieve the short-term goal of teaching, i.e. increase test scores, but not necessarily the long-term goals such as developing children's critical thinking skills (Holmstrom and Milgrom, 1991). Moreover, an incentive award based on number of students that reach a defined threshold (e.g.: number of students that pass the grade) may lead teachers to focus only on the marginal students, neglecting those that are unlikely to improve sufficiently to cross the threshold, or those that are guaranteed to pass (Behrman *et al.* 2015; Muralidharan and Sundararaman, 2011; De Fraja, G., & Landeras, 2006). Certain bonus structures, such as awarding the bonus based on the average improvement of all students rather than a threshold, may act to discourage teachers from focusing only on the brightest students (Muralidharan and Sundararaman, 2011).

It is clear that the design of the intervention can play an important role in the effectiveness of the programme. In section 7.1.2 we briefly outline the key characteristics of the included studies and the programmes they evaluate.

Figure 7.1 a: Programme theory: Teacher incentives

Inputs	Activities	Outputs	Intermediate Outcomes	Final Outcomes
(1) Budget for teacher bonuses (2) Budget and time to provide training to teachers and school principals to explain how the incentive works (3) Extra time for senior staff to monitor teachers' input (4) A monitoring system collecting information on teacher performance	(5) Agree on an incentive scheme, indicators and timeline to receive the bonus (6) Induction programmes (7) Assign responsibility to monitors (8) Monitoring of teacher performance (9) Bonus award	(1) Incentive scheme is implemented (2) Induction programmes are scheduled and take place (3) Responsibilities assigned to teachers and school principals	(1) Teachers feel they have a greater reason to perform well (immediate) (2) Teachers absenteeism decreases (3) Teacher effort increases Potential adverse effects: (4) Teachers focus on marginal children only (5) Teachers 'teach to the test'	(1) Improved student attendance (2) Improved student learning and achievement
Assumptions: (1) The government or NGO/Agency has the ability to administer the programme, in a targeted area (2) Schools are receptive to the idea of teacher incentives (3) Teachers unions do not organise campaigns or strikes against this measure	*Assumptions:* (1) The financial bonus will act as a positive incentive (2) Participation of key stakeholders to the purpose of the incentive scheme (3) The new responsibilities do not overload teachers	*Assumptions:* (1) The incentive scheme addresses schools' specific needs (2) Induction programmes are well delivered and easy to follow and all school staff can attend meetings and ask questions (3) The staff assigned the monitoring role understood their task and are able to perform it (4) The monitoring system fairly and accurately captures the monitored behaviour	*Assumptions:* (1) Teachers buy into the incentive scheme and are not demotivated (2) The incentive reinforces teachers' effort (teachers cooperate and do not game the system) (3) Teachers know how to improve their teaching strategies and have enough time/resources (4) Teachers improve their effort in an inclusive way (e.g. do not focus only on marginal students) (5) Teachers have intrinsic motivation to improve student learning rather than teaching to the test	*Assumptions:* (1) Examinations are presided over fairly by school and education ministry (2) Exams correspond to what the students have been taught (3) Students do not cheat (4) Exam marks are not altered and are a fair and accurate representation of performance (5) Poor teacher attendance or effort was a key reason for students' poor attendance and/or performance

7.1.2 Description of included studies

We included ten studies reported in eighteen different papers that evaluated the effect of a teacher incentive intervention. We use the term 'study' to refer to the research produced from an author team, which in some cases was reported in several papers. There was one instance where we included two different studies that use the same dataset but had different author teams undertaking different analysis and reporting additional outcomes. Gallego (2008a) and Contreras and Rau (2012) both report on the Sistema Nacional de Evaluación del Desempeño (SNED) in Chile. The ten included studies therefore refer to nine unique programmes.[81], [82] Table 7.1a provides and overview of the characteristics of the included studies and we summarise this below.

Setting

The included ten studies cover programmes in Latin America and the Caribbean, Sub-Saharan Africa, South Asia and East Asia and the Pacific. Two of the programmes took place in Mexico (Santibañez, 2007; Behrman *et al.,* 2012), two in India (Duflo *et al.,* 2012; Muralidharan and Sundararaman, 2011), and one in each of Pakistan (Barrera-Osorio et al. 2015), China (Loyalka *et al.* 2015), Chile (Gallego, 2008a; Contreras and Rau, 2012), Peru (Cueto *et al.,* 2008) and Kenya (Glewwe *et al.,* 2010). Five of the studies took place primarily in rural areas, two of the programmes were implemented nationwide (Gallego, 2008a; Contreras and Rau, 2012; Santibañez, 2007), while a single study was reported to have been implemented in both rural and urban areas (Barrerra-Osorio *et al.,* 2015).

Population

Five of the programmes were targeted at the primary school level only, covering a range of grades from first grade up to the end of primary school (Glewwe *et al.,* 2010; Muralidharan and Sundararaman, 2011; Cueto *et al.*, 2008; Barrera-Osorio *et al.* 2015; Loyalka 2015), although only one study reported information on age and gender of the sampled students (Glewwe *et al.*, 2010). Behrman *et al.,* (2012) evaluates outcomes for secondary schools only while three further studies evaluate outcomes for both primary and secondary schools together (Contreras and Rau, 2012; Gallego *et al.* 2008; Santibañez, 2007) A single study evaluated a programme that was implemented in non-formal education centres (Duflo *et al.* 2012).

Five of the ten programmes were targeted at public (government-run) schools only (Santibañez, 2007; Muralidharan and Sundararaman, 2011; Glewwe *et al.*, 2010; Behrman *et al.*, 2012; Barrera-Osorio *et al.,* 2015). Of these, the Aligning Learning Incentives (ALI) Programme in Mexico targeted exclusively public technical schools (Behrman *et al.,* 2012). The SNED programme in Chile was targeted at both public and publicly subsidised schools (Gallego, 2008a; Contreras and Rau, 2012). One study evaluates outcomes for students in NGO-run schools in tribal villages in India which serve as an entry point into government schools (Duflo *et al.,* 2012). Two studies do not report the type of school targeted *(Cueto et al.,* 2008; Loyalka, P., 2015).

---

[81] One of these studies tested two different conditions – group incentives and individual incentives and also reports the effects for any incentive vs. no intervention. We decided to include the latter comparison in the analysis. As a result, the two conditions are treated as one incentive programme in the analysis and write up.
[82] These two evaluations were conducted over mostly overlapping time periods using the same data set. To avoid dependency between effect estimates from these two studies, we chose to include learning outcomes from Gallego (2008) as this study used a study design that was more comparable to other studies reporting on learning outcomes. We included any additional outcomes reported in Contreras and Raul (2012).

Intervention

All nine programmes include a key feature of teacher incentive programmes – they provide an incentive or reward to teachers based on their performance, with the aim of incentivising teacher effort and recognising and rewarding excellence. However, there was considerable variation in how the teacher incentive programmes were designed.

Individual vs. group incentive structure

Five of the programmes – The ALI experiment, Carrera Magisterial (both in Mexico), the Seva Mandir Teacher incentive programme in India, the Improvers Bonus Programme for Government School Teachers in Pakistan and the ‘Pay by Design’ in China provide individual incentive schemes awarding teachers based on their performance (Behrman *et al.,* 2013; Santibañez *et al.,* 2007; Duflo *et al.,* 2012; Barrera-Osorio, F. *et al.* 2014; Loyalka, 2015). Two programmes – the SNED programme in Chile and ICS Teacher incentive programme in Kenya are group incentive schemes, awarding schools based on overall school performance, with rewards distributed among all teachers as bonuses (Gallego, 2008; Contreras and Rau, 2012; Glewwe *et al.,* 2010). The Andra Pradesh Randomized Evaluation (APRest) in India compares these two types of incentive schemes and also provides a comparison of incentive treatments vs. no intervention (Muralidharan and Sundararaman, 2011) and the META programme in Peru is a mixed individual and group incentive scheme (Cueto *et al.,* 2008).

Bonus structure

Five of the nine programmes – the ALI experiment in Mexico, APREst in India, the 'Pay by Design' programme in China, the Improvers Bonus Programme for Government School Teachers in Pakistan and the ICS programme in Kenya condition the bonus on student learning improvements (Behrman *et al.,* 2013; Muralidharan and Sundararaman, 2011; Glewwe *et al.,* 2010; Barrera-Osorio 2015; Loyalka 2015). The SNED programme in Chile, Carrera Magisterial in Mexico, the 'Pay by Design' programme in China and the Improvers Bonus Programme for Government School Teachers in Pakistan award incentives based on performance on a composite score capturing both student learning and school performance measures (Gallego, 2008; Contreras and Rau, 2012; Santibañez *et al.,* 2007; Barrera-Osorio *et al.,* 2014; Loyalka, 2015) and the META programme in Peru and Seva Mandir programme in India condition the bonus on improved teacher attendance (Cueto *et al.,* 2008; Duflo *et al.,* 2012). The Carrera Magisterial programme in Mexico differs from the other programmes as it is a competition for a permanent wage increase rather than a bonus incentive scheme (Santibañez, 2007).

Type, value and frequency of incentives

Eight out of the nine programmes provided monetary incentives (Behrman *et al.,* 2013; Gallego, 2008; Contreras and Rau, 2012; Muralidharan and Sundararaman, 2011; Santibañez *et al.,* 2007; Cueto *et al.,* 2008; Duflo *et al.;* 2012, Barrera-Osorio, F., 2014; Loyalka, P., 2015). On the other hand, the ICS Teacher incentive programme in Kenya provided in-kind rewards as these were deemed to be more culturally acceptable than monetary awards (Glewwe *et al.,* 2010).

Five programmes – the ALI experiment, APREst in India, ICS programme in Kenya, the 'Pay by Design' programme in China and the Improvers Bonus Programme for Government School Teachers in Pakistan award the bonus annually (Behrman *et al.,* 2013; Muralidharan and Sundararaman, 2011; Glewwe *et al.,* 2010; Loyalka, P., 2015; Barrera-Osorio, F., 2015), the Seva Mandir programme in India awards bonuses every two months (Duflo *et al.,* 2012) and the META programme in Peru pays bonuses every few months or annually (Cueto et al., 2008). The SNED programme in Chile provides the bonus for a period of two years in the form of a per-student subsidy that is distributed among teachers as a bonus (Gallego, 2008; Contreras and Rau, 2012). The wage increase created by Carrera Magesterial in Mexico was a permanent one (Santibañez *et al.,* 2007)

In most of the programmes, the incentives ranged between 3 and 8 per cent of teachers pay (Gallego, 2008a; Muralidharan and Sundararaman, 2011; Glewwe et al., 2010; Cueto et al., 2008). In the Seva Mandir Teacher incentive programme, teachers were awarded 5 per cent of their salary for any additional day they attended school above the basic 20 days a month (Duflo *et al.,* 2012). In the 'Pay by Design' programme in China the bonus was higher, at between 8 and 16 per cent (Loyalka, 2015). A typical bonus awarded in the ALI experiment was around 10 to 15 per cent of a teacher's annual salary (Behrman et al., 2012). The bonus awarded by the Improvers Bonus Programme provided between 17 to 56 per cent of the basic salary (Barrera-Osorio, F., 2014). The permanent wage increase programme Carrera Magisterial in Mexico had by far the largest incentives, with amounts ranging between 27 and 217 per cent of base salary (Santibañez *et al.,* 2007).

Penalty system

All included programmes primarily aimed to reward positive performance, though several also adopted some elements of a penalty system. In the ALI experiment in Mexico, teachers incurred a $125 peso penalty for each student that regressed to a lower level (Behrman *et al.*, 2012). In the Seva Mandir Teacher incentive programme in India, teachers received a 50 Rupees fine for each day they skipped work within the minimum 20 required work days a month, capped at 500 Rs (Duflo *et al.,* 2012). The APREst programme in India and ICS programme in Kenya used a system were under-performance of students affected the size of the bonus that teachers could receive but did not affect teachers' basic salaries (Glewwe *et al.,* 2010; Muralidharan and Sundararaman, 2011).

Most of the included programmes appear to have used existing systems and relied on end of year exams to assess teacher performance (Behrman *et al.,* 2012; Gallego, 2008a; Contreras and Rau, 2012; Glewwe *et al.,* 2010; Barrera-Osorio *et al.,* 2014; Loyalka, P., 2015). In the Carrera Magisterial programme, peer and supervisor feedback was considered in the decision whether to award the wage increase (Santibañez, 2007). The SNED programme in Chile used a form of self-monitoring whereby schools had to provide information about their performance in the application for the award (Gallego, 2008b). The Andra Pradesh Randomised Evaluation included low-stakes monitoring and feedback from an external evaluation team throughout the school year, though the findings of this monitoring did not affect the teachers' chances of receiving the bonus (Muralidharan and Sundararaman, 2011). Both programmes incentivising teacher attendance involved extensive monitoring, either by trained monitors (Cueto *et al.,* 2008) or through photographs taken by students with cameras with tamper-proof time and date functions to prevent cheating (Duflo *et al.,* 2012). Three of the included programmes additionally had some element of public information provision about the performance of the teachers or schools in the programme (Contreras and Rau, 2012; Glewwe *et al.,* 2010; Cueto *et al.,* 2008). Table 7.1b provides an overview of these key intervention design features.

Comparisons

Seven of the included studies compared the effect of an intervention to business as usual (that is, a comparison group with no intervention) (Contreras and Rau, 2012; Cueto *et al.*, 2008; Duflo et al., 2012; Gallego, 2008a; Glewwe et al. 2010; Barrera-Osorio, et al.; 2014;

Loyalka, 2015). The study of the permanent wage increase programme Carrera Magisterial in Mexico used a regression discontinuity design, comparing teachers who had a chance to reach the threshold for promotion with two types of comparison teachers; those who could not possibly meet the threshold (because their initial points on Carrera Magisterial precluded it), and those who were guaranteed to reach the threshold (whose initial points on CM were already above the threshold anyway) (see McEwans and Santibañez, 2005, p. 234). Two studies evaluated multiple treatment arms and provided treatment arm comparisons (Muralidharan and Sundararaman, 2011; Behrman *et al.,* 2012). For all studies, we chose the control group with no intervention as the comparison.

Outcomes

The included studies reported on a wide range of education outcomes. All ten studies reported some measure of student learning. In all cases, learning was measured through test scores: mathematics (n= 8), official language (n= 4), local language (n=1), and composite scores (n= 5). One study evaluated reading comprehension (Cueto *et al.,* 2008). Two studies presented results separately for boys and girls (Behrman *et al*. 2013 for maths test scores, Duflo *et al.,* 2012 for Maths and Hindi test scores). In addition to learning, a number of studies also reported results for enrolment (n=2), attendance (n=3), drop-out (n=4), completion (n=4).

We also collected data on secondary outcomes of interest. Four studies reported findings on teacher attendance (n=3), six studies reported findings on teacher performance. The measures of teacher performance were heterogeneous; examples include the probability of actively teaching during unannounced classroom observations (Muralidharan and Sundararaman, 2011), or the proportion of curricular content covered split by difficulty level (Loyalka, 2015) to whether teachers gave extra classes beyond normal school hours (Glewwe *et al.,* 2010). In addition to these, one study reported the effects of the intervention on teacher characteristics (Gallego *et al.,* 2008) and one study reported findings on student effort such as number of hours per week of math study and fraction of students that pay attention more than 75 per cent of the time (Behrman *et al.,* 2013).

There was some variation in the follow-up period of outcome data collection after the start of an intervention. Two studies followed up after 12 months or less of intervention exposure (Duflo *et al.,* 2012; Cueto *et al.,* 2008). Five studies collected outcome data between 12 months and 48 months after the start of the intervention (Glewwe *et al.,* 2010; Contreras and Rau, 2012; Behrman *et al.,* 2012, Loyalka, 2015; Barrera-Osorio, *et al.,* 2014). The study of the individual and group incentive treatments in the Andra Pradesh Randomised Evaluation evaluated outcomes at different stages of the programme, with first follow-up at nine months and then yearly, with final follow-up at 61 months (Muralidharan and Sundararaman, 2011). The study evaluating the Carrera Magisterial reported outcome data approximately 60 months after the start of the intervention, and one of the studies evaluating the SNED reported outcome data approximately 108 months after the intervention started (Gallego *et al.,* 2008).

A few studies also reported on additional intermediate outcomes, such as teacher perceptions about the programme and performance-based pay in general. Although we did not calculate effect sizes for these outcomes, we used the information to inform our analysis for question 2.

Study Design

Five of the included studies were cluster-randomised controlled trials (Behrman *et al.,* 2012; Duflo *et al.*, 2012; Muralidharan and Sundararaman, 2011; Loyalka, 2015; Barrera-Osorio, 2015). Two studies used a regression discontinuity design (Gallego, 2008a; Santibañez, 2007), the latter also used difference-in-differences estimation. One study used a controlled before-after study design with propensity score matching estimation (Cueto *et al.,* 2008) and one study used a controlled before-after study design with matched differences-in-differences between treated schools (public and private subsidized) and control schools (private fee-paying) implemented with three different empirical approaches (Contreras and Rau, 2012). A final study used a controlled before and after design with random effects regression (Glewwe *et al.,* 2010).

Qualitative studies, process evaluations and project documents

We identified several additional documents related to the included programmes, however, only one of these reported additional information not already provided in the included impact evaluations, several of which also included qualitative components. Therefore, the main part of our qualitative synthesis is based on the information and findings provided in the included impact evaluations. Below we describe the intervention components in more detail. We present the descriptive findings on process, implementation and context in the discussion section following the presentation of findings from the meta-analysis.

Table 7.1 a: Characteristics of included studies: Teacher incentives

Included study	Setting	Intervention summary	Included outcomes	Follow up	Study design	Sample size
Barrera-Osorio, F. 2015	Pakistan, urban/rural, public primary schools. Students were in grade 5.	**Improvers Bonus Programme for Government School Teachers in Pakistan (programme name not reported):** The programme offers yearly cash bonuses to public primary school teachers for school performance. Bonuses are offered on top of teachers' standard salaries and are a linear function of a school performance score obtained from a weighted sum of three indicators: (1) the gain in schools' enrolment in grades 1-5 (2) the gain in schools' mean score on Punjab's standardised fifth-grade exam (3) the participation rate of schools' fifth-grade students in the exam. The pilot has three treament arms. These are: (1) HT only: only head teachers are eligible for level 1 bonuses; (2) All: both head teachers and teachers are eligible for level-1 bonuses; (3) HT+: head teachers are eligible for level-2 bonuses, other teachers are eligible for level-1 bonuses. The size of the level-2 bonus was set as twice the size of the level-1 bonus for a given level of school performance.	Enrolment; Completion; Attainment	36 months	Cluster Randomised Controlled Trial	600 schools
Behrman *et al.* 2012	Mexico, rural public secondary schools. Students were in grades 10-13.	**The Aligning Learning Incentives (ALI) Programme (teacher incentives arm [T2]):** The ALI experiment incorporated various performance incentives, including a teacher incentive treatment arm (T2). Under T2 (assessed here), maths teachers were rewarded for their students' performance over the school year. Teachers' rewards were based on the sum of rewards earned by students (students in treatment arm 1 received rewards based on their performance in standardised mathematics tests at the end of the school year). Teachers received the equivalent of 5 per cent of students' bonuses. The	Learning; Drop-out; Completion; Teacher Performance	Approximately 36 months	Cluster Randomised Controlled Trial	48 schools

Included study	Setting	Intervention summary	Included outcomes	Follow up	Study design	Sample size
		intervention was designed to avoid teachers concentrating efforts on high-performing students. It did so in three ways: (1) Teachers gained more from improved performance of lower-performing students; (2) teachers incurred financial penalties if students regressed; (3) for students initially at pre-basic level, potential payments to teachers (and students) were strictly non-negative, with relatively large payments for big improvements in student test scores.				
Contreras and Rau 2012	Chile, nationwide public primary and secondary schools. Students were in grades 4, 8 and 10.	**Sistema Nacional de Evaluación del Desempeño (SNED) (National System of Performance Evaluation):** SNED awards a per-student subsidy to municipal and private subsidised schools with the best SIMCE (test) scores over the course of two years. Schools with similar characteristics are grouped into homogeneous groups to ensure that schools compete with similar schools. The award is allocated at the school level and given to all teachers, largely on the basis of hours worked. Ten per cent is allocated by the school as a differential bonus to noteworthy teachers or those who have contributed in achieving performance goals more than others. *N.B. Both Contreras and Rau (212) and Gallego et al. (2008) evaluate the SNED. See footnote 2 for more detail..*	Learning	Approximately 24 months	Controlled before-after study with matched DID estimation	8,044 schools
Cueto *et al.* 2008	Peru, rural primary schools. Students were in grades 2-6.	**Mejor Educación a través de más Tiempo en el Aula programme (META) (Better education through more time in the classroom programme):** META is a teacher attendance incentive programme providing a monetary bonus for individuals present in school for 90 per cent of days over the course of the school year. An additional	Learning	7 months	Controlled before-after study with Propensity Score Matching	178 schools, 356 teachers

Included study	Setting	Intervention summary	Included outcomes	Follow up	Study design	Sample size
		amount was also payable to each teacher in a school if 80 per cent of the schools' teachers met the individual goal. Trained monitors (typically unpaid parents of students) noted teacher attendance three times each school day. The programme provided a higher incentive for teachers in rural and remote areas.			(PSM) estimation	
Duflo *et al.* 2012	India, rural NGO-run schools. Students were ages 7-10.	**Teacher incentive programme implemented by Seva Mandir (programme name not reported):** The Seva Mandir programme is a teacher attendance incentive programme. Seva Mandir gave teachers in 57 randomly selected programme schools a camera, along with instructions to have a student take a picture of the teacher and the other students at the start and close of each school day. The cameras had tamper-proof date and time functions, allowing for the collection of precise teacher attendance data to calculate teachers' salaries. In the treatment schools, teachers received a 50 Rs bonus for each additional day they attended school in excess of the minimum 20 days expected, and a Rs 50 fine for each day of the 20 days they skipped work. Due to ethical and political concerns, Seva Mandir capped the fine at Rs 500.	Attendance; Learning; Drop out Completion; Teacher Attendance; Teacher Performance	8 months; 12 months; approximately 42 months	Cluster Randomised Controlled Trial	120 schools
Gallego *et al.* 2008	Chile, nationwide public primary and secondary schools.	**Sistema Nacional de Evaluación del Desempeño (SNED) (National System of Performance Evaluation):** SNED awards a per-student subsidy to municipal and private subsidised schools with the best SIMCE (test) scores over the course of two years. Schools with similar characteristics are grouped into homogeneous groups to ensure that schools compete with similar schools. The award is allocated at the school level and given to all	Enrolment; Drop out; Teacher Performance	Approximately 96 months	Regression Discontinuity Design	36,938 students

Included study	Setting	Intervention summary	Included outcomes	Follow up	Study design	Sample size
		teachers, largely on the basis of hours worked. Ten per cent is allocated by the school as a differential bonus to noteworthy teachers or those who have contributed in achieving performance goals more than others. *N.B. Both Contreras and Rau (212) and Gallego et al. (2008) evaluate the SNED. See footnote 2 for more detail.*				
Glewwe *et al.* 2010	Kenya, rural public primary schools. Students were in grades 4-8, with an average age of 13.	**Teacher incentive programme implemented by International Child Support:** The ICS Teacher incentive programme provides incentives to teachers of 'top-scoring schools', based on absolute performance improvements in average test scores and to 'most-improved schools', based on improvement relative to baseline performance in student exams. For each category three prizes were given to first, second, third and fourth position schools. Schools could not win more than one category and could win only one type of prize.	Attendance; Learning; Drop-out Completion; Teacher Attendance; Teacher Performance	Approximately 12 months; 24 months; 36 months (12 months following intervention end date)	Controlled before and after design with random effects regression	100 schools
Loyalka, P. 2015	China, rural primary schools. Students were in grade 6 and aged 11 on average.	**'Pay by Design' (No official programme name reported):** Teachers were randomly placed into three 'rank-order tournaments' whereby teacher rankings were determined as a function of their students' scores on standardised exams. In incentive arm 1, teacher performance was defined as the class average of student scores in end of year exams. In incentive arm 2, teacher performance was defined as the class average gain in student achievement from the start to the end of the school year. In incentive arm 3, teacher performance was defined through 'pay-for-percentile' with students grouped	Attainment	21 months	Cluster Randomised Controlled Trial	7,454 students

Included study	Setting	Intervention summary	Included outcomes	Follow up	Study design	Sample size
		according to their baseline exam and assigned a percentile score based on their performance compared to peers starting the school year at a similar achievement level. Each incentive arm was further randomly split into large or small incentive treatments. Teachers were given a cash reward at the end of the school year based on their percentile ranking compared to other teachers. Teachers in all incentive arms were presented performance pay contracts and incentive structures.				
Muralidharan & Sundararaman 2011	India, rural public primary schools. Students were in grades 1-5.	**Andhra Pradesh Randomised Evaluation: teacher performace pay treatment arms (APREst):** The APREst study evaluates two teacher incentive approaches - a group (school) based teacher incentive and an individual teacher-level incentive. The group based teacher bonuses were offered to all teachers on the basis of average school level improvement in test scores. The individual teacher level bonus was allocated based on the average performance of the students taught by the teacher. The bonus was calculated as 500 Rupees per percentage point gain in average test scores compared to baseline test scores.	Attendance; Learning; Teacher Attendance; Teacher Performance	9 months; 21 months;	Cluster Randomised Controlled Trial	300 schools

Included study	Setting	Intervention summary	Included outcomes	Follow up	Study design	Sample size
Santibañez, 2007	Mexico, nationwide public primary and secondary schools.	**Carrera Magisterial (CM):** The CM is a voluntary permanent wage increase programme. Most teachers and principals in public primary and secondary schools are eligible for substantial and permanent wage increases if they perform well in a year-long assessment process. Wage increases are calculated as a function of (a) education degrees; (b) years of experience; (c) professional development, including federal and state in-service training courses; (d) peer review; (e) teacher (or principal) knowledge, which is based on a test score; and (f) student performance, which is based on students' test scores.	Learning	Approximately 60 months	Regression Discontinuity Design	27,213 schools

Table 7.1 b: Intervention design features of included studies

Author and study ID	Country	Programme name	Implemented by	Individual vs. group incentive	Monetary vs. in kind incentive	Measure of teacher assessment	Incentive award / amount determined	Penalty for poor performance	Incentive amount as % of base salary	Frequency of award
Barrera-Osorio, F. 2015	Pakistan	Improvers Bonus Programme for Government School Teachers in Pakistan	The Provincial Government of Pakistan	Group	Monetary	Student test scores, student enrolment rates and student participation in exams	As a function of student performance, gain in enrolment rate, gain in participation rate	None mentioned	Between 17 to 56 %	Annually
Behrman *et al.* 2013	Mexico	ALI (T2)	Ministry of Education	Individual	Monetary	Student test scores (maths)	As a function of student performance	Salary reduction	Approx. 10-15%	Annually
Cueto *et al.* 2008	Peru	META	Ministry of education	Combined individual and group	Monetary	Teacher attendance	Based on threshold	No penalty	Approx. 8%	Every few months or annually
Duflo *et al.* 2012	India	Seva Mandir Teacher Incentive programme	NGO	Individual	Monetary	Teacher attendance	As a function of teacher attendance	Salary reduction	5% per additional day	Every two months
Gallego *et al.* 2008 **Contreras & Rau 2012**	Chile	SNED	Ministry of Education	Group	Monetary	Composite score (student and teacher performance)	Through competition at school level, based on hours	None mentioned	Approx. 4-8%	Once for two years

Author and study ID	Country	Programme name	Implemented by	Individual vs. group incentive	Monetary vs. in kind incentive	Measure of teacher assessment	Incentive award / amount determined	Penalty for poor performance	Incentive amount as % of base salary	Frequency of award
							worked at teacher level			
Glewwe *et al.* 2010	Kenya	ICS Teacher incentive programme	NGO	Group	In kind	Student test scores (multi-subject)	Through competition	Reduction in chance of receiving bonus	Between 1.75% and 3.6%	Annually
Loyalka, P. 2015	China	'Pay by Design' (No official programme name reported)	The research team	Individual	Monetary	Student test scores	As a function of student performance	None mentioned	Approximately 8.3 % for the small incentives group and 16.7% for the large incentives group	Annually
Muralidharan & Sundararaman 2011	India	Andhra Pradesh Randomised Evaluation	NGO	Individual Group	Monetary	Student test scores (maths and language arts)	As a function of student performance	Reduction in bonus amount	Approx. 3%	Annually
Santibañez, 2007	Mexico	Carrera Magisterial	Ministry of Education and teachers unions	Individual	Monetary	Composite score (student and teacher performance)	Based on a threshold and decision committee	None mentioned	Between 27% to 215%	Permanent wage increase

7.1.3 Synthesis of findings

The results of our synthesis are presented in two sections. First, we present the findings of the meta-analysis on the effects of *Teacher Incentive* programmes on primary and secondary outcomes and explore results according to different population sub-groups. Second, we provide a discussion of the overall findings, with reference to the qualitative synthesis of intervention and implementation features associated with relative success and failure in improving educational outcomes.

Effects of teacher incentive interventions on teacher and student outcomes

This section reports the results of the meta-analysis of the effects of teacher incentive interventions, addressing question 1a. We structure the presentation of results according to the 'ideal type' theory of change (Figure 7.1a), starting with intermediate/ secondary outcomes (teacher attendance and teacher performance), followed by education access outcomes (enrolment, attendance, drop out) and final outcomes (completion, learning outcomes: composite test scores, language arts test scores, maths test scores).

A total of 9 studies provided data for meta-analysis, but none of the studies reported on all outcomes. The number of comparisons with effect sizes range from one for teacher performance (teacher effort) to eleven for maths outcomes. All effect sizes are expressed as standardised mean difference (SMD), interpreted as the magnitude of the number of standard deviation changes in the outcome for the intervention group as compared to students in non-programme schools. SMD scores are interpreted as the number of standard deviation changes in the outcome.

Where a single study reported estimates at multiple follow-up periods, only one of those is included. Where studies provided outcomes data at more than one follow-up, we combined the most similar follow-up period from different studies in the meta-analysis. For teacher attendance, the studies included in the meta-analysis reported on outcomes after approximately a twelve month follow-up. For teacher performance (assessment in instruction and student engagement), the studies included in the meta-analysis reported on outcomes at quite different follow-ups, ranging from nine to thirty months. For teacher performance (classroom management and use of materials), the studies included in the meta-analysis reported on outcomes after approximately a twelve month follow-up. Finally, for teacher performance (preparatory sessions), the studies included in the meta-analysis reported on outcomes after approximately a twenty-four month follow-up.

For enrolment, the studies included in the meta-analysis reported on outcomes after approximately a twelve month follow-up. For attendance, the studies included in the meta-analysis reported on outcomes at quite different follow-ups, ranging from nine to thirty months. For completion, the studies included in the meta-analysis reported on outcomes after approximately a twenty-four month follow-up. The studies providing outcomes data for drop-out had quite heterogeneous follow-up periods ranging from six to ninety-six months. For composite test scores, the studies included in the meta-analysis reported on outcomes at fairly different follow-ups, ranging from nine to thirty months. For language arts and maths test scores, the majority of studies followed-up after around twelve months.

For effects related to learning, some studies reported multiple effects relating to different exams for the same outcome measure (for example, both official and NGO language exams) and for different languages (for example, English and Swahili in Kenya). For all learning outcomes, we combined the most similar tests in the meta-analysis. For language arts, we chose to combine Kiswahili test results in Kenya (rather than English results) with local language test scores from the other included studies.

Five studies included multiple treatment arms. The Andra Pradesh Randomized Evaluation (APRest) in India compares two types of incentive schemes and also provides a comparison of incentive treatments vs. no intervention (Muralidharan and Sundararaman, 2008) and the META programme in Peru is a mixed individual and group incentive scheme (Cueto *et al.,* 2008). For Muralidharan and Sundararaman (2011), we chose to include both the individual incentive treatment arm and the group incentive treatment arm in the meta-analysis. For Behrman *et al.* (2013), we chose to include treatment arm two in the meta-analysis as this is the only treatment arm that provided an incentive scheme that rewarded teachers. One study provided effect estimates for three different incentive treatment arms, each further split into two further 'large' and 'small' incentive groupings (Loyalka, 2015).[83]. Throughout, we chose to combine the 'large' incentive for the levels treatment arm in meta-analyses as it was deemed to be the most similar to the other incentive designs.

There was one instance where we included two different studies that uses the same dataset but had different author teams (Gallego, 2008a; Contreras and Rau, 2012). These studies assess the effects of the Sistema Nacional de Evaluación del Desempeño (SNED) in Chile, using different analysis and assessing different outcomes. These two evaluations were conducted over mostly overlapping time periods using the same data set. To avoid dependency between effect estimates from these two studies, we included learning outcomes from Gallego (2008) as this study was more comparable to other studies reporting on learning outcomes. We included any additional outcomes reported in Contreras and Raul (2012).

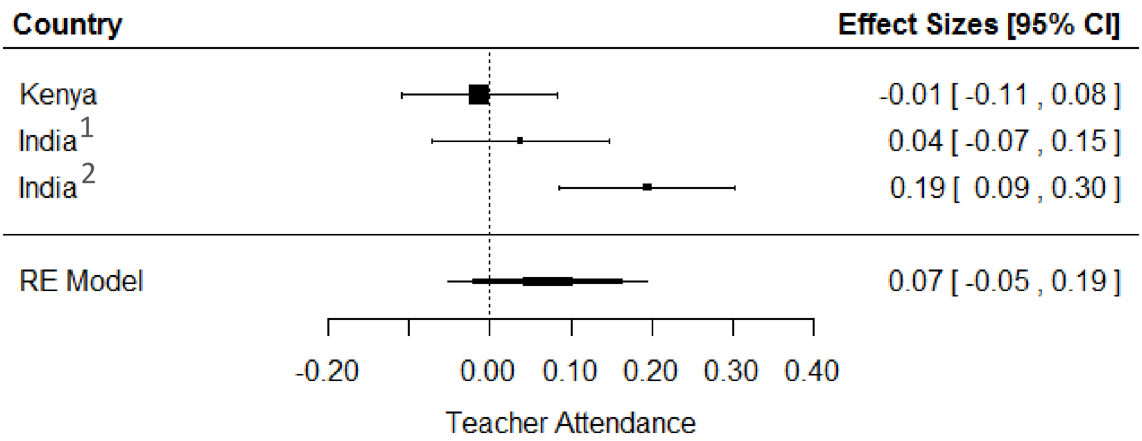
All effect estimates not included directly in the meta-analyses are reported narratively if they provide substantively different findings than those included in the meta-analysis. Estimates that are not substantively different than those already included in the meta-analysis are reported in technical Appendix H, section 7.1.

Teacher attendance

The overall average effect of teacher incentives on teacher attendance is 0.07, 95% CI [-0.05, 0.19], calculated under a random-effects model. The assessment of homogeneity suggests that the effects do not arise from a common population (I^2 = 75.73%, τ^2 = 0.0943, Q (df = 2) = 8.27, p = 0.0160). Figure 7.1b presents the forest plot with the results of the individual studies and the pooled point estimate. Only the study of the Seva Mandir Teacher incentive programme shows a positive, statistically significant effect. Removing this study reduces the effect estimate to 0.01 and increases the precision of the estimate 95% CI [-0.06, 0.08], increasing confidence in the conclusion of no effect. The study also accounts for all of the between-study heterogeneity in the meta-analysis – removing the study changes the heterogeneity estimates to I^2 = 0.00%, τ^2 = 0.000 (See Appendix H for results of all sensitivity analyses).

[83] See Table 7.1a for an overview of this intervention and the three treatment arms.

Figure 7.1 b: Teacher Attendance[84]



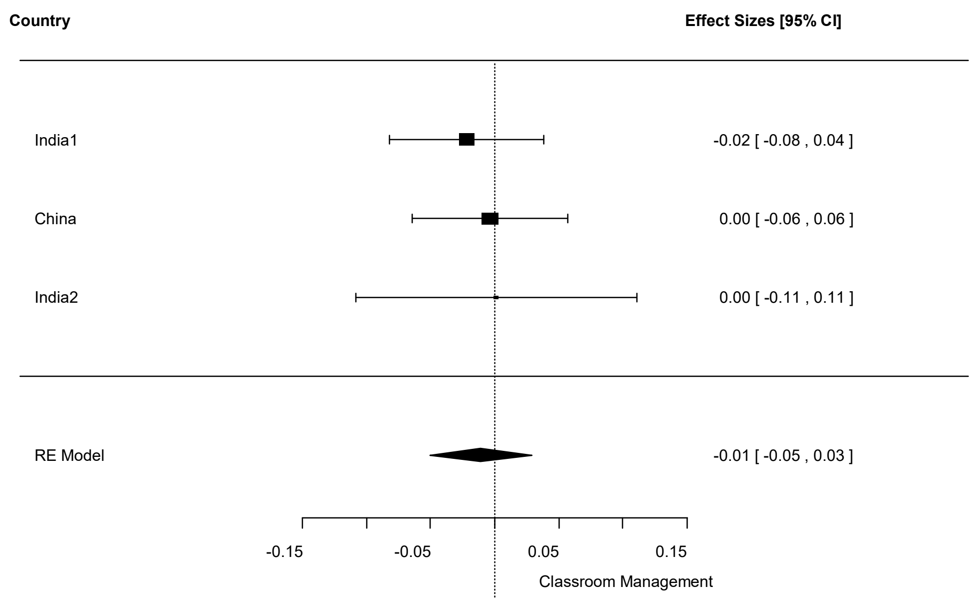
Teacher performance: classroom management

Three studies reported on this outcome. One measured teacher control of the class (Muralidharan and Sundararaman, 2011), another the number of children observed sitting within the classroom (Duflo *et al.,* 2012) and a final one asked students whether teachers could manage classrooms (Loyalka, 2015). The overall average effect of teacher incentives on teacher classroom management performance calculated under random effects is virtually zero (SMD = -0.01, 95% CI [-0.05, 0.03]). The assessment of homogeneity suggest that the only source of variation is within-study sampling error (I^2 = 0.00%, τ2 = 0.0012, Q (df = 2) = 0.2265, p = 0.8929). Figure 7c presents the forest plot with the results of the individual studies and the pooled point estimate. The forest plot shows considerable overlap between the confidence intervals of the included effect sizes, further supporting the findings of no between-study heterogeneity..[85]

[84] India1: Muralidharan & Sundararaman, 2011; Andhra Pradesh Randomized Evaluation
India2: Duflo et al., 2012, Seva Mandir Teacher Incentive programme
Kenya: Glewwe et al., 2010; ICS Teacher incentive programme

[85] It is worth noting that the effect included in the meta-analysis in Loyalka et al., (2015) was for 'incentive arm 1 – student average test scores' as this is most alike the other incentive interventions included. Loyalka et al., (2015) also evaluate two other treatment arms based on 'average student gain' and 'pay for percentile' and neither produces a substantively different effect from that reported for 'incentive arm 1 – student average test scores' in the meta-analysis.

Figure 7.1 c: Teacher Performance: Classroom Management[86]

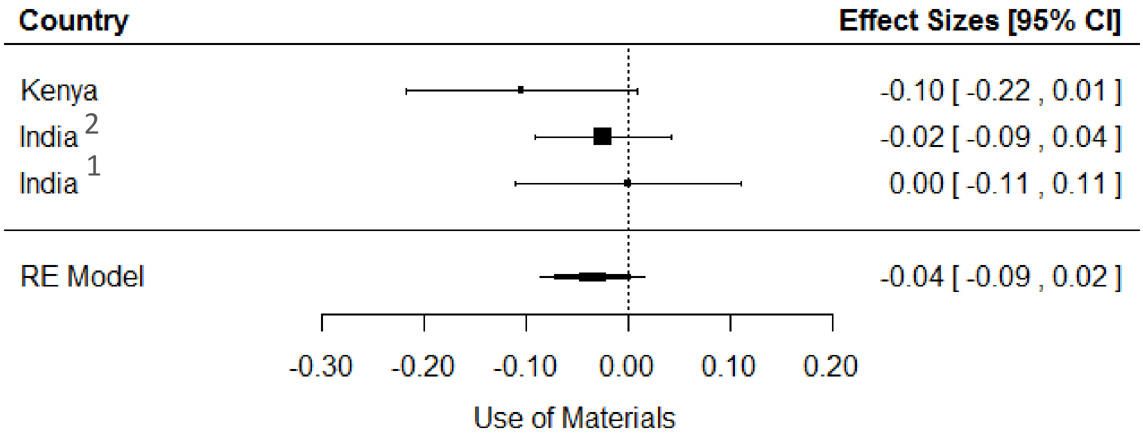


Teacher performance: use of materials

Three studies assessed teachers' use of materials, as measured by teachers' use of blackboards. The overall effect is negative but small and not statistically significant (SMD = -0.04, 95% CI [-0.09, 0.02]). The assessment of homogeneity suggest that the only source of variation is within-study sampling error (I^2 = 0.02%, τ2 = 0.0006, Q(df = 2) = 1.92, p = 0.3832). Figure 7d presents the forest plot with the results of the individual studies and the pooled point estimate. As expected the confidence intervals overlap for these studies.

[86] China: Loyalka, 2015; *Pay by Design programme*
India1: Muralidharan & Sundararaman, 2011; Andhra Pradesh Randomized Evaluation
India2: Duflo et al., 2012, Seva Mandir Teacher Incentive programme

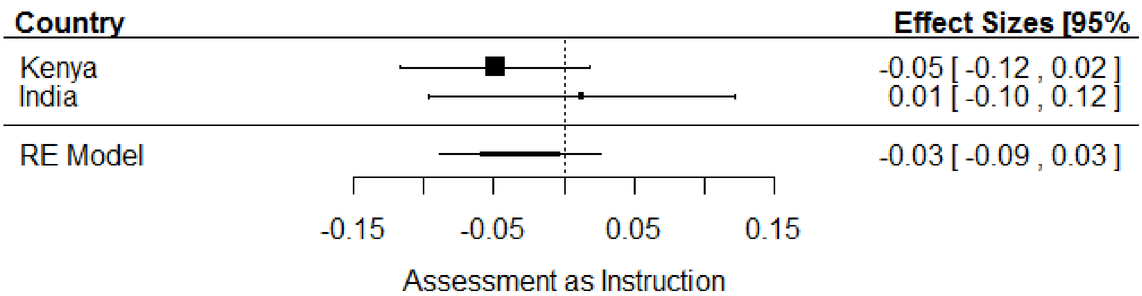
Figure 7.1 d: Teacher Performance: Use of Materials[87]



Teacher performance: use of assessment in instruction

Both studies that reported use of assessment as instruction employed measures of teachers assigning homework. The overall average effect suggest that teacher incentives had no effect on teacher's use of materials – as above the effect is negative but small and not statistically significant (SMD = -0.03, 95% CI [-0.09, 0.03]). The assessment of homogeneity suggest the only source of variation is within-study sampling error (I^2 = 0.00%, τ2 = 0.0000, Q(df = 1) = 0.87, p = 0.3508). Figure 7.1e presents the forest plot with the results of the individual studies and the pooled point estimate, as expected the confidence intervals overlap for these studies.

Figure 7.1 e: Teacher Performance: Assessment as Instruction[88]



Teacher performance: preparatory sessions

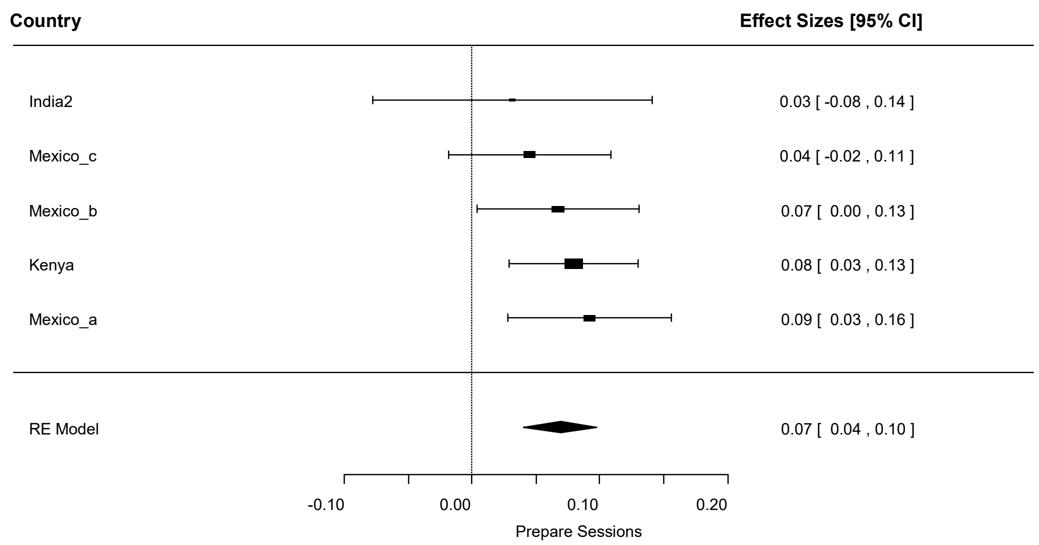
Three studies reported on preparatory session outcomes. The overall average effect of teacher incentives on teachers conducting preparatory sessions is 0.07, 95% CI [0.04, 0.10]), calculated under a random-effects model. The assessment of homogeneity suggests that the only source of variation is within-study sampling error (I^2 = 0.00%, τ2 = 0.0000, Q (df = 4) = 1.67, p = 0.7964). Figure 7.1f presents the forest plot with the results of the individual studies and the pooled point estimate, as expected the confidence intervals overlap for these studies.

[87] India1: Muralidharan & Sundararaman, 2011; Andhra Pradesh Randomized Evaluation
India2: Duflo *et al.,* 2012, Seva Mandir Teacher Incentive programme
Kenya: Glewwe *et al.,* 2010; ICS Teacher incentive programme

[88] India: Muralidharan & Sundararaman, 2011; Andhra Pradesh Randomized Evaluation
Kenya: Glewwe *et al.,* 2010; ICS Teacher incentive programme

Sensitivity analysis indicates the pooled effect estimate is not sensitive to the removal of any one of the included studies. Together with the homogeneity of the effect sizes, these findings indicate that teacher incentive programmes may have a small positive effect on teachers' effort in preparing their students for tests. Glewwe *et al.* (2010) assess the outcome of interest after 12 and 24 months. The estimate after 24 months included in the meta-analysis is positive and statistically significant (SMD = 0.08, 95% CI [0.03, 0.13]) and can be contrasted with the estimate after 12 months, which is notably smaller and not statistically significant (SMD = 0.04, 95% CI [-0.01, 0.09]).

Figure 7.1 f: Teacher Performance: Preparatory Sessions[89]

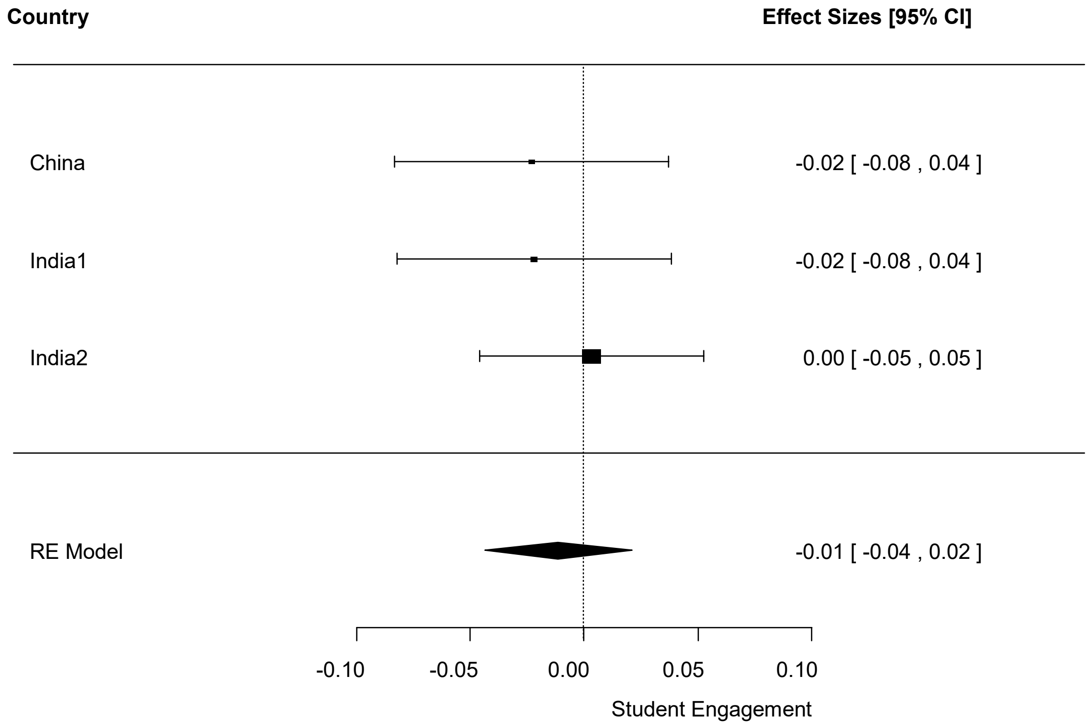


Teacher performance: student engagement

Three studies reported on student engagement outcomes. The overall average effect suggests that teacher incentives had no effect on teachers' engagement of students - the effect is negative but very small and not statistically significant (SMD = -0.01, 95% CI [-0.04, 0.03]). The assessment of homogeneity suggest that the only source of variation is within-study sampling error (I^2 = 0.00%, τ2 = 0.0000, Q(df = 1) = 0.6019 p = 0.7401). Figure 7.1g presents the forest plot with the results of the individual studies and the pooled point estimate, as expected the confidence intervals overlap for these studies.

[89] India2: Duflo *et al.,* 2012, Seva Mandir Teacher Incentive programme
Kenya: Glewwe *et al.,* 2010; ICS Teacher incentive programme
Mexico_a: Mexico: Behrman *et al.* 2013; ALI (T2); grade 10
Mexico_b: Mexico: Behrman *et al.* 2013; ALI (T2); grade 11
Mexico_c: Mexico: Behrman *et al.* 2013; ALI (T2); grade 12

Figure 7.1 g: Teacher Performance: Student Engagement[90]



Teacher performance: teacher effort

A single study provided a measure of 'teacher effort' as reported by students. Loyalka (2015) reports estimates for teacher effort for each of the three incentive arms described in the study (incentives based on: T1 student average test scores; T2 average student gain; T3 pay-for-percentile).[91]. There was a difference in the magnitude of the effect estimates reported for each treatment arm, but the confidence intervals of all estimates cross the line of no effect (T1 student average test scores SMD = 0.13, 95% CI [-0.13, 0.39]; T2 average student gain SMD = 0.01, 95% CI [-0.25, 0.26]; T3 pay-for-percentile SMD = -0.06, 95% CI [-0.31, 0.20]).

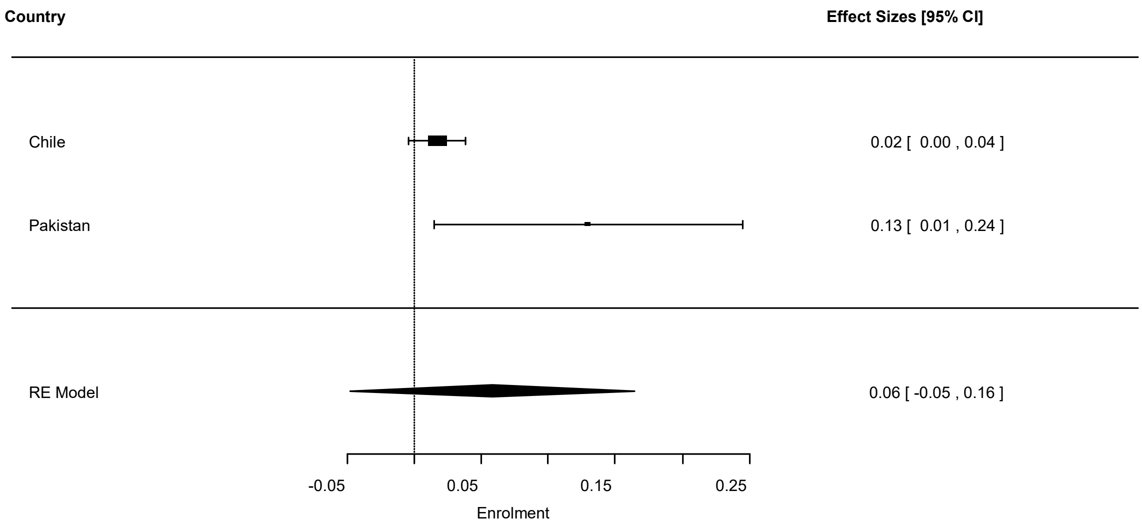
Enrolment

Two of the studies evaluating teacher incentive programmes measured student enrolment (Gallego, 2008a, Barrera-Osorio, 2015). The former (Gallego, 2008a) found no substantive effect of the programme on enrolment – the effect estimate is positive but small and not statistically significant (SMD=0.02, 95% CI [0.00, 0.004]) while the latter found positive effects on enrolment, but only in the third year of the intervention and when incentives were awarded to both teachers and head-teachers in a school (Barrera-Osorio, 2015), suggesting the length of time that incentives have been in place is important in producing an observable effect.

[90] China: Loyalka, 2015; *Pay by Design programme*
India1: Muralidharan & Sundararaman, 2011; Andhra Pradesh Randomized Evaluation
India2: Duflo *et al.,* 2012, Seva Mandir Teacher Incentive programme

[91]See Table 7.1a for an overview of this intervention and the three treatment arms.

Figure 7.1 h: Enrolment[92]

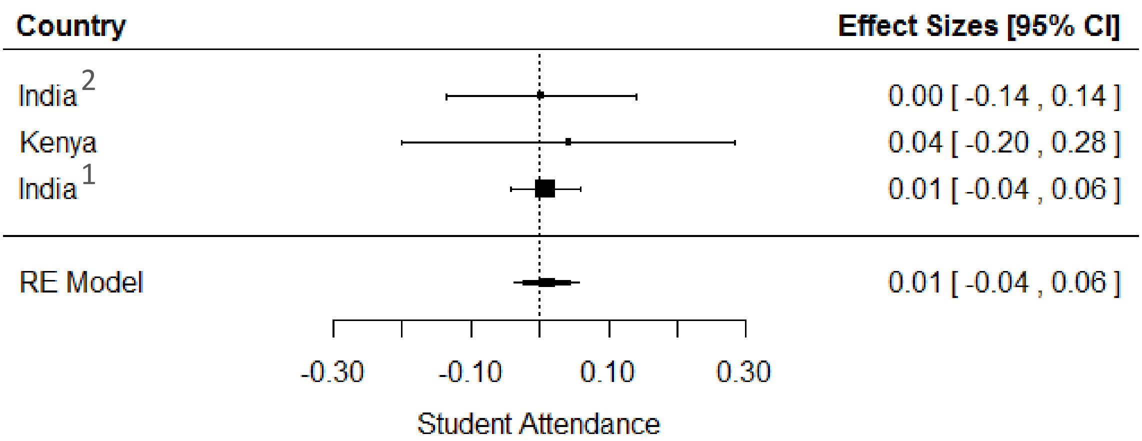


Attendance

The overall average effect of teacher incentives on student attendance is virtually zero (SMD = 0.01, 95% CI [-0.04, 0.06]), calculated under a random-effects model. The assessment of homogeneity suggest that the only source of variation is within-study sampling error (I^2 = 0.00%, $\tau^2 = 0$, Q (df = 2) = 0.08, p = 0.9605). Figure 7.1h presents the forest plot with the results of the individual studies and the pooled point estimate. As expected, the confidence intervals overlap for these studies. The average effect is not sensitive to the exclusion of any one of the included studies (See Appendix H for results of all sensitivity analyses).

[92] Chile: Gallego *et al.,* 2008; Sistema Nacional de Evaluación del Desempeño (SNED)
Pakistan: Barrera-Osorio, 2015; Improvers Bonus Programme for Government School Teachers in Pakistan

Figure 7.1 i: Attendance[93]



Completion

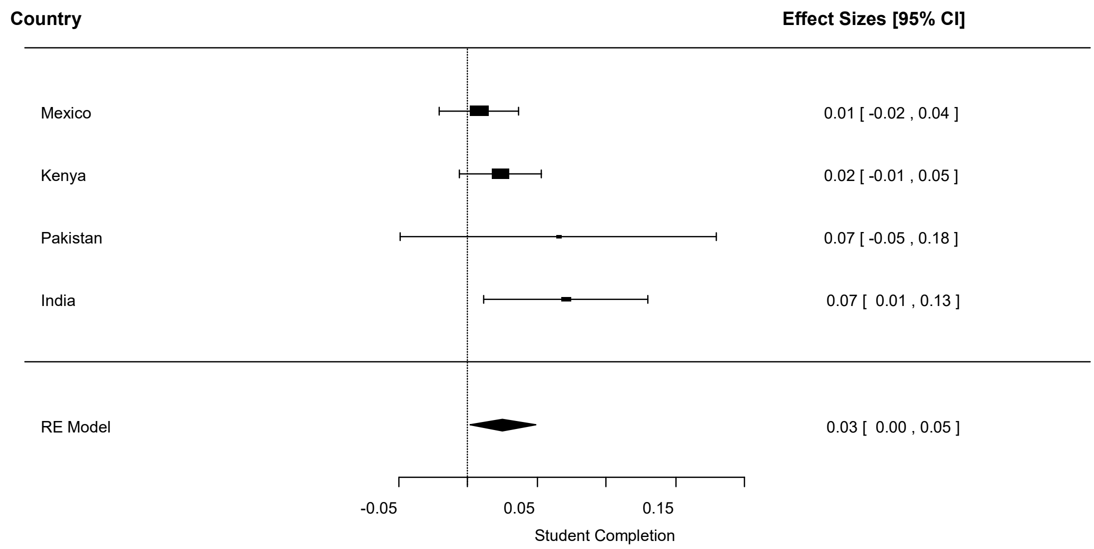
The overall average effect of teacher incentives on completion is small (SMD = 0.03, 95% CI [0.00, 0.05]). The assessment of homogeneity suggests a relatively small amount of between-studies variability (I^2 = 22.64%, τ^2 = 0.0001, Q (df = 3) = 4.0778, p = 0.2532). This can be seen when inspecting the forest plot in Figure 7.1j. Removing the study from India (Duflo *et al.,* 2012) from the meta-analysis removes the between-study variability in the model, but does not substantively change the point estimate (SMD = 0.02, 95% CI [0.00, 0.05], see Appendix H for results of all sensitivity analyses).

Barrerra-Osorio's (2015) provides various other estimates evaluating the effectiveness of the different incentive arms (head teacher incentive; head teacher and teacher incentive; high head teacher incentive, normal teacher incentive[94]). The one included in the meta-analysis is the estimate for the teacher and head-teacher incentives arm, after 24 months (the follow-up period closest to the other estimates included in the meta-analysis). The other estimates from Barrerra-Osorio (2015) indicate that incentives were had larger effects in the third year of the intervention. All three incentive arms had positive and statistically significant effects in the third year of the intervention, with the 'high head teacher incentive, normal teacher incentive' treatment arm producing the largest effect (SMD = 0.30, 95% CI [0.18, 0.42], all estimates relating to the Improvers Bonus Program for Government School Teachers in Pakistan in section 7.1 of Appendix H).

[93] India1: Muralidharan & Sundararaman, 2011; Andhra Pradesh Randomized Evaluation
India2: Duflo *et al.,* 2012, Seva Mandir Teacher Incentive programme
Kenya: Glewwe *et al.,* 2010; ICS Teacher incentive programme

[94] See Table 7.1a for an overview of this intervention and the treatment arms.

Figure 7.1 j: Completion[95]

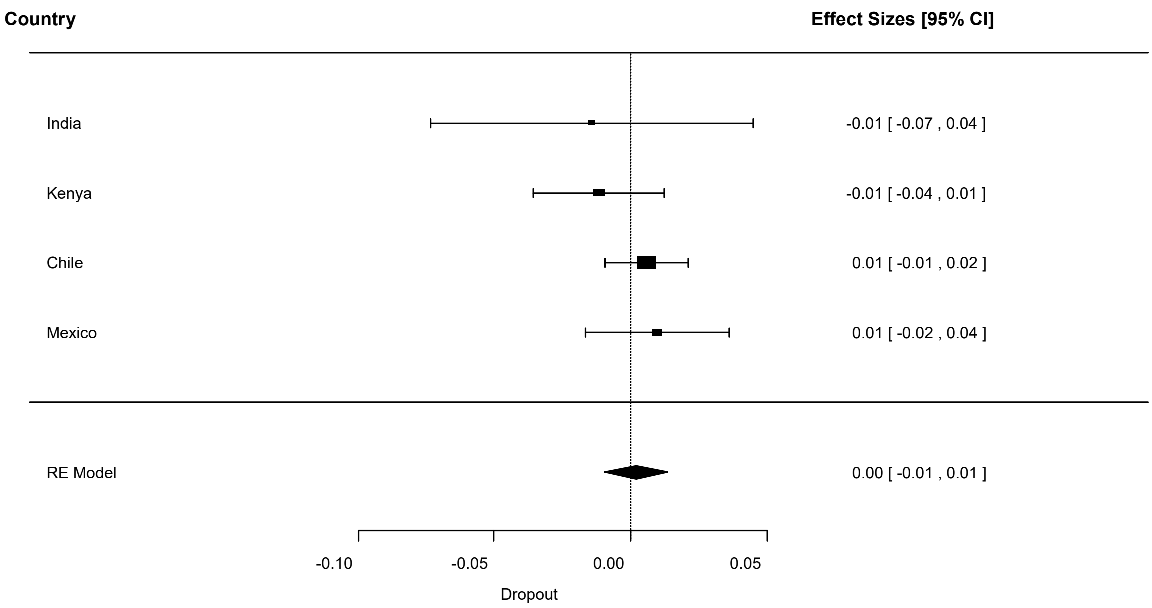


Drop-out

The overall average effect of teacher incentives on drop-out is zero (SMD = 0.00, 95% CI [-0.01, 0.01]), calculated under a random-effects model. The assessment of homogeneity suggests that the only source of variation is within-study sampling error (I^2 = 0.00%, τ2 = 0.0000, Q (df = 3) = 2.0754, p = 0.5569). Figure 7.1k presents the forest plot with the results of the individual studies and the pooled point estimate, as expected the confidence intervals overlap for these studies. These results are not sensitive to the removal of any one of the included studies from the meta-analysis (See Appendix H for results of all sensitivity analyses).

[95] India: Duflo *et al.,* 2012; ICS Teacher incentive programme
Kenya: Glewwe *et al.,* 2010; Seva Mandir Teacher Incentive programme
Mexico: Behrman *et al.* 2013; ALI (T2)
Pakistan: Barrera-Osorio, 2015; Improvers Bonus Programme for Government School Teachers in Pakistan

Figure 7.1 k: Dropout[96]



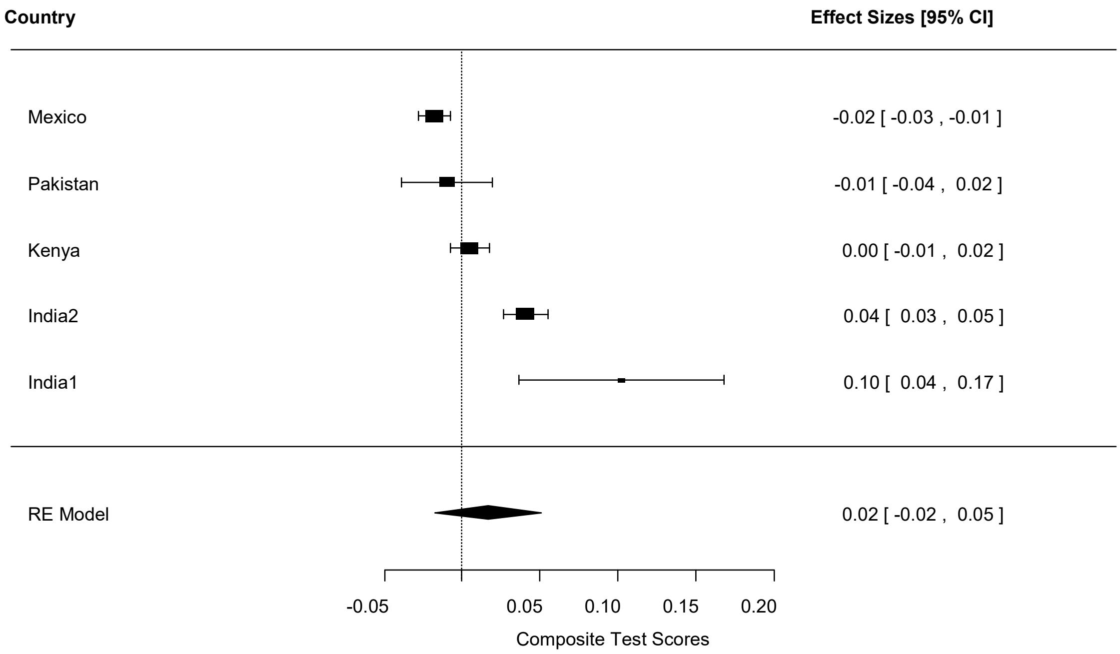
Composite test scores

The overall average effect of teacher incentives on learning outcomes as measured by a composite score is small and not statistically significant (SMD = 0.02, 95% CI [-0.02, 0.05]), calculated under a random effects model. However, the homogeneity test (I^2 = 95.06%, τ^2 = 0.0013, Q (df = 4) = 52.7876, p = <0.0001) suggests that this pooled estimate includes a large amount of between-study variability, indicating that the effects did not arise from the same population.

Figure 7.1l supports the presence of heterogeneity. The effect sizes range from -0.02, 95% CI [-0.03, -0.01] in Mexico (Santibañez, 2007) to 0.10, 95% CI [0.04, 0.17] in India (Duflo et al., 2012). Three of the studies are clustered around the line of no effect, and two of the studies suggest beneficial effects. The confidence intervals between these two groups of studies do not overlap. Nevertheless, sensitivity analysis indicates that removing any one of these studies does not make a substantive difference to the overall pooled effect (see Appendix H for results of all sensitivity analyses).

[96] Chile: Gallego *et al.,* 2008; Sistema Nacional de Evaluación del Desempeño (SNED)
India: India: Duflo *et al.,* 2012; ICS Teacher incentive programme
Kenya: Glewwe *et al.,* 2010; Seva Mandir Teacher Incentive programme
Mexico: Behrman *et al.* 2013; ALI (T2)

Figure 7.1 l: Composite Test Scores[97]



The study of the ICS teacher incentive programme in Kenya (Glewwe *et al.,* 2010) also reports results of an alternative test (the ICS NGO test)[98]. The effect estimate for this alternative test in year 2 indicates that the results are not sensitive to the type of test that is used to measure learning (SMD = 0.01, 95% CI [0.00, 0.02]).

The study of the Seva Mandir Teacher incentive programme in India (Duflo *et al.,* 2012) also reports composite test scores separately for girls and boys. The effect estimate for girls is slightly higher than that for boys with the effect for girls SMD = 0.14, 95% CI [0.04, 0.24]), while for boys it is SMD = 0.10, 95% CI [0.01, 0.20]. While girls perform marginally better on composite test outcomes, the difference is not substantive enough to conclude that teacher incentive programmes affect girls and boys differently.

Language arts test scores

The overall average effect of teacher incentives on language arts test scores is zero and not statistically significant (SMD = 0.00, 95% CI [-0.13, 0.12]), calculated under a random effects model. However, the homogeneity test (I^2 = 98.57%, τ^2 = 0.0241, Q (df = 6) = 52.00, p = <0.0001) suggests that this pooled estimate captures a lot of between-study variability, indicating that the effects did not arise from the same population.

Figure 7.1m supports the presence of heterogeneity and the effect sizes range from -0.45, 95% CI [-0.64, 0.26] for grade five students in Peru (Cueto *et al.,* 2008), to 0.11, 95% CI [-0.06, 0.29] for grade four students in Peru (same study). The confidence intervals are

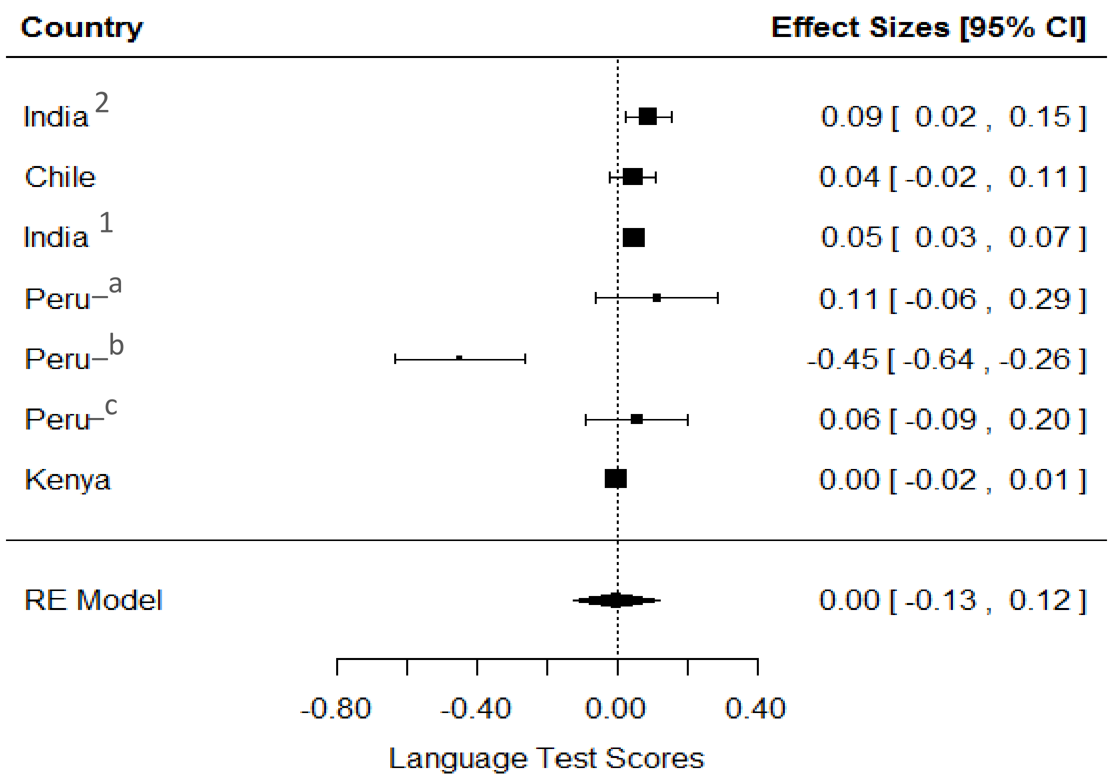
[97] India1: Duflo *et al.,* 2012, Seva Mandir Teacher Incentive programme
India2: Muralidharan & Sundararaman, 2011; Andhra Pradesh Randomized Evaluation
Kenya: Glewwe *et al.,* 2010; ICS Teacher incentive programme
Mexico: Santibañez, 2007; Carrera Magisterial
Pakistan: Barrera-Osorio, 2015; Improvers Bonus Programme for Government School Teachers in Pakistan

[98] The ICS test is an exam conducted by the implementing NGO (Glewwe *et al.,* 2003). The measure included in the meta-analysis reports the average district government test score.

overlapping between most of the included studies. The results are sensitive to the removal of the grade 5 effect estimate from Peru which changes the pooled effect estimate to a positive and marginally statistically significant effect (SMD = 0.04, 95% CI [0.00, 0.08] (see Appendix H for results of all sensitivity analyses).

Three studies report effects at a later follow up (Glewwe *et al.,* 2010; Muralidharan and Sundararaman, 2011; Contreras and Rau, 2012) and find a slight increase in the second year (Kenya: SMD= 0.02, CI 95% [0.00, 0.03]; India: SMD= 0.07, 95% CI [0.04, 0.09]; Chile: SMD= 0.12, 95% CI [0.05, 0.18]).

Figure 7.1 m: Language Arts Test Scores[99]



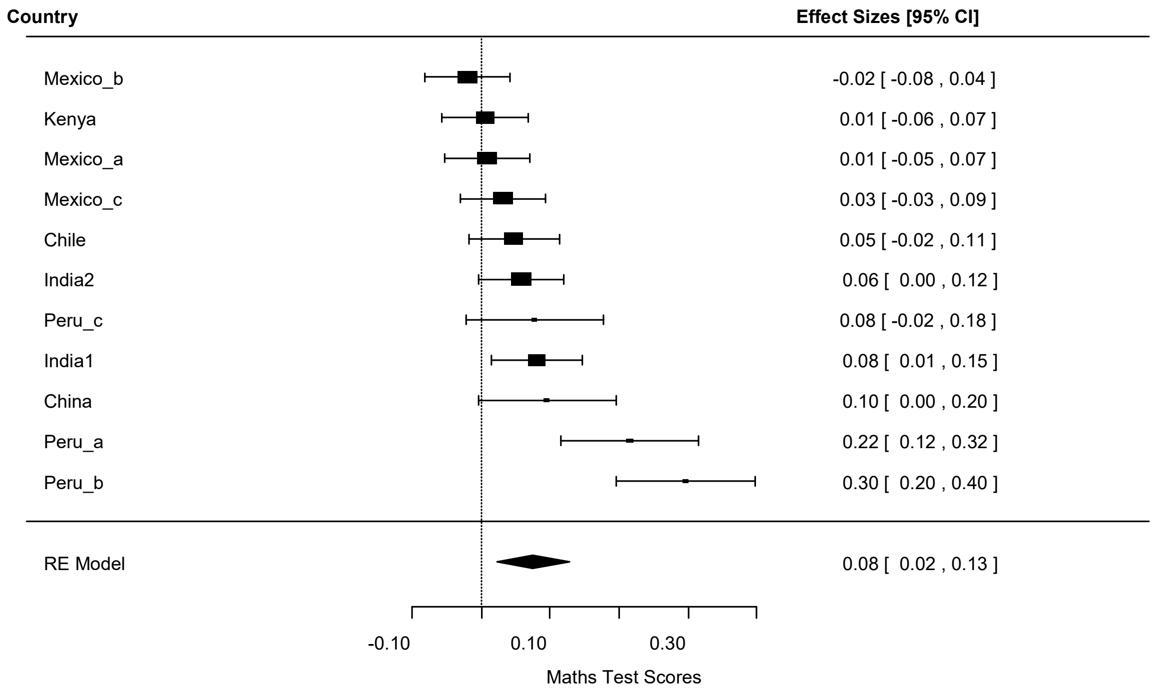
The study of the Seva Mandir Teacher incentive programme in India (Duflo *et al.,* 2012) also reports language arts test scores separately for girls and boys. Again, the effect estimate for girls is slightly higher than that for boys, with the effect for girls SMD = 0.13, 95% CI [0.04, 0.23], while for boys it is SMD = 0.10, 95% CI [0.01, 0.20]. However, the difference is not large enough to conclude that teacher incentive programmes affect girls and boys differently.

Maths test scores

The overall average effect of teacher incentives on math test scores is 0.08, 95% CI [0.02, 0.13], calculated under a random effects model. The homogeneity test suggest a high amount of between-study variability ($I^2 = 82.27\%$, $\tau^2 = 0.0063$, Q (df=10) = 44.25, p = < 0.0001). The effect sizes range from -0.02, 95% CI [-0.08, 0.04] in Mexico (Behrman et al., 2012) to 0.30, 95% CI [0.20, 0.40] in Peru (Cueto et al., 2008), and figure 7.1n supports the presence of heterogeneity. The overall results are not substantively sensitive to the removal of any of these studies.

[99] Chile: Contreras and Rau, 2012; Sistema Nacional de Evaluación del Desempeño (SNED); India1: Muralidharan & Sundararaman, 2011; Andhra Pradesh Randomized Evaluation; India2: Duflo *et al.,* 2012, Seva Mandir Teacher Incentive programme; Kenya: Glewwe *et al.,* 2010; ICS Teacher incentive programme; Peru_a: Cueto *et al.,* 2008; grade 4; Peru_b: Cueto *et al.,* 2008; grade 5; Peru_c: Cueto *et al.,* 2008; grade 6.

Figure 7.1 n: Maths Test Scores[100]



Similar to findings for composite test scores and language arts test scores there is a slight increase in effects observed in India and Kenya in the second year, although these are small in magnitude (India: SMD=0.09, 95% CI [0.03, 0.15]; Chile: SMD=0.14, 95% CI [0.07, 0.20]). However, this is not observed in the longer follow-ups reported for Mexico and Kenya.

Loyalka (2015) also reports a series of further results for the different treatment arms of the *Pay by Design programme*, each split by small and large incentives[101]. The effect reported in the meta-analysis is the 'large' incentive for the 'levels' treatment arm as this was most similar to the other included studies, but the effect of the 'pay-for-percentile large incentives' group is larger in magnitude and the estimate is more precise (SMD = 0.19, 95% CI [0.10, 0.30]).

Two studies report maths outcomes split by gender (Duflo *et al.,* 2012; Behrman *et al.,* 2012). Combining the effects for these studies for girls and boys respectively we found results that are slightly smaller in magnitude and less precise than the effect for the full sample, but not substantively different from each other (Girls: SMD=0.06, 95% CI [-0.09, 0.21], I^2 = 63.20%, τ^2 = 0.0077, Q (df = 1) = 2.71, p=0.0993; Boys, SMD=0.06, 95% CI [-0.03, 0.14], I^2 = 12.47%, τ^2 = 0.0006, Q (df = 1) = 1.14, p=0.2851).

[100] Chile: Contreras and Rau, 2012; Sistema Nacional de Evaluación del Desempeño (SNED); China: Loyalka, 2015; *Pay by Design programme;* India1: Muralidharan & Sundararaman, 2011; Andhra Pradesh Randomized Evaluation
India2: Duflo *et al.,* 2012, Seva Mandir Teacher Incentive programme; Kenya: Glewwe *et al.,* 2010; ICS Teacher incentive programme; Mexico a; Behrman *et al.* 2012; ALI; grade 10; Mexico_b; Behrman *et al.* 2012; ALI; grade 11; Mexico_C; Behrman *et al.* 2012; ALI; grade 12; Peru_a: Cueto *et al.,* 2008; grade 4; Peru_b: Cueto *et al.,* 2008; grade 5; Peru_c: Cueto *et al.,* 2008; grade 6

[101] See Table 7.1a for an overview of this intervention and the treatment arms.

7.1.4 Summary of findings and discussion

We identified ten studies of teacher incentives interventions across Latin America and the Caribbean, Sub-Saharan Africa, South Asia and East Asia and the Pacific. We were able to examine effects on teacher attendance, teacher performance, enrolment, student attendance, completion, drop-out, maths, language arts and composite test scores using meta-analysis.

The overall average effects range from -0.04, 95% CI [-0.09, -0.02] for teacher performance (use of materials) to 0.08, 95% CI [0.02, 0.15] for maths test scores. Overall, the results show that while incentives interventions can produce positive outcomes, many fail to improve outcomes of interest, while in some cases this type of intervention may produce negative effects. There is also a large amount of between study variability for many of our outcomes of interest. Individual effect sizes range from -0.45, 95% CI [-0.64, -0.26] for language arts test scores to 0.30, 95% CI [0.20, 0.40] for completion.

The large amount of heterogeneity is not surprising given the variability in our included incentive interventions' design, context and populations. Due to the limited number of included studies for any single outcome we were not able to undertake any moderator analysis to assess which factors may be driving the results. However, we discuss the results below incorporating any findings from the qualitative synthesis, which may help to explain these findings (Section A7.1 in appendix J provides the results of the full qualitative synthesis, tables 7.1c and 7.1d provide the summary findings from this synthesis).

The findings of the meta-analysis suggest teacher incentive programmes do not improve teacher attendance, with the exception of the Seva Mandir Teacher incentive programme in India which finds a positive effect on teacher attendance (Duflo *et al.,* 2012). This could potentially be due to the different design features of the studies included in the meta-analysis. The Kenyan ICS Teacher incentive programme and the Indian Antra Pradesh Randomised Evaluation of teacher incentives that showed no effect on teacher attendance were intended to increase student test scores by awarding incentives based on improvements in student learning. Their focus was not on increasing teacher attendance and it is not clear whether absenteeism was a problem in either of these two cases (Duflo et al., 2012, Glewwe *et al.,* 2010, Muralidharan and Sundararaman, 2011). However, the Seva Mandir Teacher Incentive programme, which had a positive effect on teacher attendance, was specifically aimed and designed to incentivise teacher attendance. The programme monitored teacher attendance three times a day using photographs taken by students using tamper-proof cameras, rewarded every additional day of attendance beyond the minimum 20 days a month, and imposed a fine of 50 Rs (equivalent to 5% or base salary) for each day the teacher skipped work within the minimum 20 required work days (Duflo *et al.,* 2012).

The findings also suggest that teachers do not seem to alter their teaching approach and instruction techniques in response to incentives, with the exception of an increased focus on preparatory sessions for tests. The authors of the evaluation of the Carrera Magisterial reported evidence suggesting teachers dedicated extra time to test preparation and hypothesise that this additional effort might have been partly responsible for the positive effects observed in some programmes (Santibañez, 2007). However, the effect appears small (the pooled effect is 0.07, 95% CI [0.04, 0.10]), and none of the effects included in the meta-analysis exceeds 0.09 SMD. All three programmes that contributed effect sizes to the meta-analysis on preparatory sessions (Behrman *et al.,* 2012, Glewwe *et al.,* 2010, Muralidharan and Sundararaman, 2011) awarded incentives to teachers based on student performance on test scores, rather than a more comprehensive score that would take into account other teacher performance dimensions. These findings align with some of the theories of change for teacher incentive programmes that suggest that teachers may focus on improving the observable measures of their performance for which they are rewarded.

The meta-analysis findings indicate that teacher incentives do not seem to have an effect on access-related outcomes such as student attendance or drop-out and only small effects on enrolment and completion. There are few studies reporting on these outcomes but among those that do, the findings of homogeneity tests indicate that there is no or small between-study variability and where relevant, results of sensitivity analysis indicate that these results are not sensitive to the removal of any of the included effect sizes. The findings of the qualitative analysis do not provide any insight for why this might be the case. However, the lack of sizeable effects is not surprising given that the primary goal of incentives interventions is to improve teacher performance and ultimately student test scores.

The findings about the effect of teacher incentives on learning outcomes are mixed. The meta-analysis indicates that teacher incentive programmes are on average effective in improving maths learning outcomes, though the pooled effect size is relatively small. The effect sizes for language arts and composite test scores are very heterogeneous across the included studies, so the pooled effect estimates for these two outcomes may not be meaningful. However, even with consideration of this heterogeneity, the size of the effects is generally small, with the largest effect size not exceeding 0.11 for language arts test scores and 0.10 for composite test scores. Findings from measures taken at different follow-up periods indicate that the effect of teacher incentives on all three types of learning outcomes may be larger when measured over longer follow-up periods. There is no strong evidence that teacher incentive programmes differentially affect girls' and boys' learning. However, as very few studies reported findings by sub-group, our ability to draw reliable conclusions based on this evidence is limited.

Various other factors identified in the qualitative synthesis might explain why the interventions had limited effects on teacher effort and student learning, particularly in the first year of implementation. Three studies note that information about the intervention was not provided in a timely manner or was not sufficiently clear *(Behrman et al., 2012, Glewwe et al.,* 2010, Gallego, 2008b), possibly limiting take-up. One study reports that the incentive programme was not implemented as intended, incentivising seniority and levels of education instead of performance (Santibañez, 2007) and one study note problems implementing the monitoring component, discouraging continuation of the programme (Cueto *et al.,* 2008). Two studies note that the way the intervention was delivered made it difficult for participants to succeed (Santibañez, 2007, Gallego, 2008b) and two studies noted a lack of transparency about the decision process used to award the incentives may have discouraged participation and teacher effort (Behrman *et al.,* 2012, Gallego, 2008b). Three studies report general support among teachers for both the specific incentive programmes implemented and the idea of performance-based pay in general (Gallego, 2008b; Glewwe *et al.,* 2010; Muralidharan and Sundararaman, 2008), though one case highlights that teachers did not feel comfortable with the idea of having their behaviour monitored (Gallego, 2008b). Finally, these interventions were generally implemented in the context of limited teacher accountability structures and norms and within systems where teacher performance is not typically rewarded (Cueto, 2008; Duflo *et al.,* 2012; Gallego, 2008b; Glewwe, 2010; Muralidharan and Sundararaman, 2008, 2011).

Table 7.1 c: Descriptive findings: Process and implementation

Descriptive findings: Process and implementation	Context	Intervetion	Citation (info type)
Delayed information provision about the programme may account for limited effectiveness in the first year	Mexico	ALI programme	Behrman *et al.,* 2012 (impact evaluation)
Poor messaging about how the programmes works resulted in schools / teachers feeling that the bonus was not achievable, possibly limiting take up, skepticism among teachers about the actual award of the bonus in the first year	Chile, Kenya, Pakistan	SNED programme, ICS Teacher incentive programme	Gallego, 2008b, Glewwe *et al.,* 2010, Barrera-Osorio *et al.* 2015 (impact evaluations)
The criteria for succeeding were difficult to meet due to time constraints or limited opportunities to demonstrate required competencies , possibly limiting take up over time	Mexico, Chile	Carrera Magisterial, SNED programme	Santibañez, 2007, Gallego, 2008b, Contreras and Rau, 2012 (impact evaluations)
Lack of transparency about decision process used to award the incentive may have discouraged take up	Mexico, Chile	ALI programme, SNED programme	Behrman *et al.,* 2012, Gallego, 2008b
The evaluation system in the first few years of the programme was not fully functional, resulting in promotions being based on seniority and education levels rather than teacher performance	Mexico	Carrera Magisterial	Santibañez, 2007 (impact evaluation)
Problems with implementing the daily monitoring of classes discouraged continuation of the programme	Peru	META	Cueto *et al.,* 2008 (impact evaluation)
Teachers disliked the idea of having their performance monitored despite the incentives	Chile	SNED programme	Gallego, 2008b (impact evaluation)
Participating teachers supported the use of incentives to motivate teachers	Kenya, Chile, India	ICS programme, SNED programme, Andra Pradesh Randomized Evaluation	Glewwe *et al.,* 2010, Gallego, 2008b, Muralidharan and Sundararaman,

Descriptive findings: Process and implementation	Context	Intervetion	Citation (info type)
			2008 (impact evaluations)
Additional interventions delivered in treatment and/or control group may have affected the effectiveness of the programme	Mexico, Chile, Kenya, India	ALI experiment, SNED programme, ICS programme, Seva Mandir programme	Behrman *et al.,* 2012, Contreras and Rau, 2012, Glewwe *et al.,* 2003, 2010, Duflo *et al.,* 2012 (impact evaluations)
Teachers dedicated extra time to student test preparations	Mexico	Carrera Magisterial	Santibañez, 2007 (impact evaluation)
Teacher incentives did not affect the characteristics of teachers in beneficiary schools with the exception of reducing the percentage of teachers working in other colleges	Chile	SNED programme	Gallego, 2008a (impact evaluation)
Teacher incentive programmes did not increase student efforts or change their behaviour in class	Mexico, India	ALI programme, Andra Pradesh Randomized Evaluation	Behrman *et al.,* 2012, Muralidharan and Sundararaman, 2008 (impact evaluations)
Teacher incentives with an additional component aimed to provide information to parents increased the percentage of parents that were informed about the academic results of their children	Chile	SNED programme	Gallego, 2008a (impact evaluation)
Households did not adjust their own inputs into their children's education	India	Andra Pradesh Randomized Evaluation	Muralidharan, 2011 (impact evaluation)

Descriptive findings: Process and implementation	Context	Intervetion	Citation (info type)
Monitoring component did not have an effect on learning outcomes when teacher and school performance was not made publically available	India	Andra Pradesh Randomized Evaluation	Muralidharan and Sundararaman, 2008 (impact evaluations)

Table 7.1 d: Descriptive findings: Contextual Factors

Descriptive findings: Contextual factors	Context	Intervetion	Citation/ info type
Teacher incentive programmes were implemented in contexts of weak teacher accountability structures where teacher unions are strong, parent committees and attendance monitoring weak, and disciplinary actions rare	India, Kenya	Andra Pradesh Randomized Evaluation, Seva Mandir Programme, ICS Programme,	Muralidharan and Sundararaman, 2008, Duflo *et al.,* 2012, Glewwe *et al.* 2010 (impact evaluations)
Teacher incentive programmes were implemented in contexts of weak teacher accountability norms and high teacher absenteeism	Peru, India	META programme, Seva Mandir programme, Andra Pradesh Randomized Evaluation	Cueto, 2008, Duflo *et al.,* 2008, Muralidharan and Sundararaman, 2008, 2011 (impact evaluations)
Teacher incentive programmes implemented in contexts where existing incentive structures did not typically reward performance	Peru, India, Chile, Kenya	META programme, Andra Pradesh Randomized Evaluation, SNED programme, ICS programme	Cueto, 2008, Muralidharan and Sundararaman, 2008, Glewwe *et al.,* 2010 (impact evaluations) Mizala and Urquiola, 2013 (Mixed methods study)
Teacher incentive programmes implemented in contexts where student performance standard were low	Peru, India	META programme, Andra Pradesh Randomized Evaluation,	Cueto, 2008, Muralidharan and Sundararaman, 2011 (impact evaluations)
Teacher incentive programmes implemented in contexts where student performance standard varied	Chile	SNED programme	Gallego, 2008b (impact evaluation)

7.2 Teacher hiring

Teacher hiring interventions are designed to increase the number and quality of teachers in schools. Some of them focus on hiring additional teachers, so that class size, the incidence of multigrade teaching (whereby teachers have to teach multiple grades in one class) and pupil-teacher ratios can be reduced, with the result that students receive more and better quality contact time. Others promote the employment of contract teachers instead of permanent civil-service teachers, with the guiding principle that employing teachers on short-term contracts can be economical, will increase incentives for teacher attendance and performance, and will still ensure that teachers are qualified and capable. Other teacher hiring interventions are designed to introduce new hiring and promotion processes that will increase the quality of new appointees.

7.2.1 How may teacher hiring interventions affect education outcomes?

Hiring additional teachers

Governments and donors have responded to increased student enrolment rates and high pupil-teacher ratios, particularly in hard-to-reach areas, by recruiting new, additional staff (Vegas *et al.,* 2013; Kingdon *et al.,* 2012). Hiring additional teachers is intended to reduce pupil-teacher ratios, the incidence of multi-grade teaching (where students in different grades are taught together) and the number of single-teacher schools. The central goal is to increase the quality of teaching; smaller class sizes facilitate more targeted tuition and students may receive increased individual attention and opportunities for participation in classes (Banerjee *et al.,* 2007). In the case of single-teacher schools, an additional teacher may also increase the number of days the school will be open by reducing the likelihood that teacher absence will affect school opening.

For additional teachers to have the desired effect, it is important that funding be additional and schools are not forced to reallocate resources from within their existing budgets. It is, of course, crucial that teachers are appointed to the schools they are intended for, especially in the case of one-teacher schools (Chin 2005). It is also important that the addition of a new teacher to a school does not lead to the reallocation of existing teachers or a reduction in pre-existing recruitment plans (Bold *et al.* 2013). The benefits produced by the recruitment of additional teachers may increase student satisfaction, reduce dropout, improve parental perceptions to education and thus result in an increase in enrolment and attendance (Krishnaratne *et al.,* 2013). These factors may create a feedback loop with the result that the pupil-teacher-ratio is not ultimately reduced (Bold *et al.* 2013).

Contract teachers

Some programmes recruit teachers on renewable contracts rather than to permanent civil-service positions, with the intention of recruiting teaching staff with similar educational qualifications at much lower cost. Contract teachers may also have greater incentives for good attendance rates and increased teacher effort as a result of performance-related contract renewal or permanent appointment (Bold et *al., 2013*). However, in practice contract teachers may be younger on average, less experienced and may not have pre-service training. As a result, these teachers may actually be less effective in the classroom (Atherthon *et al. 2010*). An effective mechanism that supports the supervision, training and coaching of contract teachers may be a crucial determinant of success. Contract teachers typically have the same responsibilities as civil-service teachers, but poorer employment conditions could potentially negatively impact their performance. This can result in a 'disgruntled worker effect' that negatively impacts student learning (Vegas and Laat, 2003).

Some interventions also monitor and assess performance (for example, by training school committees) in order to ensure that teaching quality is upheld (Duflo *et al.* 2009).

Methods of hiring or awarding promotion

Some teacher hiring interventions seek to have an effect on educational outcomes by adapting the hiring process used to appoint teachers and/or the processes used to decide how teachers can gain promotion. This may involve switching from traditional hiring or promotion procedures to examination based assignment or decentralising appointment and promotion decisions so that local authorities have more decision-making power.

The introduction of competitive testing as a means of appointing candidates is designed to increase the quality of teaching by ensuring that positions are filled by the most capable candidates (Estrada 2013). The theory assumes that the test is an adequate means of assessing potential quality of teaching. Making promotion contingent on teachers' ability to pass exams is equally intended to create a meritocracy in which the best and most dedicated teachers are rewarded (Ome, 2012).

Teaching staff may be appointed locally with involvement from the community and/or school committee. This approach aims to increase accountability and reduce teacher absenteeism by hiring teachers that are embedded in their local communities. Locally hired teachers are also likely to be socially and culturally similar to the students they teach, which may make them more effective (Kingdon *et al.,* 2012). Hiring female teachers is regarded as a strategy to break political and cultural barriers to female schooling (Kim *et al.,* 1998); parents may be more likely to send girls to school when female teachers are hired.

Clearly these theories are not mutually exclusive and many teacher hiring programmes combine a number of the elements described above to promote effectiveness. Increasing the quantity of teaching staff available in schools may only be of limited effectiveness if no consideration is made of how they are deployed, for how long and to what extent they address the unmet needs of the students (Banerjee *et al.* 2007). As a result, hiring additional teachers has been combined with a mix of incentives to change teacher behaviour (Duflo *et al.,* 2009). For example, many programmes combine the recruitment of additional teachers with new classroom space and teaching materials, arguably a prerequisite for additional teachers to be effective.

7.2.2 Description of included studies

Description of studies

We included eight studies reported in fourteen different papers that evaluated the effect of a teacher hiring intervention in a low-or-middle-income-country. These described eight unique programmes. We use the term 'study' to refer to a unique evaluation of a programme, which in one case was reported in more than one paper. In the following section, we describe the characteristics of these studies in detail. Table 7.2b provides an overview of the key characteristics of the included studies and table 7.2c gives an overview of intervention design features, including whether the programme meant that additional teachers were hired, the contractual status of new teachers and whether hiring of local or female staff was encouraged.

Population

All of the programmes targeted teachers; three were designed explicitly to target the recruitment of local teachers (Duflo, Dupas, Kremer, 2012; Bold *et al.,* 2013; Muralidharan and Sundararaman, 2013), while one study promoted the recruitment of female teachers (Chin, 2005). One intervention providing additional teachers to schools (Muralidharan and Sundararaman, 2013) reported that the new teachers were predominantly female (68 percent) and aged 26 on average. A further study examining multiple interventions with treatments involving contract and civil-service teachers reported the following; contract teachers were on average 27 years old, had 1.5 years of experience and 48 per cent of them were female; civil-service teachers were on average 42 years old, had over 15 years of experience and 62 per cent of them were female (Duflo, Dupas, Kremer, 2012). Interventions focussed on primary school children, with only one targeting secondary school students (Estrada, 2013). Only two studies provided information on the gender balance of students in treatment groups, with Estrada (2013) reporting that 42 per cent of students were female and Duflo, Dupas, & Kremer (2012) reporting that 49 per cent of students were female.

Setting

The included programmes were implemented in a broad range of settings. Three in South Asia, three in Sub-Saharan Africa and one in Latin America and the Caribbean. Two of these programmes were in India (Chin, 2005; Muralidharan and Sundararaman, 2013), one in Pakistan (Bau & Das, 2014), two in Kenya.[102] (Duflo, Dupas and Kremer, 2012 and 2007; Bold et al., 2013), one in Togo (Vegas and de Laat, 2003) and one in Mexico (Estrada, 2013). Three of the programmes targeted purely rural schools, while the others targeted a broader range of rural and urban settings.

Intervention

The interventions in this category are quite varied, employing a variety of mechanisms to achieve their goals. As a result, it is difficult to split them into clearly defined and mutually exclusive categories. Two studies that are a little easier to define both examined national policy changes introducing new teacher hiring procedures (and in one case also new procedures for determining promotion). Estrada (2013) examines an intervention whereby teachers would be selected according to their achievement in a competitive exam (Estrada, 2013). Ome (2012) evaluates a similar programme that made teacher hiring and promotion contingent on their passing a series of examinations. The other six studies are discussed in detail below, with Table 7.2c providing an overview of the different components involved in each intervention.

Four studies involved the recruitment of additional teachers for schools (Chin, A., 2005; Duflo, Dupas and Kremer, 2012; Bold *et al.,* 2013; Muralidharan and Sundararaman, 2013). Essentially, this meant providing budget for and recruiting a new teacher with the goal of reducing teacher-pupil ratios in treatment schools and thereby improving the quality of teacher-pupil interactions. In two of these cases, the additional teachers were contract teachers. In the other, schools were randomly assigned either an additional contract teacher or an additional civil-service teacher.

[102] These two studies both cover similar programmes based on the same concept of hiring additional contract teachers. However, they are counted as two separate interventions here as they were undertaken by different implementing agencies. One was located in western Kenya only while the other was implemented in all districts, while they were also differentiated by various other factors.

Five studies involved the employment of contract teachers with the aim of providing schools with lower cost but similarly qualified teachers (Vegas and de Laat, 2003; Duflo, Dupas and Kremer, 2012; Bold *et al.,* 2013; Muralidharan and Sundararaman, 2013; Bau & Das, 2014). As already stated, in three cases, these contract teachers were also additional (Duflo, Dupas, Kremer, 2012; Bold *et al.,* 2013; Muralidharan and Sundararaman, 2013).

Three studies involved contract teachers being recruited locally (Duflo, Dupas, Kremer, 2012; Bold *et al.,* 2013; Muralidharan and Sundararaman, 2013). Typically this meant that the selection process was undertaken by a local selection committee and local teachers were purposively targeted due to their potentially higher accountability to local communities. One study involved an effort to ensure that at least some of the new teachers were female, as part of an attempt to encourage girls to participate in schooling (Chin, 2005). In several cases, complementary components provided additional teaching resources or class space for new teachers (Chin, 2005; Duflo, Dupas and Kremer, 2012; Muralidharan and Sundararaman, 2013).

Comparisons

The included studies are based on a diverse range of interventions and the comparisons they use are also quite varied. Three interventions had more than one treatment arm (Vegas and de Laat, 2003; Duflo, Dupas and Kremer, 2007 and 2012; Bold *et al.,* 2013). Of these, two provided a direct comparison between treatment groups – Vegas and de Laat (2003) comparing contract teachers to civil-service teachers and Duflo, Dupas and Kremer (2007 and 2012) comparing additional contract teachers to additional civil-service teachers. Six studies provided a comparison between interventions and a no-intervention comparison group. These included Duflo, Dupas and Kremer's (2007 and 2012) study of the Extra Teacher Programme in Kenya which provided a comparison between each treatment arm and the no-intervention business as usual comparison group. Table 7.2a summarises the comparisons provided by each study.

Table 7.2 a: Comparisons provided by teacher hiring studies

Study ID	Comparison(s) provided
Bau & Das, 2014	Contract teacher Vs civil-service teacher
Bold *et al.,* 2013	- Additional contract teacher Vs no intervention - Additional contract teacher NGO implemented Vs no intervention - Additional contract teacher government implemented Vs no intervention
Chin, A., 2005	Additional civil-service teacher Vs no intervention
Duflo, Dupas, Kremer, 2012; Duflo, Dupas, Kremer, 2007	- Additional civil-service teacher Vs no intervention - Additional contract teacher Vs no intervention - Contract teacher Vs civil-service teacher
Estrada, 2013	Civil-service teacher recruited by competition Vs teachers recruited by committee
Muralidharan and Sundararaman, 2013	Additional contract teacher Vs no intervention
Ome, 2012	Teacher hiring and promotion by competition Vs no intervention
Vegas and de Laat, 2003	Civil-service teacher Vs contract teacher

Outcomes

The included studies reported on a variety of outcomes. A single study reported a measure of teacher attendance[103], teacher performance[104] and student attendance[105] (Duflo, Dupas and Kremer, 2012). Muralidharan and Sundararaman (2013) provided effect sizes relating to class size, pupil-teacher ratio and multigrade teaching. Two studies provided a measure of completion with Duflo, Dupas and Kremer (2007) measuring it as the number of students that had progressed to grade three after 24 months of the programme and (Chin, 2005) providing a measure, split by gender, and defined as the number of students completing primary school (Chin, 2005). Three studies provided measures relating to drop-out, with both Ome (2012) and Duflo, Dupas and Kremer (2007) measuring it as the incidence of student drop-out (contrasting this to students transferring from one school to another). Estrada (2013) provides a measure of retention measured as the number of students taking the end of year exam over the number of students registered at the beginning of the academic year[106].

Five studies reported a measure of learning outcomes. Four reported a composite maths and language arts test score (Duflo, Dupas & Kremer, 2012; Muralidharan and Sundararaman, 2013; Bold *et al.,* 2013; Bau and Das, 2014). Five reported separate language arts scores (Duflo, Dupas and Kremer, 2012; Ome, 2012; Estrada, 2013; Muralidharan and Sundararaman, 2013; Bau and Das, 2014), while six reported separate

[103] There were multiple observations per teacher with unannounced school visits made by enumerators on a quarterly basis and teachers' presence in school was recorded.
[104] There were multiple observations per teacher with unannounced school visits made by enumerators on a quarterly basis to observe whether teachers were observed in class actively teaching.
[105] There were multiple observations per pupil with unannounced school visits were made by enumerators on a quarterly basis and students' presence in school was recorded.
[106] In this case exam results did not determine student completion of the school year.

maths test scores (Vegas and de Laat, 2003; Ome, 2012; Duflo, Dupas and Kremer, 2012; Estrada, 2013; Muralidharan and Sundararaman, 2013; Bau and Das, 2014).

The follow-up period between the start of the intervention and final data collection varied quite widely between studies. To a large extent, this was because some interventions were field trials, while several were the result of national policy changes. One study followed up after 16 months (Bold *et al.,* 2013). Both Muralidharan and Sundararaman, (2013) and Duflo, Dupas and Kremer (2007; 2012) measured outcomes after 12 and 24 months, though the latter also collected some outcomes data after 18 months. Information is limited for Vegas and de Laat (2003) though it appears the data used in the study was collected over a period of 12 months. Chin (2005), Ome (2012) and Bau and Das (2014) all evaluate national-level programmes and have a relatively long follow-up period. Ome has a follow-up period of 84 months, while endline data collection for Chin (2005) and Bau and Das (2014) was after approximately 120 months.

Study Design

The eight studies identified are based on a mix of experimental and quasi-experimental designs. Three were cluster- randomised control trials where the intervention was assigned at the school level (Duflo, Dupas and Kremer, 2007 and 2012; Bold *et al.,* 2013, Muralidharan and Sundararaman, 2013). Four studies employed a controlled before-after study designs (Vegas and de Laat, 2003; Chin, A., 2005; Ome, 2012 and Estrada, 2013). A final study used a fuzzy regression discontinuity design approach in a natural experiment context (Bau & Das, 2014).

Qualitative studies, process evaluations and project documents

We identified eight additional documents that present qualitative, process and project information for six of the included programmes. These documents were used to provide additional background information and to inform our qualitative synthesis of intervention and implementation features associated with interventions' relative success and failure. This is a relatively large number of documents, and probably resulted from the fact that our included interventions include relatively large field trials and government policy changes. Below, we describe the intervention components in more detail, before presenting the descriptive findings on process, implementation and context.

Table 7.2 b: Characteristics of included studies for teacher hiring

Study ID	Setting	Intervention summary	Included outcomes	Follow- Up*	Study design	Sample Size
Bau & Das, 2014	Pakistan, rural primary schools. Students were in grades 3, 4 and 5.	**Key feature: hiring contract teachers.** Following an unexpected budget crisis in 1998, the Pakistani province of Punjab moved almost entirely from hiring a majority of civil service teachers on permanent contracts to hiring teachers on temporary contracts. Under the Pakistan Contract Appointment Policy (CAP), the government first decreased teacher hiring and then replaced the hiring of permanent teachers with contract teachers.	Composite test scores; language arts test scores; maths test scores	Approx. follow-up 120 months	Natural experiment with fuzzy regression discontinuity design (RDD)	823 schools
Bold *et al.*, 2013	Kenyan primary schools in all eight Kenyan provinces. Students were in grades 2 and 3.	**Key feature: hiring additional contract teachers.** Under this pilot of the National Teacher Programme, additional contract teachers were assigned to schools, either under a government-run programme, or the coordination of World Vision Kenya. Head teachers were asked to use new teachers to split existing classes in target grades, maximising reduction in class size.	Composite test scores	16 months	Cluster-RCT	192 schools, approximately 15,000 students
Chin, A., 2005	India, rural and urban primary schools. Students were	**Key feature: hiring additional civil-service teachers.** Under Operation Blackboard, an extra teacher was provided to every one-teacher primary school in the country following the 1986 All-India Education Survey. These	Completion	12, 24, 96 and 120 months	Controlled before and after study design with	Approximately 84,900 students

	in grades 1-5 and aged 6-10	new teachers were to be permanent members of staff and result in at least one female teacher being present in the school. There was also a one-time grant for learning equipment and a requirement that primary schools have at least two rooms.			DID estimation	
Duflo, Dupas and Kremer, 2012; Duflo, Dupas and Kremer, 2007	Primary schools in three districts of Western Kenya. Students were in grades 1 and 2, had an average age of around eight, but ranged from age 5-14.	**Key feature: hiring additional contract and civil-service teachers.** Under the Extra Teacher Programme (ETP) in Kenya, schools were randomly assigned an additional civil-service teacher or contract teacher. Contract teachers were hired locally and required to have the same academic qualifications as regular civil-service teachers but were paid less and had fewer benefits. Schools reduced current grade one class sizes, providing a new class for the new teachers. Additional classroom facilities were provided across schools.	Teacher attendance; teacher performance; student attendance; completion; drop-out; composite test scores; language arts test scores; maths test scores	12, 18 and 24 months	Cluster-RCT	210 schools, approximately 21,000 students
Estrada, 2013	Mexican rural public secondary schools. Students were in grade 9.	**Key feature: competitive teacher recruitment.** The intervention introduced a standardised test that was used in preference to traditional committee-based recruitment. The teacher test measures cognitive skills, knowledge of the teaching subject and mastery of teaching methods. Candidates are ranked by state and teacher-type according to their exam results or, if states opt for it, a weighted average of the test score and other criteria (often undergraduate GPA). Teaching positions	Drop-out; language arts test scores; maths test scores	60 months	Controlled before and after study design with DID estimation	1,232 students

		and exam results were publicised by media and published on a dedicated webpage. Schools offered newly opened places to teachers based on their ranking. Civil-society organisations participated as monitors at various stages.				
Muralidharan & Sundararaman 2013	Indian rural, public primary schools. Students were aged 8 on average.	**Key feature: hiring additional contract teachers.** Under the Extra Contract Teacher (ECT) Intervention schools selected by a lottery were authorised to hire an additional contract teacher. They were expected to follow the same procedures and guidelines for hiring a contract teacher as they would normally do. This meant that a selection committee of three members, typically comprised of the head teacher, a member of the local elected body, and another teacher, chose new teachers. It also meant that local candidates were recruited. Additional classroom facilities were also provided across treatment schools.	Class size; pupil-teacher ratio; multi-grade teaching; composite test scores; language arts test scores; maths test scores	12 and 24 months	Cluster-RCT	200 schools, approximately 16,000 students
Ome, 2012	Colombia, public primary, secondary and high schools, students in grades 5, 9 and 11	**Key feature: competitive teacher hiring and promotion processes.** El Estatuto de Profesionalizacion Docente (EPD) or the 'Teacher Professionalisation Statute' introduced a series of examinations for teachers that they would have to pass in order to get hired or promoted.	Dropout; maths test scores; language arts test scores	84 months	Natural experiment with fixed effects regression	102,431 students (dropout) Approx. 525,700 students (test scores)

Vegas and de Laat, 2003	Togo primary schools. Students were in grade 5.	**Key feature: comparison of (non-additional) contract versus civil-service teachers.** Students were non-randomly assigned either a permanent civil-service teacher or a contract teacher in their classroom. Students were tested for mathematics achievement at the beginning of the school year and then at the end of the school year.	Maths test scores	Unclear. Nationwide programme, start date unclear. Data used in study collected over 12 months.	Controlled before and after study with maximum likelihood and Bayesian estimation methods	233 schools and 2,846 students

*** Follow-up was calculated as the time between the start of the intervention and data collection, in months.**

Table 7.2 c: Intervention design features of included studies for teacher hiring

Study ID	Key design feature	Additional teachers?	Local teacher recruitment?	Female teacher recruitment?	Contract teachers?	Innovative method of hiring or promoting teachers?	Other component?
Bau & Das, 2014	Hiring contract teachers	N/A	N/A	N/A	✓	N/A	N/A
Bold *et al.*, 2013	Hiring additional contract teachers	✓	✓	N/A	✓	Centralised versus local hiring and payment of teachers	N/A
Chin, 2005	Hiring additional civil-service teachers	✓	N/A	✓	N/A	N/A	One-time grant for learning equipment provided
Duflo, Dupas, Kremer, 2012; Duflo, Dupas, Kremer, 2007	Hiring additional contract and civil-service teachers	✓	✓	N/A	✓	Civil-service teachers hired by Ministry of Education. Contract teachers hired by Parent-Teacher Associations.	Additional classroom facilities provided across schools
Estrada, 2013	Competitive teacher recruitment	N/A	N/A	N/A	N/A	Teachers recruited following national competition	Civil-society organisations participated as monitors

Study ID	Key design feature	Additional teachers?	Local teacher recruitment?	Female teacher recruitment?	Contract teachers?	Innovative method of hiring or promoting teachers?	Other component?
Muralidharan & Sundararaman, 2013	Hiring additional contract teachers	✓	✓	N/A	✓	N/A	Additional classroom facilities provided across schools
Ome, 2012	Performance based recruitment and promotion	N/A	N/A	N/A	N/A	Teacher recruitment and promotion decisions based on exam performance	N/A
Vegas and de Laat, 2003	Comparison of (non-additional) contract versus civil-service teachers.	N/A	N/A	N/A	✓	N/A	N/A

7.2.3 Synthesis of findings

The results of our synthesis are presented in two sections. First, we present the meta-analysis findings relating to the effects of *teacher hiring* on the review's primary and secondary outcomes of interest and explore results according to different population sub-groups. Second, we provide a discussion of the overall findings, with reference to the qualitative synthesis of intervention and implementation features associated with relative success and failure in improving educational outcomes.

Effects of teacher hiring

This section reports the results of the meta-analysis of the effects of *teacher hiring*, addressing question 1a. We structure the presentation of results according to the theory of change, starting with intermediate/ secondary outcomes (teacher attendance), followed by education access outcomes (attendance, drop-out) and final outcomes (completion, learning outcomes: composite test scores, language arts test scores, maths test scores).

Eight studies provided outcomes data. None of the studies reported on all outcomes, though Duflo, Dupas and Kremer (2007 and 2012) provided data for most outcomes. The number of effect sizes per outcome is just one in several cases, but as many as three for completion and composite test scores. As the studies included in are fairly diverse in terms of the types of teacher hiring interventions that they report on and the comparisons they make, we only include studies comparing an *additional contract teacher* with a no-intervention comparison group in the pooled meta-analysis and forest plots.

Outcomes data for studies reporting on all other interventions and comparisons are reported narratively in the appropriate section. All effect sizes are expressed as standardised mean difference (SMD), interpreted as the magnitude of the number of standard deviation changes in the outcome for the intervention group as compared to students in comparison schools. SMD scores are interpreted as the number of standard deviation changes in the outcome. All pooled meta-analyses employ random effects models.

Not all effect estimates from a given study are included in each meta-analysis. For example, where a single study reported estimates at multiple follow-up periods, only one of those is included. Where studies provided outcomes data at more than follow-up, we always tried to combine the most similar follow-up period from different studies in the meta-analysis. For some meta-analyses, we only have data from a range of different follow-up periods and we include these in the meta-analysis and conduct sensitivity analysis. The two studies providing outcomes data for the meta-analysis of completion outcomes had quite heterogeneous follow-up periods with one being 24 months (Duflo, Dupas and Kremer, 2012) and the other being 120 months (Chin, 2005). For composite test scores, the studies included in the meta-analysis reported follow-ups ranging from sixteen to twenty-four months, while for maths and language arts the two included studies had follow-ups after eighteen (Muralidharan and Sundararaman, 2013) and twenty-four months (Duflo, Dupas and Kremer, 2012).

Teacher attendance and performance

Only a single study (Duflo, Dupas and Kremer, 2007) provided an effect size relating to teacher attendance. The study compared the effect of an *additional contract teacher* to a no intervention comparison. The result suggest no difference between additional civil-service teachers' attendance and that of comparison school (SMD = -0.02, 95% CI [-0.07, 0.03]). The results are the same for teacher performance (SMD = -0.02, 95% CI [-0.07, 0.03]).

Class size, pupil-teacher ratio and multigrade teaching

A single study, Muralidharan and Sundararaman (2013), provided two effect sizes each for class size, pupil-teacher-ratio and multi-grade teaching.[107]. For each outcome, there were effect sizes after one year and two years of the intervention. The study compared an *additional contract teacher* treatment group to a no-intervention comparison group. For class size, the year one effect was -0.44, 95% CI [0.64, 0.25] and at the end of year two it was -0.55, 95% CI [-0.75, -0.35]. For pupil-teacher ratio, after one year the effect was -0.62, 95% CI [-0.82, -0.42] and at the end of year two it was -0.78, 95% CI [-0.98, -0.57]. For multi-grade teaching, the effect after one year was -0.62 [-0.83, -0.42] and after two years was -0.36, 95% CI [-0.56, -0.17]. All six of these measures were statistically significant. This suggest the Indian Extra Contract Teacher Intervention assessed by Muralidharan and Sundararaman (2013) was successful in its intermediate goal of reducing class sizes, pupil-teacher ratios and the incidence of multi-grade teaching.

Student attendance

A single study (Duflo, Dupas and Kremer, 2007) assessed the effect of teacher hiring on student attendance. The study compared the effect of an *additional contract teacher* to a no intervention comparison and the effect is not much different from zero (SMD = 0.02, 95% CI [0.00, 0.06]).

Completion

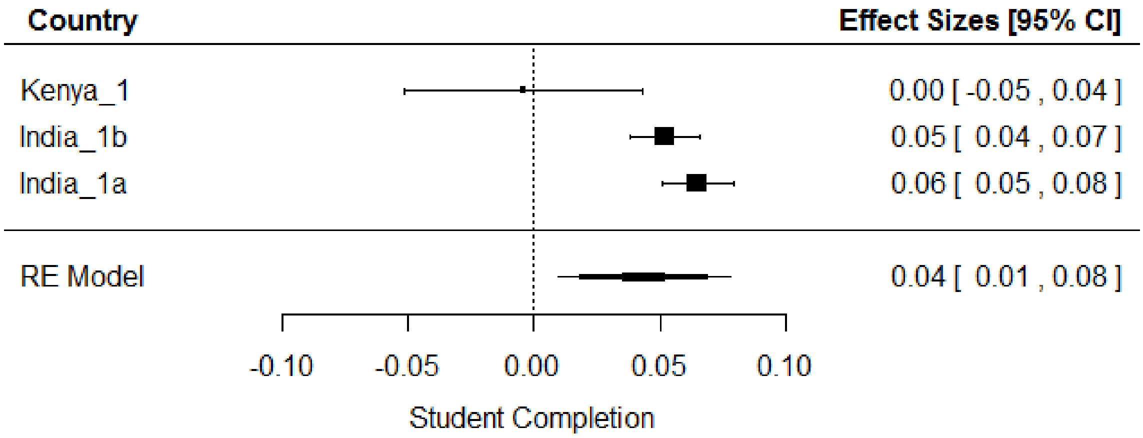
The following meta-analysis results and forest plot relates to the effect of *hiring an additional contract teacher compared to a no-intervention comparison group*, calculated under a random effects model. Other teacher hiring interventions and comparisons are then discussed narratively.

The overall average effect of hiring contract teachers on completion is positive but small in magnitude (SMD = 0.04, 95% CI [0.01, 0.08]). The assessment of homogeneity suggests a large amount of between-studies variability (I^2 = 89.08%, τ^2 = 0.0007, Q(df=2)= 8.1503, p = 0.0170). This is also apparent when inspecting the forest plot in figure 7.2a**.**

The effect sizes range from 0.00, 95% CI [-0.05, 0.04] in Kenya (Duflo, Dupas and Kremer, 2012), to 0.06, 95% CI [0.05, 0.08] in India (Chin, 2005). Note that the two Indian effect sizes represent girls (India_1a) and boys (India_1b respectively), with girls performing relatively better (Chin, 2005). The average effect is slightly sensitive to the inclusion of the study from Kenya, which when removed increases the positive average effect to 0.06, 95% CI [0.05, 0.07] (see Appendix H for results of all sensitivity analyses).

[107] 'Multi-grade teaching' is the incidence of teachers teaching multiple grades of children simultaneously instead of being able to focus on a single grade.

Figure 7.2 a: Completion[108]



Duflo, Dupas and Kremer's (2012) study in Kenya reports effects on student completion for two different comparisons. One for the hiring of an *additional contract teacher* compared to a no-intervention comparison group (included in the forest plot) and one for an *additional civil-service* teacher compared to a no-intervention comparison group (not included in the forest plot). There is no effect for an *additional contract teacher* SMD = 0.00, 95% CI [-0.05, 0.04]), whereas the effect of an additional *civil-service teacher* is positive and statistically significant *(*SMD = 0.09, 95% CI [0.05, 0.14]).

Drop-out

Four studies provided effect sizes for drop-out, though we chose not to meta-analyse them as they are drawn from quite different types of intervention that also imply different types of comparisons.

Duflo, Dupas and Kremer (2007) provide two estimates, both evaluating the effect of a teacher hiring intervention on drop-out. The first assess the effect of hiring an *additional contract teacher*, whereas the second provides the effect of an *additional civil-service* teacher. Both the effects are small, but with opposite signs (SMD = -0.04, 95% CI [-0.09, 0.01]; SMD = 0.03, 95% CI [-0.01, 0.08]).

Estrada (2013) estimates the effect on student retention of *civil-service teachers recruited by competition versus civil-service teachers recruited by committee*. They find a positive effect for competitive recruitment on student retention, though again, the effect is very small (SMD = 0.02, 95% CI [-0.07, 0.03]). Ome (2012) estimates the effect of *teacher recruitment and promotion based on performance in exams* on drop-out from primary school, secondary school and high school respectively. The effects are relatively small in magnitude, although these appears to have been an improvement in drop-out rates among secondary school students (Primary: SMD= -0.02, 95% CI [-0.06, 0.01]; Secondary: SMD -0.08, 95% CI [-0.11, -0.04]; High school: SMD= -0.04, 95% CI [-0.08, 0.00]).

[108] India_1a: Chin (2005): girls only
India_1b: Chin (2005): boys only
Kenya_1: Duflo, Dupas and Kremer (2012): full sample

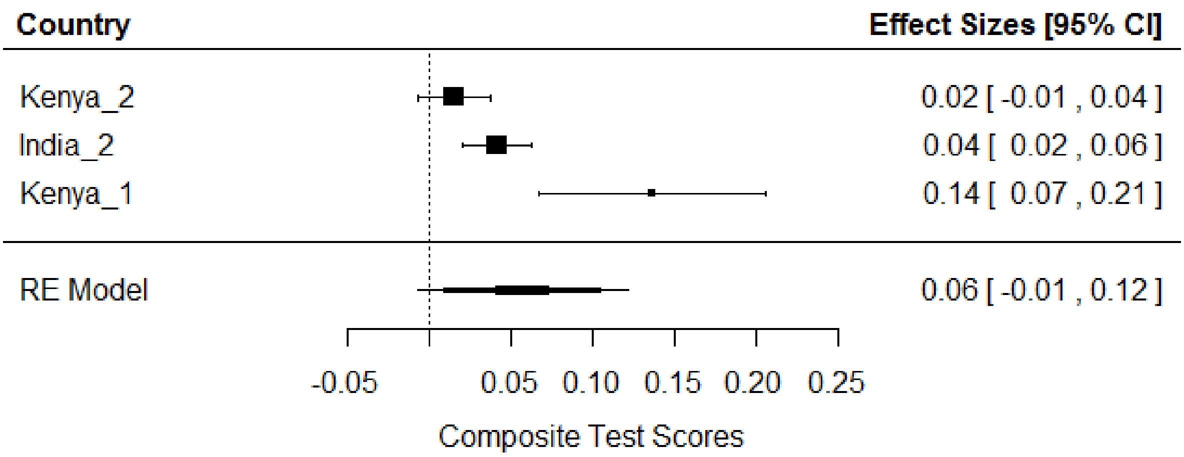
Composite scores

The overall average effect of hiring an *additional contract teacher* as compared to a no-intervention comparison group on learning outcomes measured by a composite score is 0.06, 95% CI [-0.01, 0.12]. The homogeneity tests (I^2 = 92.79%, τ^2 = 0.0531, Q (df = 2) = 11.5772, p = 0.0031) indicate that the effects did not arise from the same population, which is supported by the forest plot in figure 7.2b. Effect sizes range from –0.02, 95% CI [-0.01, 0.04] in Kenya (Bold et al., 2013), to 0.14, 95% CI [0.14, 0.21] (Duflo, Dupas and Kremer, 2012). Sensitivity analysis indicates that removing the largest effect size for Kenya_1 reduced the pooled effect size slightly to 0.03, 95% CI [0.00, 0.06] (see Appendix H for results of all sensitivity analyses).

The effect size from Bold *et al.,* (2013) included in the forest plot captures the overall effect of the programme. However, Bold *et al.* also provide sub-group analysis exploring how *government-implemented* and *NGO-implemented* programme arms performed relative to one-another. The effect size for the additional contract-teacher programme implemented by the government compared to a no-intervention comparison group was -0.03, 95% CI [-0.06, 0.01] whereas when implemented by an NGO the effect was slightly larger 0.03, 95% CI [-0.06, 0.01].

Muralidharan and Sundararaman (2013, India_2) also measure effects after the shorter period of one year, finding a slightly smaller effect (SMD = 0.03, 95% CI [0.01, 0.06]).

Figure 7.2 b: Composite Test Scores[109]



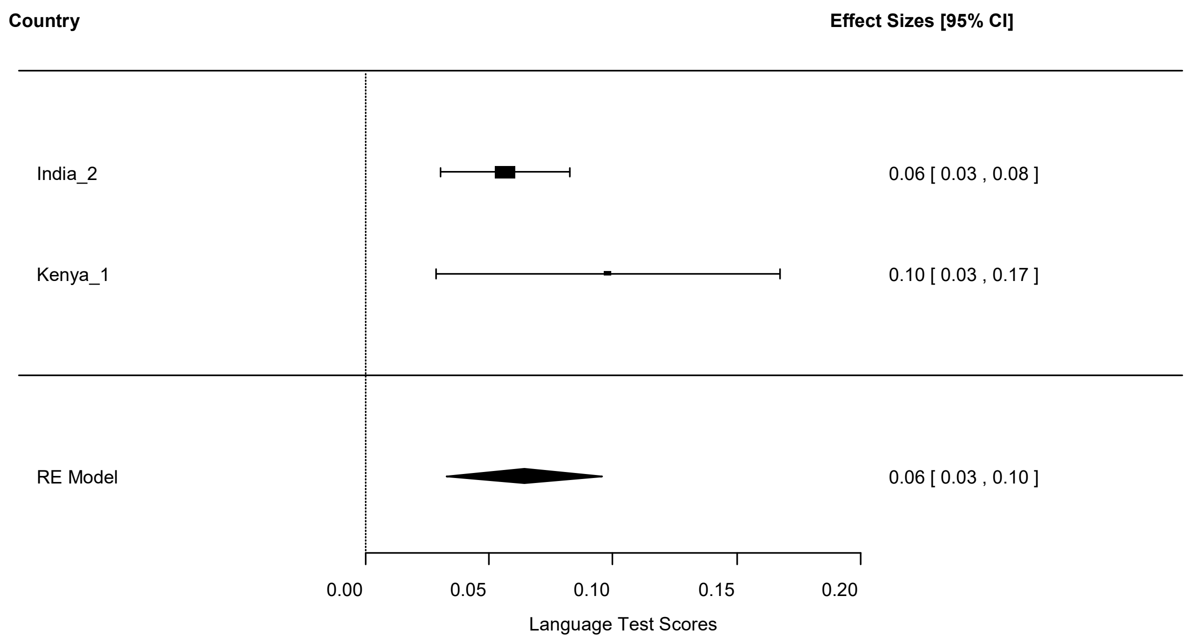
Duflo, Dupas and Kremer (2012) provided a comparison of hiring an *additional civil-service teacher*, as compared to a no-intervention comparison group (0.05, 95% CI [-0.02, 0.12]). This can be indirectly compared to the effect size reported in the same study for an additional contract-teacher (and included in the forest plot) compared to a no-intervention comparison group (SMD = 0.14, 95% CI [0.14, 0.21]). The results suggest that the effects of an additional contract teacher may be larger than the effect of an additional civil-service teacher, though it is not clear whether this difference is statistically or substantively significant. Bau & Das (2014) compare (non-additional) contract teachers directly to civil-service teachers, and similarly find that students of contract teachers performed better on their composite test scores (SMD = 0.70, 95% CI [0.64, 0.78]).

[109] India_2: Muralidharan and Sundararaman (2013); Kenya_1: Duflo, Dupas and Kremer (2012); Kenya_2: Bold *et al.* (2013)

Language arts scores

The overall average effect of *hiring an additional contract teacher* on language arts test scores is 0.06 (95% CI [0.03, 0.10]). The homogeneity tests (I^2 = 16.47%, τ^2 = 0.0001, Q (df = 1) = 1.1972, p = 0.2739) indicate limited heterogeneity. The effect size included in the meta-analysis and represented in the figure 7.2c forest plot from Muralidharan and Sundararaman (2013) was after two years of the intervention (India_2). They also provide an effect size after the shorter period of one year, which is smaller (SMD = 0.04, 95% CI [0.00, 0.07]).

Figure 7.2 c: Language Arts Test Scores[110]



Seven other effect sizes for language arts test scores were not included in the pooled meta-analysis or forest plot as they related to different interventions and/or made different comparisons. Duflo, Dupas and Kremer (2012) measure the effect of hiring an additional civil-service teacher (0.04, 95% CI [-0.03, 0.11]). This can be indirectly compared to the effect size reported in the same study (and included in the forest plot) for an additional contract-teacher compared (SMD=0.10, 95% CI [0.03, 0.17]). The results suggest the effect of an additional contract teacher may be larger than the effect of an additional civil-service teacher, though it is not clear whether this difference is statistically or substantively significant.

Bau & Das (2014) compare (non-additional) contract teachers directly to civil-service teachers, and find that students of contract teachers performed better on their language arts test scores in both Urdu (SMD=0.69, 95% CI [0.61, 0.77]) and English (SMD=0.85, 95% CI [0.77, 0.93]). Estrada (2013) compares civil-service teachers recruited by competition versus civil-service teachers recruited by committee, finding little difference (SMD=-0.03, 95% CI [-0.11, 0.05]). Ome (2012) estimates the effect of teacher recruitment and promotion based on their performance in exams versus a no intervention comparison group and provides three separate estimates for students' language arts test scores, split primary, secondary and high

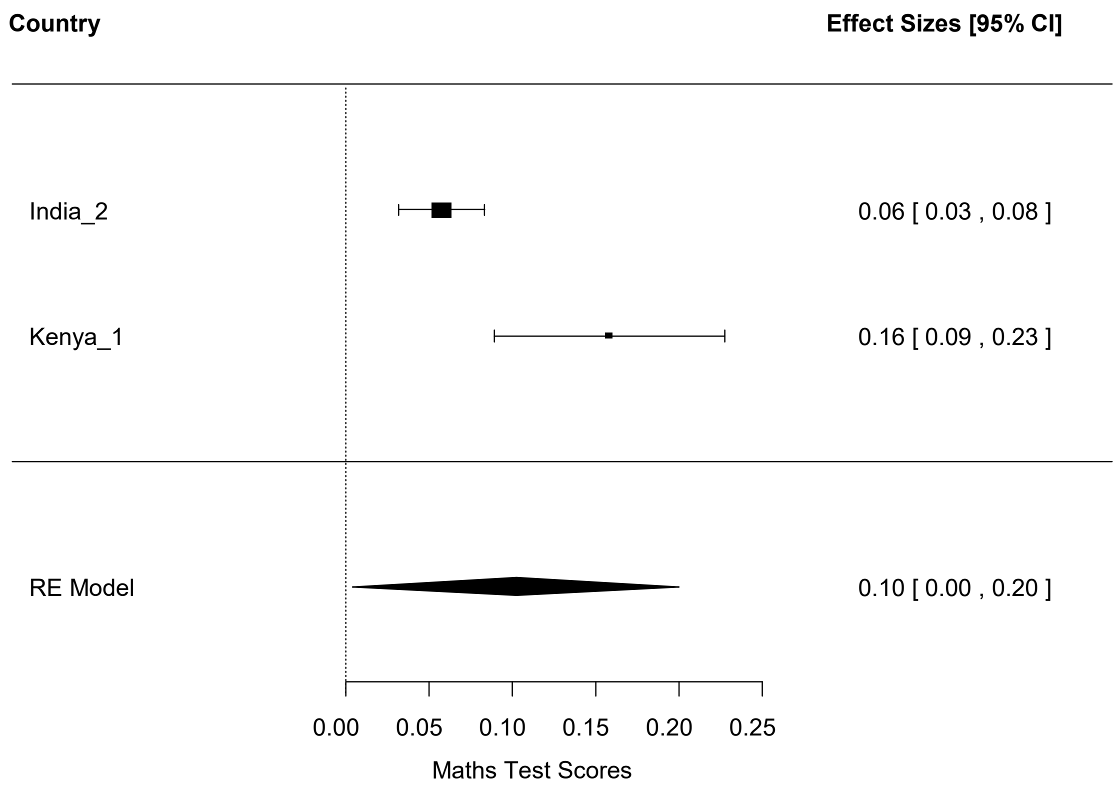
[110] India_2: Muralidharan and Sundararaman (2013), Kenya_1 Duflo, Dupas and Kremer, (2012)

school. There appears to be a substantial effect on secondary school students, but not primary school and high school (Primary: SMD = 0.02, 95% CI [-0.01, 0.05]; Secondary school: SMD = 0.10, 95% CI [0.06, 0.14]; High school students: SMD = 0.00, 95% CI [-0.03, 0.02]).

Maths test scores

The overall average effect of *hiring an additional contract teacher* on maths test scores is positive (SMD = 0.10 95% CI [0.00, 0.20]) but only marginally statistically significant. The homogeneity tests (I^2 = 85.90%, τ^2 = 0.0043, Q (df = 1) = 7.0922, p = 0.0077) again indicate substantial heterogeneity between the included studies, which means the average results of the meta-analysis should be interpreted cautiously. Only two studies are included, both of which find positive and statistically significant effects, though their confidence intervals do not overlap.

Figure 7.2 d: Maths Test Scores[111]



The effect size included in the meta-analysis and represented in the figure 7.2d forest plot from Muralidharan and Sundararaman (2013) was after two years of the intervention (India_2). They also provide an effect size after the shorter period of one year, which is slightly smaller (0.05, 95% CI [0.03, 0.08]).

Eight other effect sizes for maths test scores were not included in the meta-analysis as they related to different interventions or made different comparisons. Duflo, Dupas and Kremer (2012) also assess the effect of hiring an *additional civil-service teacher* (SMD=0.06, 95% CI [-0.01, 0.13]). Comparing this to the effect of an *additional contract-teacher* (as included in

[111] India_2: Muralidharan and Sundararaman (2013)
Kenya_1 Duflo, Dupas and Kremer, (2012)

the meta-analysis: SMD=0.16, 95% CI [0.09, 0.23]) suggest an additional contract teacher may be more beneficial than the effect of an additional civil-service teacher.

Bau & Das (2014) compare (non-additional) contract teachers directly to civil-service teachers, and find that students of contract teachers performed better on maths (SMD=0.62, 95% CI [0.53, 0.72]). Vegas and de Laat (2003) also compare (non-additional) civil-service teachers directly to contract teachers. Unlike Bau & Das (2014), they find that students of civil-service teachers performed better on maths (SMD = 0.27, 95% CI [0.13, 0.41]).

Estrada (2013) compares civil-service teachers recruited by competition versus civil-service teachers recruited by committee, finding little difference (SMD = -0.01, 95% CI [-0.08, 0.05]). Ome (2012), evaluating the effect of a programme that compared *recruitment and promotion of teachers based on teacher exam scores* versus the status quo, again reports separate estimates for primary, secondary and high schools. As for composite scores and maths scores they observe a positive effect on secondary school students (Primary: SMD = 0.03, 95% CI [0.01, 0.06]); Secondary: SMD = 0.10, 95% CI [0.06, 0.13]; High school: SMD = 0.00, 95% CI [-0.02, 0.02].

7.2.4 Summary of findings and discussion

We identified eight studies of teacher hiring programmes across five different countries in South Asia, Latin America and Sub-Saharan Africa. We were able to examine effects on completion, maths, language arts and composite test scores using meta-analysis. We only included studies comparing an additional contract teacher to a no-intervention comparison group in the meta-analyses. All other types of teacher hiring intervention or comparisons were discussed narratively.

There is a large amount of between study variability for most outcomes. Below we discuss the results, incorporating any relevant findings from the qualitative synthesis (reported in full in appendix J) that may help explain this heterogeneity.

One finding that emerged from the qualitative evidence is that teacher hiring interventions are not always easy to implement due to the fact they can threaten existing jobs or provoke opposition because they may mean lower pay, fewer privileges and less job security. In Kenya, the government ended a contract teacher intervention early and acquiesced to union demands to absorb contract teachers into civil-service employment (Bold *et al.,* 2013). The authors conclude that contract teacher hiring may be effective on a small-scale, but a large cohort of teachers employed at wages far below levels for civil-service peers can be difficult to implement, particularly where teachers are unionised and politicised. In Mexico, recruitment by competitive exam rather than traditional recruitment by committee was opposed by teaching unions (Estrada, 2013).

Additional contract teachers compared to a no-intervention comparison group

The overall average effects range from 0.04, 95% CI [0.01, 0.08] for completion, to 0.06, 95% CI (-0.01, 0.12) for composite test scores, 0.06, 95% CI (0.03, 0.10) for language arts test scores and 0.10, 95% CI (0.00, 0.20) for maths test scores. These results should be interpreted cautiously as they are based on just a few studies and the average scores for composite test scores and maths test scores are not statistically significant. Nevertheless, our analysis provides tentative evidence that hiring additional contract teacher interventions can have beneficial effects on student outcomes.

Looking at composite test scores, Muralidharan and Sundararaman (2013) and Duflo, Dupas and Kremer (2012) both report a positive and statistically significant effect on test scores, while Bold et al. (2013) report a relatively smaller positive, but statistically insignificant effect. Intermediate outcomes and wider process and implementation factors provide some possible reasons for this difference in performance. The Indian Extra Contract Teacher Intervention assessed by Muralidharan and Sundararaman (2013) was successful in its intermediate goal of reducing class sizes, pupil-teacher ratios and the incidence of multi-grade teaching. Duflo, Dupas and Kremer (2012) also find qualitative evidence that additional contract teachers were able to reduce class sizes. However, Bold et al. (2013) find that reallocation of teachers undermined the potentially beneficial effect of additional teachers. Although teachers were typically placed in the correct grade, they were often asked to cover other grades while some teachers were reallocated within schools. As a result, the actual reduction in class sizes was quite small. Both Duflo, Dupas and Kremer (2012) and Muralidharan and Sundararaman (2013) also reported that schools were able to fill the vacancies created by programmes relatively quickly. On the other hand, Bold *et al.* (2013) report that although vacancies were created and funding provided for additional teachers, posts were not always filled (with some remaining open and others initially being filled but teachers then leaving without being replaced).

In addition to the factors outlined above, Bold *et al.* (2013) also report that salary delays occurred and were significantly correlated with relatively lower test score improvement, while recruitment mechanisms were prone to elite capture and many teachers hired for the programme had a connection to existing teachers.

Comparing contract teachers and civil-service teachers

Three studies compared contract teachers to civil service teachers, either directly or indirectly. Of these, two studies reported that overall performance of contract teachers (in terms of their students' test scores) was superior to that of civil-service peers. Bau and Das (2014) and Duflo, Dupas and Kremer (2012) find that the students of contract teachers perform better than students of civil service teachers in composite test scores, maths and language arts. However, there was also one study that found that students of contract teachers performed worse than those of civil-servant teachers on maths tests (Vegas and de Laat, 2003)[112].

One possible explanation may lie in the way that contract teachers perceive their pay and conditions relative to that of other teachers in a given context. Contract teachers in all three interventions received less pay and fewer benefits than their civil-service peers. However, Duflo, Dupas and Kremer (2012) report that despite these relatively poorer conditions, unemployed teachers still actively seek contract teaching positions. Bau and Das (2014) further report that contract teachers were still better paid than teachers working in low-cost private schools and this may have been a positive motivating factor. Vegas and de Laat (2003) comment that contract teachers were worse paid and more likely to report that they received their pay on a very irregular basis than civil-service teachers. They test the theory that this would have resulted in comparatively less qualified candidates filling the contract teacher vacancies and conclude that this was the case. The theory behind contract-teacher interventions predicts that teachers on contracts can provide a more economical, equally qualified and better motivated workforce. However, there is also the danger that their poorer employment conditions may result in a 'disgruntled worker effect' that negatively effects performance and the quality of candidates. The evidence presented here suggest both these

[112] Maths is the only learning outcome reported on by this study.

scenarios may occur in different contexts and that contract teachers' relative pay and conditions are important determinants of their job satisfaction and potentially also their commitment and performance.

Competitive teacher recruitment

Two studies examined competitive teacher recruitment interventions and reported on drop-out and maths and language arts test scores (Estrada, 2013; Ome, 2012). Estrada's evaluation of the introduction of standardised tests in Mexico, found a positive effect on student retention, but small reductions in maths and language arts test scores, though none of these were statistically significant. Ome's evaluation of the 'Teacher Professionalisation Statute' in Colombia was more conclusive, finding a small and statistically significant reduction of student drop-out for secondary school students. Estimates for primary and high school students also indicated small reductions in drop-out rates, though not statistically significant. Ome (2012) found evidence to suggest improvements among secondary school students, but not primary school and high school. There is no clear qualitative evidence that can explain the differences in findings between these two studies.

Table 7.2 d: Key descriptive findings: Process and implementation

Descriptive finding	Citation and context
Teacher recruitment and allocation	
Some teacher hiring programmes were able to fill the vacancies created by programmes relatively quickly.	Muralidharan & Sundararaman, 2013 (India, ECT) Duflo, Dupas, Kremer, 2007 and 2012 (Kenya, ETP)
Though vacancies were created and funding provided for teachers, posts were not always filled with some remaining open and others initially being filled but then teachers leaving without being replaced.	Bold *et al.,* 2013 (Kenya)
Teachers were not always allocated to the right schools. Operation Blackboard was meant to target additional civil-service teachers to one-teacher schools. However, misallocation meant that only 1 in 4 teachers appointed under OB were actually sent to a one-teacher school.	Chin, 2005 (China, OB)
Additional teachers enabled schools to reduce class sizes substantially.	Duflo, Dupas, Kremer, 2007 and 2012 (Kenya, ETP)
Reallocation of teachers within schools can undermine the desired effect of additional teachers. In Kenya, reductions in class sizes were very small - although teachers were typically placed in the correct grade, they were also asked to cover other grades while other teachers were reallocated within schools.	Bold *et al.,* 2013 (Kenya)
The recruitment process for additional civil-service teachers bypassed existing waiting lists and resulted in delays, court cases and rushed recruitment of teachers that had previously been set aside because their credentials did not meet state standards.	Dyer, 2012 (China, OB)
Additional materials and classroom equipment	
Additional materials and classroom equipment was not delivered to schools or used.	Chin, 2005 (China, OB)
Civil-service teachers were able to 'pull rank' and obtain greater access to additional materials and physical classroom infrastructure compared to their contract teacher peers.	Duflo, Dupas, Kremer, 2007 and 2012 (Kenya, ETP)
Programme implementation	

Descriptive finding	Citation and context
Implementation by an NGO-led version the National Teacher Programme pilot, including monitoring and evaluation, vacancy filling success rate and teacher attendance was better than a government-led version, potentially accounting for the better performance of NGO-led schools on primary outcomes such as test scores.	Bold *et al.,* 2013 (Kenya)
Salary delays occurred and led to poorer performance. They were more severe in government-led schools than NGO-implemented schools. The salary delays were significantly and negatively correlated with test score improvement.	Bold *et al.,* 2013 (Kenya)
Teacher hiring processes can be prone to local capture with selected teachers often those with a connection to existing teachers. Reported nepotism was higher for the NGO-led version of the National Teacher Programme than the government-led version.	Bold *et al.,* 2013 (Kenya)
There was little monitoring of implementation and it would not have been difficult for states to use the funds in unintended ways.	Chin, 2005 (China, OB)
The introduction of new teacher recruitment processes disrupted existing ones and initially led to delays in recruitment.	Estrada, 2013 (Mexico)
Less than a third of prospective teachers taking Mexico’s new standardised teacher exam passed.	Estrada, 2013 (Mexico)
Contract versus civil-service teachers	
Contract teachers were worse paid and more likely to report that they received their pay on a very irregular basis compared to civil-service teachers. As a result, vacancies may have attracted less qualified candidates.	Vegas and de Laat, 2003 (Togo)
Contract teachers were treated worse than civil-service teachers and more likely to feel discriminated against due to lower pay, fewer non-monetised benefits and fewer desirable responsibilities.	Bau & Das, 2014 (Pakistan, CAP)
Contract teachers still received better pay than private school teachers and this may have been a positive motivating factor.	Bau & Das, 2014 (Pakistan, CAP)

Descriptive finding	Citation and context
In Kenya, despite low pay and lack of job security, unemployed teachers still actively seek after contract positions.	Duflo, Dupas, Kremer, 2007 and 2012 (Kenya, ETP)

Table 7.2 e: Key descriptive findings: Context

Descriptive finding	Citation and context
Context	
Unions have opposed new teacher hiring approaches, either because they threaten existing jobs or because they will mean lower pay, fewer privileges and less job security.	Bold *et al.,* 2013 (Kenya) De Pascual Pola, 2009 (Mexico)
Limited involvement of local stakeholders may result in low buy-in for an intervention	Dyer, 1999 (China, OB)

7.3 Diagnostic Feedback

Diagnostic feedback interventions use 'low-stakes' student tests to provide teachers with information on student achievement that will enable them to target their efforts in the classroom more effectively. 'Low stakes' tests have been described as 'assessments for learning' and can be compared to 'high stakes tests', which can be described as 'tests of learning'. Low stakes tests allow teachers to monitor student progress and tailor their teaching approach to promote learning, without subjecting students to the stress of high stakes exams.

7.3.1 Description of included studies

We included two studies that evaluated the effect of a programme that provided diagnostic feedback to teachers. Table 7.3a provides an overview of the characteristics of the studies, which are described in more detail in the following section.

Populations

Both included studies evaluated programmes in India. The Continuous and Comprehensive Evaluation (CCE) programme was implemented in Haryana state in northern India (Duflo *et al.,* 2015). It assessed the effect of the CCE on students in grades 1-8 of rural public primary and upper primary (equivalent to secondary) schools. The second programme provided diagnostic feedback to teachers and was implemented in Andhra Pradesh in southern India (Muralidharan and Sundararamen, 2010). It evaluated the effect of the programme on rural public primary schools in grades 1-5.

Interventions

The CCE replaced 'high-stakes' end-of-year exams with frequent 'low-stakes' assessments of student achievement (Duflo *et al.,* 2015). The intention was to provide teachers and students with detailed and frequent feedback on performance with the aim of allowing teachers to tailor their teaching to the learning needs of individual students. Teachers were trained on how to conduct regular evaluations of students and maintain student progress records. In order to implement the programme, schools were also provided with manuals, evaluation sheets, and report cards.

The diagnostic feedback intervention evaluated by Muralidaran and Sundararaman (2010) also involved low-stakes diagnostic tests and feedback to teachers. Tests were independently administered at the start of the school year and teachers were given detailed diagnostic feedback on student performance together with an explanation of how to interpret and use the performance reports and benchmarks. Teachers were informed that their teaching would be monitored over the course of the year and an end-of-year test administered to students, though no individually attributable information would be made public and there would be no negative consequences for poor performing teachers.

Comparisons

Both of the included studies compared the effect of the intervention to business as usual (no intervention). However, Duflo *et al.* (2015) also evaluate whether a combination of the CCE programme with another intervention called the Learning Enhancement Programme (LEP) was any more or less effective than CCE alone. In schools receiving the two programmes together, the on-going assessment and feedback components of the CCE were combined with LEP components which involved setting aside a portion of the school day to teach students according to their ability level, regardless of age or grade. The LEP is included elsewhere in this report in its own right as part of the chapter on 'school participation by ability'. In the synthesis section later in this chapter we briefly examine the difference between the effectiveness of the CCE alone and the CCE/LEP combination.

Outcomes

Duflo et al. (2015) report on maths and language arts outcomes, providing separate analysis for upper primary students (equivalent to secondary school students) and primary students. They also undertake subgroup analysis by gender for primary school students only. Muralidaran and Sundararaman (2010) report maths, language arts and composite test scores as well as a measure of teacher performance. The follow-up periods for both of our included studies are relatively short with Duflo *et al.* (2015) following up approximately 15 months after baseline and Muralidaran and Sundararaman (2010) following up after 12 months.

Study Design

Both included studies were cluster randomised controlled trials.

Table 7.3 a: Characteristics of diagnostic feedback programmes

Included study	Population	Intervention summary	Included outcomes	Follow up	Study design	Sample size
Muralidharan & Sandararaman, 2010	India, rural primary public schools in Andhra Pradesh. Students were in grades 1-5.	Low stakes tests and diagnostic feedback on student performance. Teachers were provided with performance reports, benchmarks and an explanation of how to interpret and use them to tailor teaching to students' needs. Teaching was monitored over the year and an end-of-year test administered to students. Teachers were told that no performance information would be made public and there would be no negative consequences for poor performing teachers.	Teacher performance; Separate maths and language arts test scores; a composite test score for incorporating both maths and language arts	12 months	Cluster RCT	400 Schools
Duflo *et al.*, 2015	India, rural public primary and upper primary schools in Haryana, Mahendragarh, & Kurukshetra. Students were in grades 1-8.	The Continuous and Comprehensive Evaluation (CCE) programme introduced frequent low-stakes tests and diagnostic feedback on student performance. Teachers received manuals, evaluation sheets and report cards and trained in how to conduct frequent student evaluations and tailor teaching to meet students' needs.	Separate maths & language arts test scores	Approx. 15 months	Cluster RCT	400 Schools, 12,663 students

7.3.2 Synthesis of findings

In the following section, we report on the effects of *diagnostic feedback* interventions on the review's primary and secondary outcomes of interest. There are only two studies included, so it was only possible to synthesise evidence through meta-analysis for two outcomes (language arts and maths test scores). In all other cases, we report results narratively. We structure the presentation of results according to the theory of change, starting with intermediate/ secondary outcomes (teacher performance), followed by final outcomes (learning outcomes: composite test scores, language arts test scores, maths test scores).

Effects of diagnostic feedback interventions on teacher performance and student composite, language arts and maths scores

Neither of the included studies reported on all outcomes. Both Duflo *et al.* (2015) and Muralidharan & Sandararaman (2010) provide data for separate language arts and maths test scores. Muralidharan & Sandararaman (2010) also provide outcomes for composite test scores and teacher performance. All effect sizes are expressed as standardised mean difference (SMD), interpreted as the magnitude of the number of standard deviation changes in the outcome for the intervention group as compared to students in comparison schools. SMD scores are interpreted as the number of standard deviation changes in the outcome. The meta-analyses combine similar follow-up periods, with Duflo *et al.* (2015) following up approximately 15 months after baseline and Muralidaran and Sundararaman (2010) following up after 12 months.

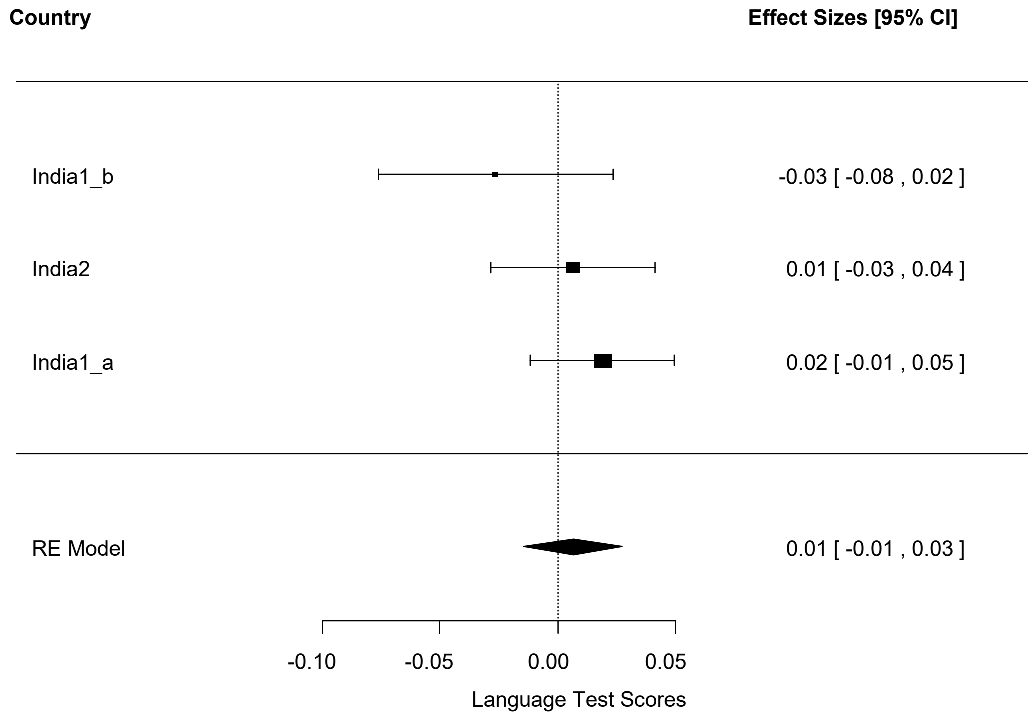
Teacher performance and composite scores

Muralidharan & Sandararaman (2010) find an effect of 0.18 (95% CI [0.15, 0.21]) on teacher performance but they find no effect on composite test scores (SMD = 0.00, 95% CI [-0.02, 0.02]).

Language arts scores

The overall average effect of diagnostic feedback interventions on language arts test scores is close to zero (SMD = 0.01, 95% CI [-0.01, 0.05]). The homogeneity tests ($I^2 = 0.04\%$, $\tau^2 = 0.0000$, Q (df = 2) = 2.3321, p = 0.3116) indicate minimal heterogeneity between included estimates. Sensitivity analysis indicates that removing any single study does not substantively change the results.

Figure 7.4 a: Language Arts Test Scores[113]



There were six other estimates of language arts test scores from Duflo *et al.*'s (2015) evaluation of the Continuous and Comprehensive Evaluation (CCE) programme that were not included in the meta-analysis. The effect size presented in Figure 7.4a was a synthetic effect size combining Hindi reading and written test scores. Duflo *et al.* (2015) also provide separate scores for each of these exams, split by gender. The only effect that appear substantively different from the main sample is the effect on girls' written Hindi test scores (SMD = 0.07, 95% CI [0.02, 0.12]), although the effect is still small in magnitude.

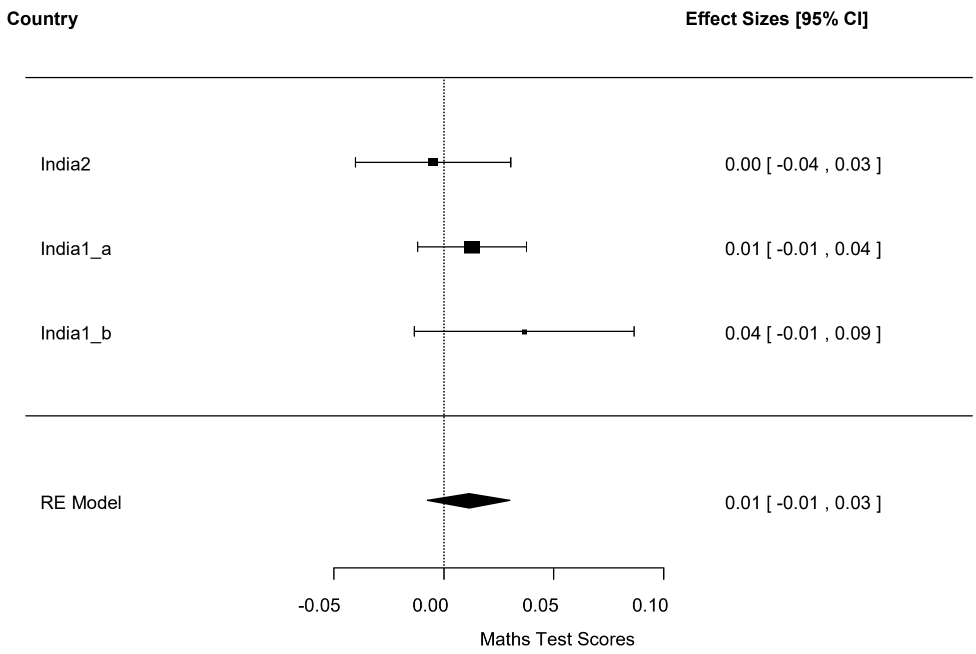
Duflo *et al.* (2015) also further evaluate whether a combination of the CCE programme with another intervention called the Learning Enhancement Programme (LEP) was any more or less effective than CCE alone. In schools receiving the two programmes together, the on-going assessment and feedback components of the CCE were combined with LEP components which involved setting aside a portion of the school day to teach students according to their ability level, regardless of age or grade. The results are not substantively different from CCE only.

Maths scores

The overall average effect of diagnostic feedback interventions on maths test scores is 0.01 (95% CI [-0.01, 0.03). The homogeneity tests (I^2 = 0.06%, τ^2 = 0.0003, Q (df = 2) = 1.8054, p = 0.4055) indicate minimal heterogeneity between included estimates. As expected, sensitivity analysis indicates that the average effect is not sensitive to the removal of any one estimate.

[113] India1_a: Duflo *et al.,* (2015) Hindi language, lower primary school
India1_b: Duflo *et al.,* (2015) Hindi language, upper primary school
India2: Muralidharan & Sandararaman (2010) Telugu language, primary schools

Figure 7.4 b: Maths Test Scores[114]



There were six other estimates of maths test scores from Duflo *et al.'s* (2015) evaluation of the Continuous and Comprehensive Evaluation (CCE) programme that were not included in the meta-analysis. The effect size presented in Figure 7.4b was a synthetic effect size combining ASER oral maths test scores with test scores for a written maths exam. Duflo *et al.* (2015) also provide separate scores for each of these exams, further subdivided by gender. As above the effects are not substantively different, although they are slightly larger in magnitude for girls' written maths test scores (SMD = 0.05, 95% CI [0.00, 0.10]). As before, Duflo *et al.* (2015) also evaluate whether a combination of the CCE programme with another intervention called the Learning Enhancement Programme (LEP) was any more or less effective than CCE alone. The results are not substantively different from CCE only.

7.3.3 Summary of findings and discussion

We identified two studies of diagnostic feedback programmes, both in India. We were able to examine effects on maths and language arts test scores using meta-analysis. All other outcomes were reported narratively. There is a very limited amount of between study variability for meta-analysed outcomes. This degree of homogeneity is unsurprising given the similarities in intervention design, context and populations in the two included studies. Overall effects from meta-analyses for language arts (SMD = 0.01, 95% CI [-0.01, 0.05]) and maths (SMD = 0.01, 95% CI [-0.01, 0.03) suggest there is no substantive effects of these programmes on learning outcomes.

There was a positive effect on teacher performance reported by a single study (Muralidharan and Sundararaman, 2010), but no evidence that this led to improved student composite, maths and language arts test scores. Muralidharan and Sundararaman (2010) did find some tentative evidence of positive effects on girls' written Hindi and Maths test scores, though this improvement was not observed for oral maths or language arts test scores and they

[114] India1_a: Duflo *et al.,* (2015), lower primary school
India1_b: Duflo *et al.,* (2015), upper primary school
India2: Muralidharan & Sandararaman (2010), primary schools

found no substantive effects on boys' test scores. The follow up period for both studies was relatively short (12 months for Muralidharan and Sundararaman, 2010; 15 months for Duflo *et al.,* 2015) and it is possible this may have been too short a period for student test scores to improve. The diagnostic feedback intervention was also combined with a complementary programme setting aside a portion of the school day to teach students according to their ability level, but the effects on test scores were not different for this programme (Muralidharan and Sundararaman, 2010).

7.4 Teacher training

7.4.1 Description of included study

We included a single study that reported on a teacher training programme. Elsewhere in this report, the chapter on pedagogical interventions brings together many interventions that include a teacher training component. However, in those cases training typically took the form of a short period of topic-specific learning designed to facilitate the introduction of new classroom sessions. In contrast, the study we report on here focused on teacher training as professional development designed primarily to build teachers' professional capabilities.

We included two papers from two papers that relate to a single study of the Learning to Read in a Healing Classroom (LRHC) programme (Halpin, Torrente & Aber, 2014; Wolf *et al.,* 2015). The programme was implemented in public primary schools in the South-eastern Katanga province of the Democratic Republic of Congo. LRHC targeted 346 teachers teaching grades one to six of primary schools. The majority of these teachers were male (71.7 %) and teachers were on average 37.4 years old and had 11.7 years of teaching experience.

The LRHC intervention consisted of two key components. The first was a teacher professional development support system made up of teacher trainings and Teacher Learning Circles (TLCs) that provided peer support and focused on effective teaching. The second component consisted of instructional guides with integrated social-emotional and literacy practices. The study used a cluster-RCT study design that compared the intervention group to a no-intervention comparison group made up of schools that received the intervention at a later date. The study followed up on outcomes around 12 months after the programme was implemented.

7.4.2 Findings

The authors reported the effect of the Learning to Read in a Healing Classroom (LRHC) teacher training programme on learning outcomes. There were two separate estimates for different maths tests and one language arts tests.

All the effects are positive, but relatively small in magnitude (Early Grade Maths Assessment (EGMA): SMD = 0.04, 95% CI [-0.02, 0.11]; Geometry test: SMD = 0.07, 95% CI [0.01, 0.14]); Language arts: SMD = 0.08, 95% CI [0.02, 0.15]). Overall, the findings for the single teacher-training study identified suggest this type of intervention may be beneficial as compared business as usual. However, the evidence is limited and the effects appear small.

8. System-level interventions

Systems interventions are programmes aiming to improve education through changes to the education system at either the community, local government and district/ state or national level. The interventions taking place at this level are primarily related to the management, governance and financing of education. We reviewed the evidence on three types of interventions, namely School-based management (SBM), Community Based Monitoring (CBM) and Public private partnerships (PPP). At the core of these initiatives is the decentralisation of decision making authority to local levels and greater involvement of communities in making decisions and monitoring service providers.

This chapter provides the findings of our synthesis of the 37 included studies evaluating the effect of SBM, CBM and PPP on learning outcomes and access to schooling. The chapter is organised by interventions. Each sub-section starts with a description of the intervention type and its theory of change, followed by descriptive results and the findings addressing our research questions.

8.1 School-based management (SBM) interventions

School-based management (SBM) interventions involve decentralising authority to the school level to improve the quality of school administration and leadership. SBM reforms take on many different forms but there are two key dimensions to the decentralisation of authority: (a) the degree of decision-making authority that is being devolved to the school level and (b) who is given the responsibility for the devolved functions (Barrera- Osorio *et al.,* 2009). The transfer of responsibility is usually made to a combination of principals, teachers and community members who may work through a school management committee. Some SBM programmes specifically encourage parental and community participation in the management of the school. The decision- making authority that is being devolved to the school may include any of the following: budget allocation, employment and remuneration of teachers, curriculum development, and procurement of educational material, infrastructure improvement and monitoring and evaluation of teacher and student performance (Barrera-Osorio *et al.* 2009).

School management committees may also devise school improvement plans and receive funds to finance implementation of these plans. For instance, the Education Quality Improvement Project in Cambodia encouraged school committees to identify their school's needs, suggest improvements and then carry out reforms using cash grants from the Ministry of Education (WDR, 2004).

8.1.1 How may SBM effect education outcomes?

SBM has gained increased popularity in developed and more recently in developing countries as it is regarded to have the potential to provide a low- cost means to increase the efficiency and accountability of education. However, SBM programmes are far from uniform and they differ in terms of their objective, strategies, stakeholders that are targeted and the specific policy and social context in which the SBM intervention is being implemented.

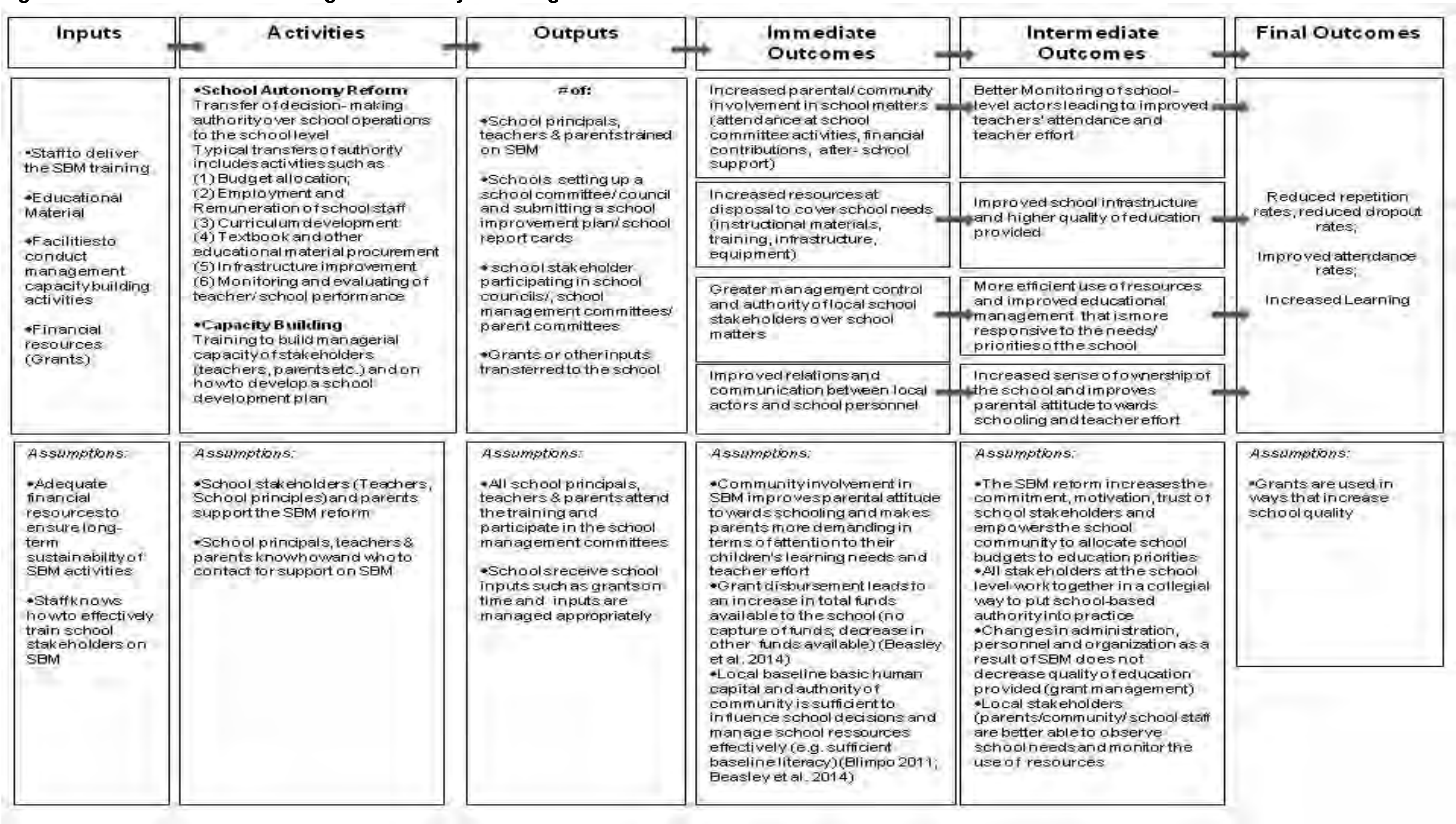
Although the type of SBM interventions included in our review are very diverse, the following programme theory as outlined in Figure 8.1a, will attempt to provide an overview of the different pathways through which SBM might lead to increased learning outcomes.

The type of inputs vary depending on the intervention; in some cases only funds are provided while in others the recruitment of trainers may also be required. One of the first activities will be introducing SBM to the school stakeholders to gain their support followed by capacity building activities to improve their management skills. Other typical inputs include learning materials and financial resources (payment of staff, grants, learning material).

The outputs of these activities will see the participants trained and aware of how SBM works. It is important that participants believe that SBM is a framework that helps schools to improve and that they are consequently are genuinely committed to it. Also, there may exist different power relations between participants. The model therefore assumes that all stakeholders at the school level work together in a collegial way to put SBM into practice. Immediate outcomes may be a greater level of parents' involvement in the school activities (in cash or in- kind) and in monitoring of school personnel. A second potential pathway to better education outcomes is increased empowerment and commitment of school stakeholders as a result of increased resources, knowledge and autonomy.

Finally, further along the causal chain, these will result in changes in practices at the school level such as (a) improved school management that is more responsive to the needs and priorities of the school; (b) more efficient use of resources by making school personnel more accountable and enhancing transparency. Increased participation of all school stakeholders is assumed to lead to a better school environment in general (more collegiate relationships etc.). As a result, higher quality of education is provided (Teachers' pedagogy/ attendance/ performance improves; improved school infrastructure) leading to improved students learning. Further assumptions include that grants and training result in increased spending on school infrastructure (not as a substitute to a community's own contribution) and that local decision makers are better able to understand their own school needs and have a strong incentive to demand high quality services (Beasley *et al.,* 2014; Carnoy *et al.,* 2008). At the same time, local baseline human capital and authority of the community needs to be sufficient to influence school decisions and manage school resources (Blimpo 2011; Beasley *et al.,* 2014).

Figure 8.1 a: School based management theory of change



8.1.2 Description of included studies

We included 14 studies reported in 21 different papers that evaluated the effect of SBM interventions in a low- or middle- income country. These studies evaluated 12 individual programmes, policy reforms or interventions. On several occasions we included more than one study that evaluated the same programme but had a different study team, undertook different analysis or reported different outcomes. For example, there are two studies that evaluated the Quality Schools Programme and two studies that evaluated the Apoyo a la Gestión Escolar initiative in Mexico. Table 8a provides details of the characteristics of included studies, summarised below.

Setting

The included studies evaluated 12 different programmes in countries in Latin America, East Asia, South Asia and Sub- Saharan Africa. Three programmes took place in Mexico (Bando, 2010; Gertler *et al.,* 2012; Murnane *et al.,* 2006; Skoufias *et al.* 2006; Santibanez *et al.,* 2014), three interventions were implemented in the Philippines (Khattri *et al.,* 2012; Yamauchi 2014; San Antonio (2008), and one each in Niger (Beasley *et al.,* 2014), Indonesia (Pradhan *et al.,* 2014), Brazil (Carnoy *et al.,* 2008), Senegal (Carneiro *et al.,* 2015), Sri Lanka (Aturupane *et al.,* 2014) and The Gambia (Blimpo *et al.,* 2015).

The majority of programmes are national programmes or reforms and several authors included samples from both rural and urban areas within a country (Skoufias *et al.* 2006; Beasley *et al.,* 2014; Blimpo *et al.,* 2011; Santibanez *et al.*, 2014; Carneiro *et al.*, 2015). Pradhan *et al.*, (2014) and Gertler *et al.*, (2012) restrict their sample to schools located in rural areas only whereas Murnane (2006) include schools exclusively located in urban areas. The remaining studies do not make that distinction (Carnoy *et al.*, 2008; Bando 2010; Khattri *et al.*, 2012; Yamauchi 2014; San Antonio 2008; Aturupane *et al.*, 2014).

Population

All programmes were targeted at public (government run) schools only. The majority of programmes were targeted at the primary school level (n= 10) and included students at different ages and grades. One programme in the Philippines (San Antonio; 2008) targets the intervention at the secondary school level only. The Apoyo de la Gestion programme in Mexico targeted both primary and secondary school levels (Gertler *et al.*, 2012; Bando, 2010). The study by Gertler *et al.*, (2012) specifically included a sample of non- indigenous primary schools whereas the study sample used by Bando (2010) consisted of television based lower secondary schools called *telesecundarias* where classes are taught on television and a teacher in the classroom reinforces the material.

Intervention

SBM programmes take many different forms, but all 12 programmes evaluated in the included studies contained some of the main components of a typical SBM as it is described in the literature (Barrera-Osorio, 2009). Apart from San Antonio (2008) and Pradhan *et al.* (2014) all the programmes in the included studies were implemented by national governments. All the studies include one or more of the following key SBM intervention components: decentralised decision making, grant provision, capacity building and development of a school improvement plan. Table 8b provides details of the intervention design components of included studies as detailed below.

Decision-making

In the majority of the included programmes, decision- making authority over school operations and funds were transferred to a school management committee. Although it is not always clearly specified, the committees typically consisted of the school principal, teachers and community representatives. In the case of the Apoyo a la Gestión Escolar (AGE) programme in Mexico authority over school matters were given exclusively to parents by providing monetary grants directly to parent associations (subject to audits) (Bando, 2010; Gertler *et al.*, 2012).

Grants

Apart from the Advisory School Council programme in the Philippines (SanAntonio, 2008) and the Programme for School Improvement in Sri Lanka (Aturupane *et al.*, 2014) all programmes gave school stakeholders control over funds from central or other relevant level of government. Programmes differed in terms of the amount and frequency of funds that provided to the school. For instance, the Quality Schools Programme in Mexico provided schools with an annual grant over five years while the COGES programme in Niger distributed a one- off grant of US$209 on average per school (US$1.83 per student). Several authors point out that the grant provided to the school only represented a small contribution relative to the school budget (Beasley *et al.*, 2014; Blimpo *et al.*, 2015; Pradhan *et al.*, 2014). In the case of the Whole School Development programme in The Gambia for instance the grant of US$500 provided in the first year of the programme represented less than 5 per cent of the average school budget. Funding of SBM activities in consequent years was expected to come from the school budget and funds raised at the local level (Blimpo *et al.*, 2015).

Some programmes also established certain requirements to receive the grant. For example, for parents to receive the grant in the AGE programme in Mexico a parent and student association had to be formed which had to attend trainings and develop a school improvement plan (Gertler *et al.*, 2012; Bando 2010). To qualify for the PEC programme, school stakeholders had to prepare school improvement plans that include an assessment of the school's problems and needs, specific objectives for improvement, and an annual improvement plan (Murnane *et al.*, 2006, Skoufias *et al.*, 2006). The main goal of the school grants programme in Senegal was to improve school quality by improving pedagogical resources in the school. Schools therefore had to complete a grant application for a school project addressing a particular pedagogical issue faced by the school (Carneiro *et al.*, 2015).

The extent to which schools were able to decide on what to use the grant for also differed between programmes. While some programmes did not pose any restriction on the usage of funds as in the case of the COGES programme in Niger, others limited usage to specific school activities. Common categories include training, materials or equipment and infrastructure improvements. In the case of the CP programme in Indonesia the grants received were mainly spent on holding school committee meetings (Pradhan *et al.*, 2014).

Capacity building

The majority of programmes (n= 10) include a capacity building component targeted at different school stakeholders. The training typically took the form of orientation workshops and seminars with the aim to build the skills necessary to fulfil assigned management responsibilities. Topics covered in the training included training on financial management, project planning or participatory skills. Most programmes also trained school stakeholders specifically on how to develop a school improvement plan or produce annual report cards. The exception is the Plano de Desenvolvimento da Escola (PDE) programme in Brazil where a training component is not specified (Carnoy, 2008).

Additional components

Several included programmes provided interventions in addition to the grant and capacity building component. Blimpo's study in The Gambia included the distribution of school management manuals in addition to the management training. The TEEP and BESRA programmes in the Philippines required schools to develop annual report cards on school performance to be shared with the community at the end of the school year (Khattri *et al.*, 2012, Yamauchi, 2014).

The Advisory School Council programme in the Philippines and the Programme for School Improvement in Sri Lanka differ from the other programmes in that they were limited to the establishment of advisory school councils/ school development committees. In the case of the ASC, the councils then conducted monthly meetings, which offered opportunities for school stakeholders to discuss school management concerns the principal referred to them (SanAntonio, 2008). As part of the PSI school had to establish a school development committees that are comprised of the principal (chair) and an elected group of teachers, parents and students and a school management team that included all school staff members. The committees were supposed to meet on a monthly basis. (Aturupane *et al.*, 2014).

Comparisons

The majority of included studies compare the effect of an intervention to a comparison group with no intervention (n= 13). The exception is SanAntonio's (2008) study in the Philippines where the control group received a different education intervention to the treatment group. Four studies reporting on three different programmes also evaluated more than one treatment arm (Skoufias 2006; Murnane *et al.*, 2006, Blimpo *et al.*, 2011; Pradhan *et al.*, 2014). Skoufias' and Murnane's evaluations of the Quality Schools Programme in Mexico both differentiate between two treatment groups based on the number of years the schools were exposed to the programme. For Blimpo *et al.*'s (2015) evaluation of the Whole School Development programme in The Gambia we chose the first treatment arm (WSD treatment) as it includes a capacity building component in addition to the grant. Pradhan *et al.* (2014) evaluate seven different combinations of SBM components (provision of grant blocks, training, elections of school committee members and linking school committees to village councils). The treatment arm that just provided a block grant was the only one compared to a business as usual comparison group so this is the treatment arm we included in our meta-analysis.

Outcomes

The included studies report effects on a wide range of education outcomes. Four studies assess enrolment, measured as the number of students listed in the school register or enrolled in first- grade. Two studies assess student attendance, either measured as the percentage of students absent on the day of the survey or as the average number of absences reported during the school year. Completion was assessed in eight studies, measured as either the number of students who passed the grade in a given school year or passed an end of primary school test. Seven studies assess student drop- out, measured either as the total number of students who have dropped out of school by the end of the school year or the number of students enrolled in grade x at the end of the school year. Finally, 11 of the included studies measured student learning, with the majority of studies using results from written tests in individual subjects (language arts, math) or a composite score. The tests were either standardized national achievement tests or developed by the study team. Two studies also report scores from orally administered reading and comprehension tests that measure, for example, a students' ability to correctly read letters and words (Beasley *et al.,* 2014; Blimpo *et al.,* 2015).

Three included studies also reported data on secondary outcomes of interest such as teacher attendance and teacher performance (Beasley *et al.,* 2014; Blimpo *et al.,* 2015; Aturupane *et al.* 2014). Teacher attendance in the three studies was measured as the percentage of teachers either present and absent respectively on the day of survey or the number of days a teacher took leave for vacation, medical, maternity, no pay or other purposes. Teacher performance measures include teaching practice variables such as whether the teacher has a written lesson plan or encourages the children to participate during class or frequency that the teacher gives homework to students.

The follow- up periods of outcome data collection after the start of a programme vary between the different studies. Four studies followed up 12 months or less after being exposed to the programme (SanAntonio, 2008; Beasley *et al.,* 2014; Blimpo *et al.,* 2015; Santibanez *et al*., 2014). The study evaluating the COGES programme in Niger has the shortest follow- up period, measuring effect of the programme only six months after the initial grant disbursement (Beasley *et al.,* 2014). Seven studies followed up after between 12 and 24 months of intervention exposure. The remaining three studies collected outcome data between 24 and approximately 36 months after the start of the intervention (Murnane *et al.*, 2006; Yamauchi, 2014; Carnoy *et al.*, 2008).

Study Design

Six of the included studies were cluster- randomised control trials where the intervention was assigned at the school level (SanAntonio, 2008; Beasley *et al.*, 2014; Pradhan *et al.*, 2014, Blimpo *et al.*, 2015; Bando, 2010; Aturupane *et al.*, 2014). Five studies are quasi-experimental studies, combining data before and after the intervention with matching and/ or regression techniques (Murnane *et al.*, 2006; Skoufias *et al.*, 2006; Yamauchi *et al.*, 2014; Carnoy *et al.,* 2008; Gertler *et al.,* 2012). Khattri *et al.*'s study in the Philippines is a natural experiment. The authors make use of a pipeline approach where schools selected for the programme participation, but not yet treated, are chosen for the control group. Using panel data, they then use a difference- in- difference strategy and with PSM to estimate programme impact.

Qualitative studies, process evaluations and project documents

We identified 15 additional documents that present qualitative, process and project information for five of our included programmes (PDE, PEC, TEEP, PSI, PEC- FIDE). The large amount of additional documents may be due to the fact that these are nationwide programmes or policy reforms that have been implemented over a number of years. We found a lack of additional documents for cluster- RCTs that are evaluating a one- off trial, often to inform future scale-up of the intervention to the rest of the country. Several of the impact evaluations also included qualitative components and therefore a main part of our qualitative synthesis is based on that. The intervention components are described below in more detail, followed by the descriptive findings on process, implementation and context.

Table 8.1 a: Characteristics of Included Studies School Based Management

Included study	Setting	Intervention summary	Included outcomes	Follow- Up	Study design	Sample Size
Murnane *et al.* (2006)	Mexico (urban), Primary school	Quality Schools programme, Programa Escuelas de Calidad (PEC). PEC combines increased resources for schools with decentralisation to allow school principals to make management decisions using the increased resources. Participation in PEC contains four different components: (1) School improvement plan; (2) School grants; (3) Parental involvement; (4) Professional development.	Completion; Dropout;	36 months	CBA with PSM	17,274 schools
Skoufias *et al.* (2006)	Mexico (rural & urban & peri-urban), Primary school	Quality Schools programme, Programa Escuelas de Calidad (PEC). PEC combines increased resources for schools with decentralisation to allow school principals to make management decisions using the increased resources. Participation in PEC contains four different components: (1) School improvement plan; (2) School grants; (3) Parental involvement; (4) Professional development.	Drop-out; Completion;	24 months	CBA with DID, PSM,	74,700 schools
Gertler *et al.* (2012)	Mexico (rural), Primary school	AGE (Apoyo a la Gestión Escolar)provides a small monetary grant (US$500- 700 depending on school size) to parent association provided certain requirements are met to cover variable costs (excluding wages, water, electricity, infrastructure; including materials, maintenance). Parents also receive training in the management of these funds and in participatory skills to increase their involvement in school activities.	Completion; Dropout; Enrolment	12 months & 24 months+	CBA with DID	6,027 schools

Bando (2010)	Mexico (not clear whether urban or rural), Secondary school	AGE (Apoyo a la Gestión Escolar) provides a small monetary grant (US$500- 700 depending on school size) to parent association provided certain requirements are met to cover variable costs (excluding wages, water, electricity, infrastructure; including materials, maintenance). Parents also receive training in the management of these funds and in participatory skills to increase their involvement in school activities.	Learning; Completion	Approx. 17 months	Cluster- RCT	57,386 schools
Khattri *et al.* (2012)	Philippines (not clear whether urban or rural), Primary school	Third Elementary Education Project (TEEP) includes the following key components: (1) School Development Plan; (2) Training for Principles and head teachers; (3) School grants; (3) Parental Involvement in developing the SIP.	Learning	24 months	Natural experiment	5,167 schools
Yamauchi *et al.* (2014)	Philippines (not clear whether urban or rural), Primary school	Basic Education Reform Agenda (BESRA) - School based management reform with the following key components: (1) Community Involvement; (2) Training for principal and other school staff; (3) School Improvement Plan: (4) School Grant	Learning	Approx. 36 months	CBA with DID/ PSM	3255 schools (not clear)
SanAntonio (2008)	Philippines (not clear whether urban or rural), Secondary school	The intervention involved introducing democratic school leadership. Main components of the intervention are: (1) Training of principles and other stakeholder groups; (2) Advisory School Councils	Learning, Teacher performance	12 months	RCT	76 schools

Carnoy *et al.* (2008)	Brazil (not clear whether rural or urban)	Plano de Desenvolvimento da Escola (PDE). Schools engage in a self-evaluation, develop a school plan focusing on two or three "efficiency factors" (one of which has to be effective teaching and learning), and design actions intended to address those factors.	Attendance; Completion; Learning; Dropout	24- 36 months	CBA	172 schools
Beasley *et al.* (2014)	Niger (urban & rural)	COGES (Comité de Gestion de l'Etablissement Scolaire) established by ministry of education. The COGES programme includes three major components: (1) Training for school committee members; (2) School Improvement Plan; (3) School Grants.	Attendance; Drop-out; Learning; Enrolment; Completion, Teacher Attendance	5- 6 months	Cluster- RCT	1000 schools
Pradhan *et al.* (2014)	Indonesia (rural), Primary school	Encouraging School Committee Participation- The treatment arm included in our review provided grant and Facilitation: Treatment schools received a block grant of USD 870 to help school committees catalyse change. The school committee was expected to develop a plan for expenditure with the help of facilitators and to be transparent by posting expenditure on a notice board. The block grant was transferred directly from the Ministry into school committee's bank account.	Completion; Drop-out; Learning	10 months	Cluster- RCT	520 schools

Blimpo *et al.* (2015)	The Gambia (appears to cover both rural and urban)	Whole School Development (WSD) included 3 components: (1) training; (2) Grant; (3) Distribution of management manuals addressing six specific topics pertaining to the management and functioning of schools: school leadership and management, community participation, curriculum management, teacher professional development, teaching and learning resources (e.g., textbooks and libraries), and the school environment.	Attendance; Enrolment; Learning; Teacher Attendance; Teacher Performance	36 months	Cluster- RCT	273 schools
Santibanez *et al.* (2014) **EER247**	Mexico (rural & urban), Primary School	As a spinoff of the PEC, The Programme to Strengthen and Invest Directly in Schools (PEC-FIDE), seeks to improve student achievement by providing cash grants to schools in exchange for collaborative school planning and shared decision making. The school councils are required to draft a five- year school improvement plan and a one-year work plan. Parent association may be involved in the school council. School principals and president of the school council receive basic training on SBM.	Learning; Completion; Dropout	12 months	CBA (PSM-, DID)	3675 students (3rd grade sample); 3575 students (6th grade sample)
Aturupane *et al.* (2014) **EER223**	Sri Lanka (not clear); Primary School	The Programme for School improvement (PSI) includes three main components: (1) Establishment of a School Development Committee (2) Establishment of a School Management Team (SMT). (3) Development of a medium-Term Plan and an Annual Implementation Plan	Learning; Teacher Attendance; Teacher performance	24 months	Cluster- RCT	100 schools (PSI treatment arm + control arm)
Carneiro *et al.* (2015) **EER209**	Senegal (rural & urban); Primary School	The School Grant Programme (SGP)- Through this programme, every elementary school could apply for funds for a specific school project that seeks to improve the quality of learning and teaching, with the best proposals being selected through a competitive process. Grants were to be prepared by a committee of parents, teachers and local officials.	Learning	12 & 18 months	Cluster- RCT	633 schools

Table 8.1 b: Intervention Design Features of included studies

	Murnane PEC	Skoufias PEC	Gertler AGE	Santibanez PEC-FIDE	Bando AGE	Khattri TEEP	Yamauchi BESRA	SanAntonio ASC	Carnoy PDE	Beasley COGES	Pradhan CP	Blimpo WSD	Aturupane PSI	Carneiro SGP
Intervention design features														
Decision- Making **(1) Community/ Parents** **(2) Schools (Committees/ principals)**	2	2	1	2	1	2	2	2	2	2	2	2	2	2
Grant Provision	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓
Capacity building activities	✓	✓	✓	✓	✓	✓	✓	✓	Not clear	✓	✓	✓		
School improvement plan	✓	✓	✓	✓	✓	✓	✓		✓		✓	✓	✓	
Additional Features						✓	✓	✓			✓	✓		
Implementation features														
Requirements to receive grants	✓	✓	✓	✓	✓	✓	Not clear	n/a	✓		✓	Not clear	n/a	✓
Restrictions on activities to be supported by grant	✓	✓	✓	✓	✓	Not clear	Not clear	n/a	✓		✓	✓	n/a	✓

8.1.3 Synthesis of findings

The results of our synthesis are presented in two sections. First, we present the findings of the meta-analysis on the effects of SBM on primary and secondary outcomes (Questions 1a and 1b) Second, we present a discussion of the findings incorporating evidence from our descriptive qualitative synthesis to assess factors related to intervention design, implementation and context which might act as barriers or facilitators of the effectiveness of PPP (Questions 2a and 2b).

Effects of SBM interventions on enrolment, attendance, dropout rates, completion and learning outcomes

This section reports the results of the meta-analysis of the effects of SBM, addressing question 1a. We structured the presentation of results according to the 'ideal type' theory of change (Figure 8.1a), starting with intermediate/ secondary outcomes (teacher attendance), followed by education access outcomes (enrolment, attendance, drop out) and final outcomes (completion, learning outcomes: composite test scores, language arts test scores, maths test scores).

The studies include a range of different follow up periods, with the majority of studies including a data point within the 18 month-24 month periods. Therefore we selected these for the meta-analysis when available. For some studies we only have data on a shorter follow up period. We have included these in the meta-analysis and conduct sensitivity analysis.

For Blimpo *et al.*'s (2015) study we selected the whole school development treatment arm for the meta- analysis, as it is most similar to the design of other included SBM programmes. Aturupane *et al.* (2014) evaluate the impact of two programmes; the Programme for School Improvement (PSI) and the School Report Card Programme. We only included the PSI program in our meta-analysis as the school report card programme was never implemented as stated by the authors. For Carnoy *et al.*'s (2004) study did not contain the necessary data for us to calculate an effect size for language arts outcome.[115] Both Skoufias *et al.* (2006) and Murnane *et al.* (2006) evaluating the PEC programme both measure the same completion outcome based on the same dataset. We used the completion data from Murnane *et al.* as the follow up period for the two groups is clearer compared to Skoufias *et al.* Pradhan *et al.* (2014) evaluate seven different combinations of SBM components (provision of grant blocks, training, elections of school committee members and linking school committees to village councils). The treatment arm that just provided a block grant was the only one compared to a business as usual comparison group so this is the treatment arm we included in our meta-analysis.

Fifteen studies provided data for meta-analysis, but none of the studies reported on all outcomes. The number of comparisons with effect sizes range from one for attendance to seven for maths outcomes[116]. All effect sizes have been expressed as standardised mean difference (SMD), interpreted as the magnitude of the number of standard deviation changes in the outcome for the intervention group as compared to students in non-SBM schools. SMD scores are have been interpreted as the number of standard deviation changes in the outcome.

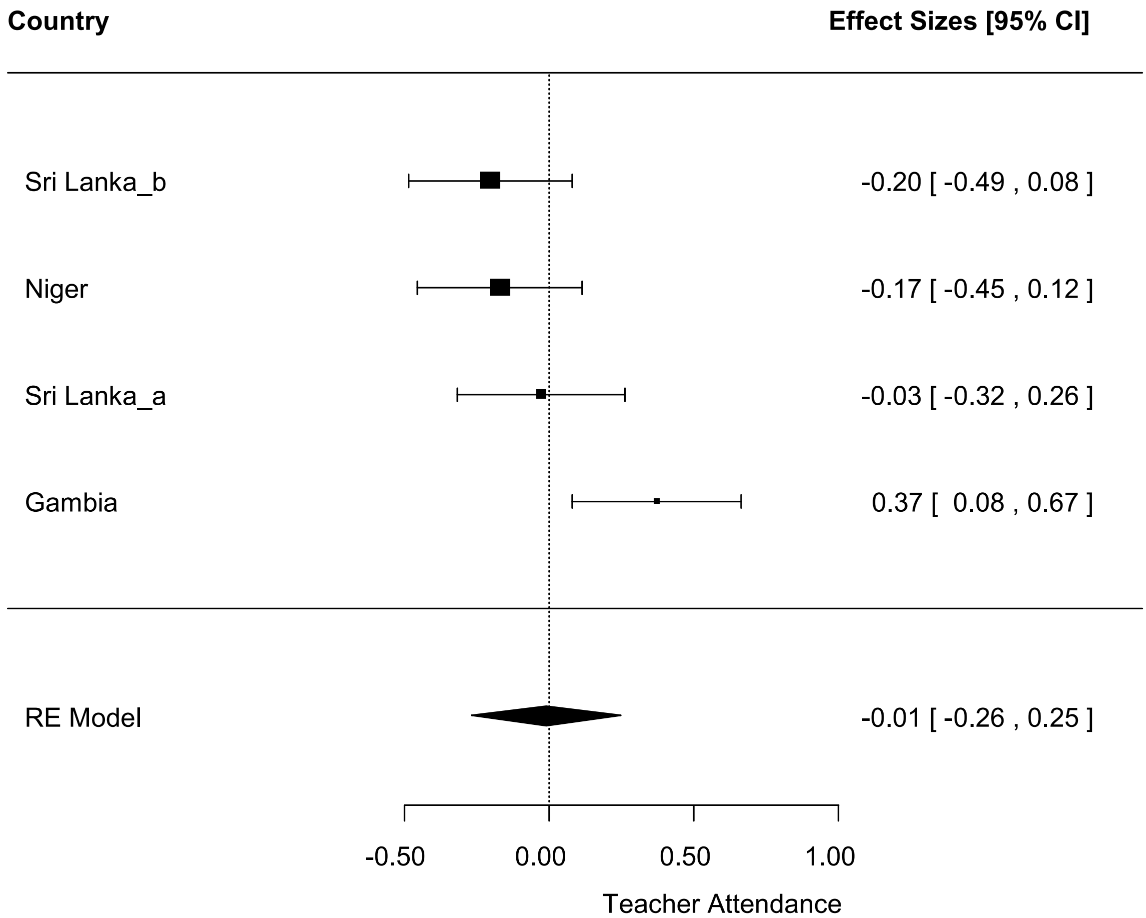
[115] The authors were contacted but we received no clarification

[116] Three studies reported on a range of teacher performance measures. We did not combine them in a meta- analysis as they are too different.

Teacher attendance

The overall average effect of SBM on teacher attendance is -0.01, 95% CI [-0.26, 0.25], calculated under random-effects model. The assessment of homogeneity suggests that the effects do not arise from a common population (I^2 = 68.65%, τ^2 = 0.0472, Q(df = 3) = 9.5090, p-val = 0.0232). Figure 8.1B presents the forest plot with the results of the individual studies and the pooled point estimate. The effects range from -0.20, 95% CI -0.49, 0.08] for the grade 8 sub-sample in Sri Lanka (Aturupane *et. al.,* 2014) to 0.37, 95 % CI [0.08, 0.67] in the Gambia (Blimpo *et al.,* 2015). As expected, the results are sensitive to the inclusion of the study from the Gambia. When removing this study from the analysis the negative effect becomes substantively larger in magnitude, although it remains statistically insignificant (SMD= -0.13, 95% CI [-0.30 0.03]).

Figure 8.1 b: Teacher Attendance[117]

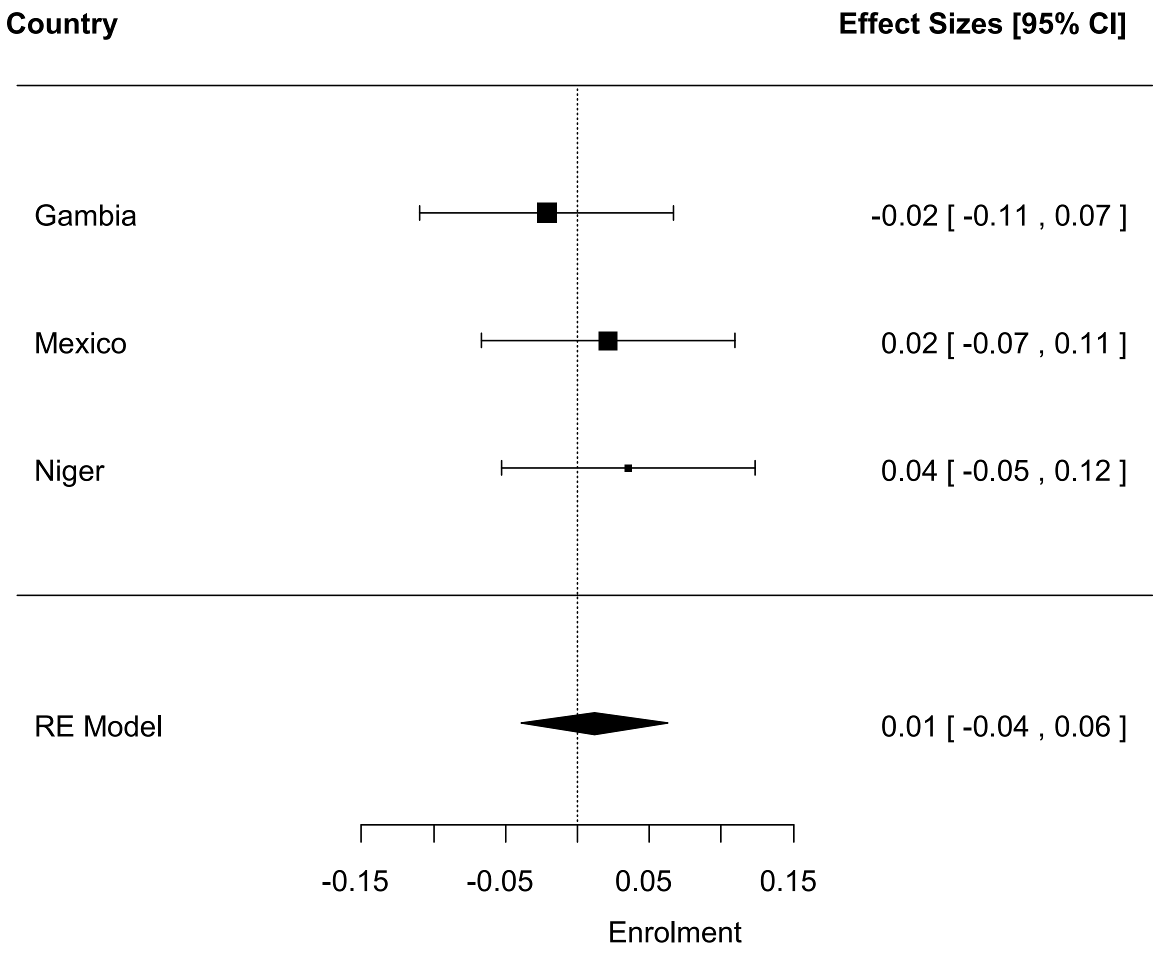


Enrolment

The overall average effect of SBM on enrolment is virtually zero (SMD = 0.01, 95% CI [-0.04, 0.06]), calculated under fixed-effect model. The assessment of homogeneity suggests that the only source of variation is within-study sampling error (I^2 = 0.00%, τ^2 = 0, Q(df = 2) = 0.8525, p = 0.6530). Figure 8.1C presents the forest plot with the results of the individual studies and the pooled point estimate, as expected the confidence intervals overlap for these studies.

[117] Sri Lanka_a corresponds to the grade 8 sub-sample and Sri Lanka_b corresponds to the grade 4 sub-sample.

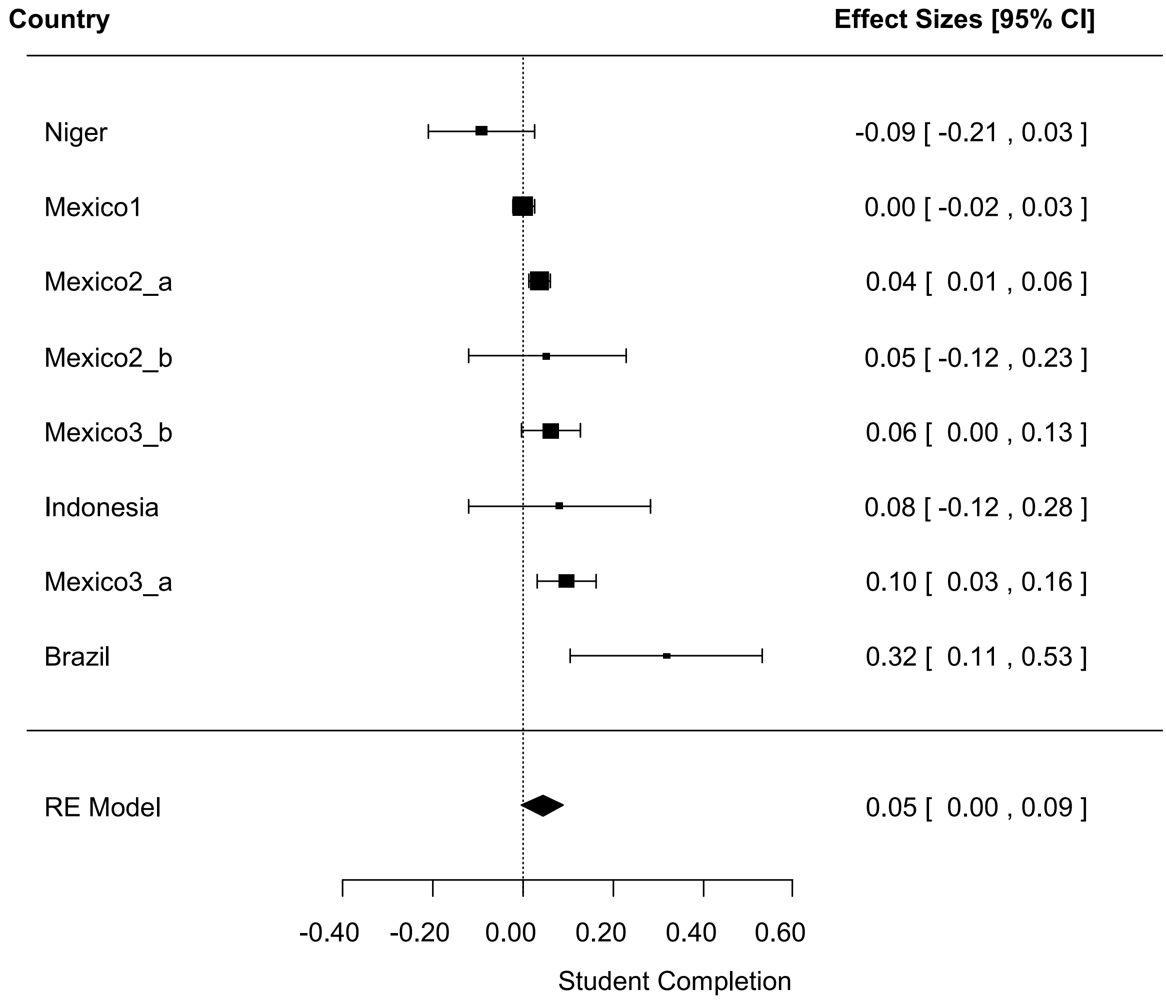
Figure 8.1 c: Enrolment



Completion

The overall average effect of SBM on completion is not different than zero under random-effects model (SMD = 0.05, 95% CI [0.00, 0.09]). The assessment of homogeneity suggest large amount of between-studies variability (I^2 = 77.18%, τ^2 = 0.0024 Q(df = 7) = 22.6108, p-val = 0.0020). This is also apparent when inspecting the forest plot in Figure 8.1D. The effect sizes range from -0.09 [-0.21, 0.03] in Niger, to 0.32 [0.09, 0.55] in Brazil. The average effect is sensitive to the inclusion of both the study from Indonesia (Pradhan *et al.,* 2014) and Brazil (Carnoy *et al.,* 2008). When removing any of these studies the result changes into a small and negative effect, although in both cases the effect remains statistically insignificant (See Appendix H for results of all sensitivity analyses).

Figure 8.1 d: Completion[118]

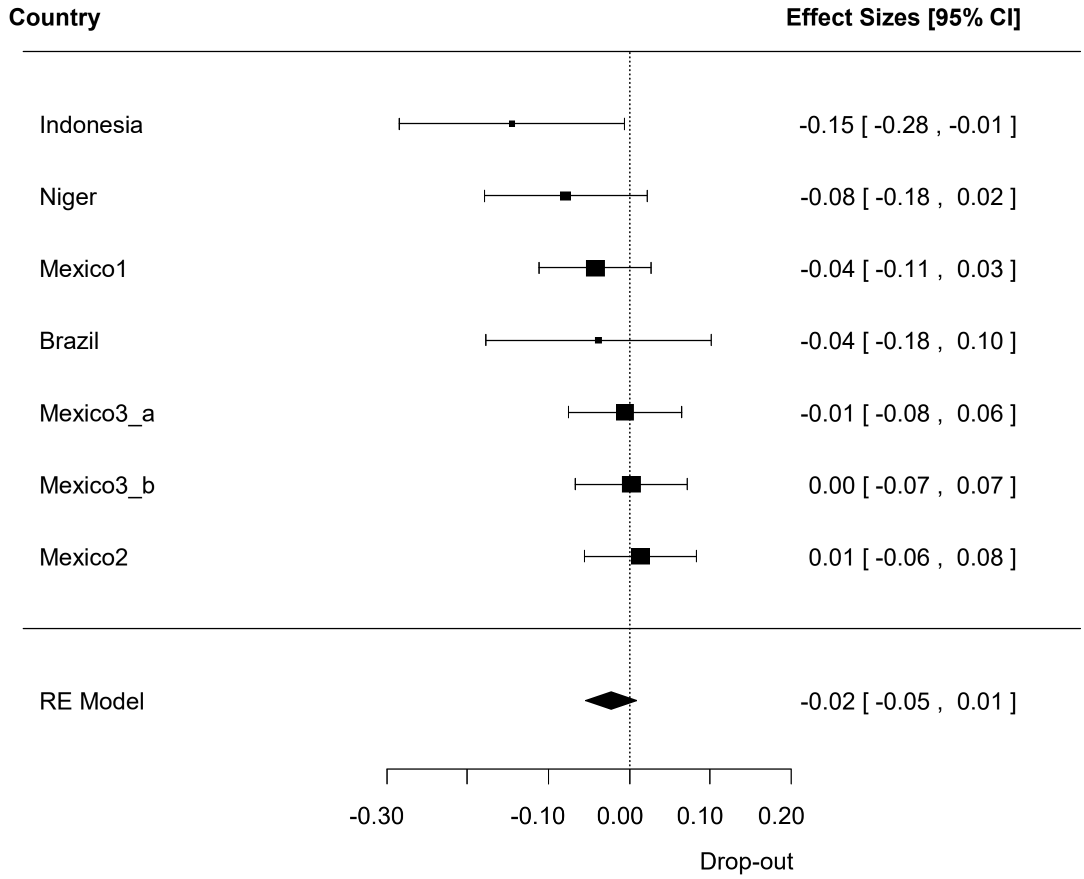


Drop-out

The overall average effect of SBM on drop-out is small and negative, but not statistically significant (SMD = -0.02, 95% CI [-0.05, 0.01]), calculated under random-effects model. This negative effect represents a positive result in terms of desirability. The assessment of homogeneity suggests little between-studies variation ($I^2 = 0.44\%$, $\tau^2 = 0.0000$, Q(df = 6) = 6.2609, p-val = 0.3946). The overall average effect is not sensitive to the removal of any one study (see Appendix H for results of all sensitivity analyses).

[118] Mexico1 refers to Murnane *et al.,* 2006 (average student failure rate in school year t); Mexico2_a refers to Gertler *et al.,* 2012 (Proportion of students that fail their grade); Mexico2_b refers to Bando, 2010 (failure rate); Mexico3_b refers to Santibañez *et al.,* 2014 (Pass rate for 6th graders); Mexico3_a refers to Santibañez *et al.,* 2014 (Pass rate for 3rd graders).

Figure 8.1 e: Drop-out[119]

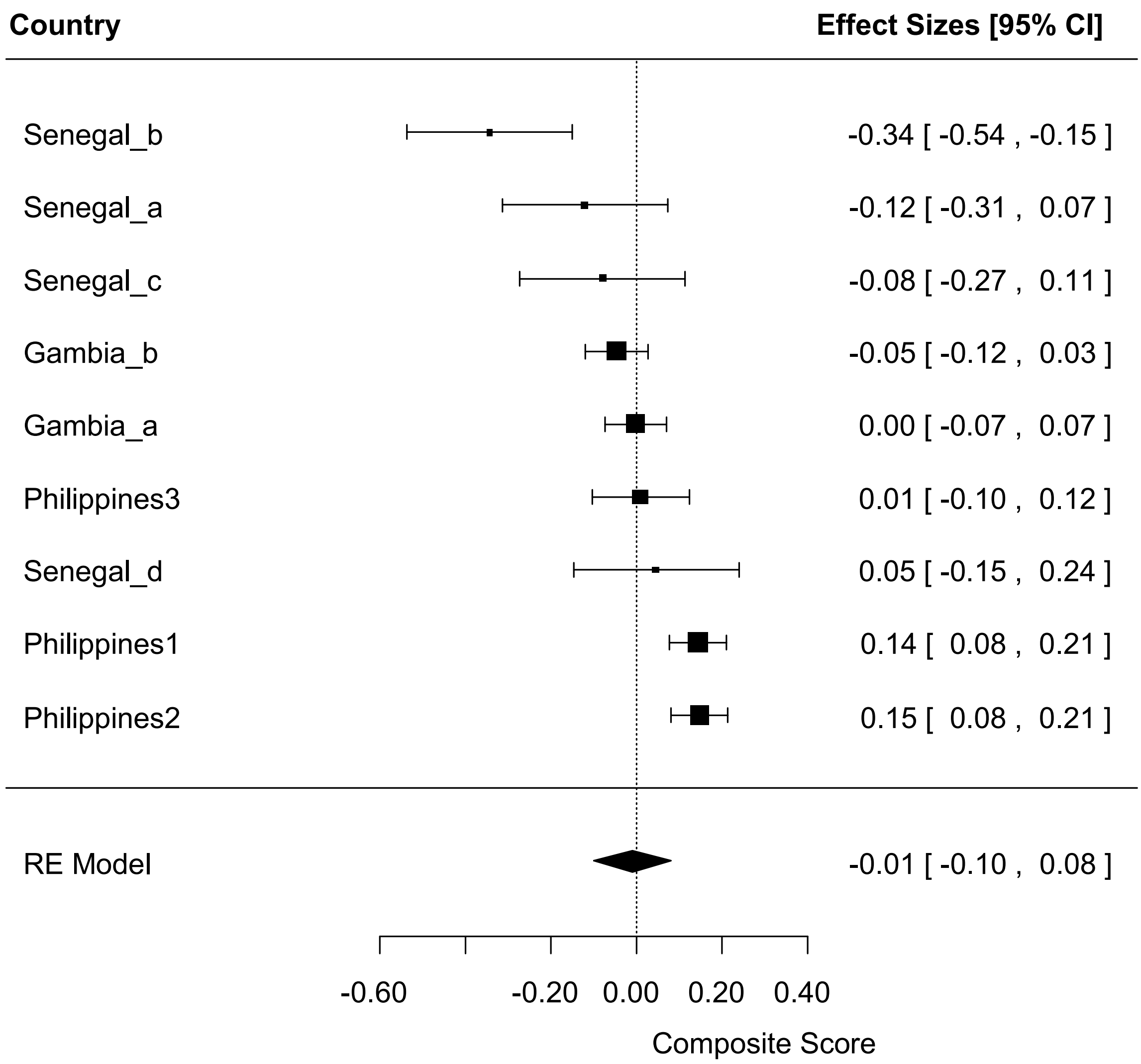


Composite scores

The overall average effect of SBM on learning outcomes as measured by a composite score is -0.01, 95% CI [-0.10, 0.08], calculated under a random effects model. The homogeneity tests (I^2 = 85.65%, τ^2 = 0.0148, Q(df = 8) = 45.2978, p-val= < .0001) indicate that the effects did not arise from the same population. The forest plot in Figure 8H supports the presence of heterogeneity. The effect sizes range from -0.34, 95% CI [-0.54, 0.15] in Senegal to 0.15, 95% CI [0.08, 0.21] in one of the studies from the Philippines. Nevertheless, sensitivity analysis indicates that removing any one of these studies does not make a substantive difference to the overall pooled effect (see Appendix H for results of all sensitivity analyses).

[119] Mexico2 refers to Gertler 2012 (full sample); Mexico1 refers to Murnane 2006 (full sample); Mexico3_a refers to Santibañez
et al. 2014 (grade 3); Mexico3_b refers to Santibañez et al. 2014 (grade 6)

Figure 8.1 f: Composite Scores[120]



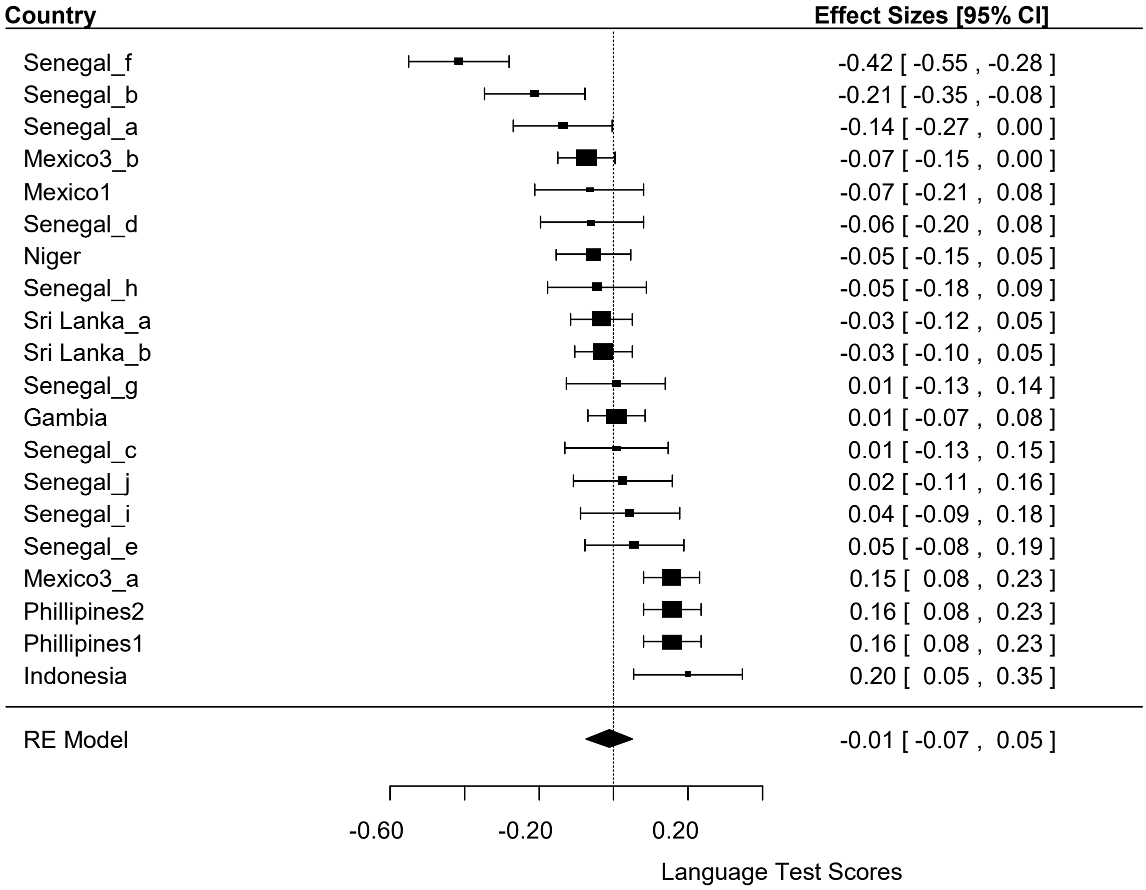
Language arts scores

The overall average effect of SBM on language arts test scores is -0.01, 95% CI [-0.07, 0.05]), calculated under a random effects model. The homogeneity test (I^2 = 84.87%, τ^2 = 0.0153 , Q(df = 19) = 114.9422, p-val < .0001) suggest large amount of between-studies variability. Figure 8.1G supports the presence of heterogeneity and shows a similar pattern to the forest plot for composite test scores. The effect sizes range from -0.42, 95% CI [-0.55, - 0.28] in Senegal (Carneiro *et al.,* 2015) to 0.20, 95% CI [0.05, 0.35] in Indonesia (Pradhan *et al.,* 2014). There is a cluster of ten comparisons where the point estimate is negative, and a cluster of four comparisons with relatively large, statistically significant effects. The confidence intervals between these two groups do not overlap. However, the removal of any

[120] Senegal_b refers to Carneiro *et al.* 2015 (Grade 3 female at second follow-up); Senegal_a refers to Carneiro *et al.* 2015 (Grade 2 male at second follow-up); Senegal_c refers to Carneiro *et al.* 2015 (Grade 5 male at second follow-up); Gambia_b refers to Blimpo *et al.* 2015 (Full sample); Gambia_a refers to Blimpo *et al.* 2015 (Full sample); Philippines3 refers to San Antonio 2008 (Full sample); Senegal_d refers to Carneiro *et al.* 2015 (Grade 5 female at second follow-up); Philippines1 refers to Khattri *et al.* 2010 (Full sample); Philippines2 refers to Yamauchi 2014 (Full sample)

one study does not make a substantive difference to the overall pooled effect (see Appendix H for results of all sensitivity analyses).

Figure 8.1 g: Language arts test scores[121]

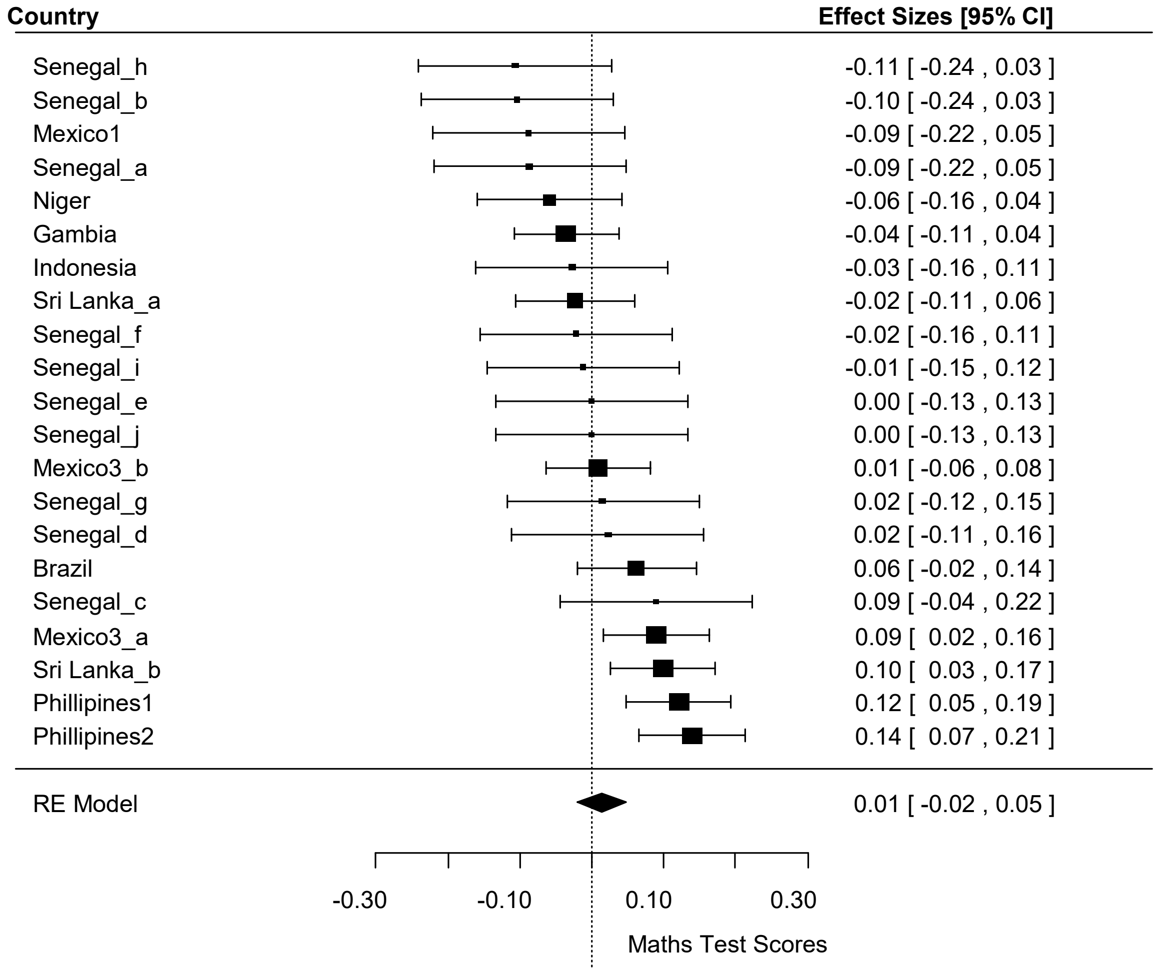


Maths test scores

The overall average effect of SBM on maths test scores is 0.01, 95% CI [-0.02, 0.05], calculated under a random effects model. The homogeneity test suggests a moderate amount of between-studies variability (I^2 = 56.40%, τ^2 = 0.0033, Q(df = 20) = 45.9563, p-val = 0.0008). Figure 8.1H provides the forest plot with the pooled effect size, as well as the effects from different studies, which range from -0.11, 95% CI [-0.24, 0.03] in Mexico (Carneiro *et al.,* 2015) to 0.14, 95% CI [0.07, 0.21] in the Philippines (Yamauchi, 2014). The confidence intervals are overlapping between most of the studies, and the overall results are not substantively sensitive to the removal of any of these studies (see Appendix H for results of all sensitivity analyses).

[121] Senegal_f refers to Carneiro *et al.* 2015 (Grade 2 female); Senegal_b refers to Carneiro *et al.* 2015 (Grade 3 female at second follow-up); Senegal_a refers to Carneiro *et al.* 2015 (Grade 3 male at second follow-up) ; Mexico3_b refers to Santibañez *et al.* 2014 (Grade 6); Mexico1 refers to Bando 2010 (Main sample); Senegal_d refers to Carneiro *et al.* 2015 (Grade 5 female at second follow-up); Senegal_h refers to Carneiro *et al.* 2015 (Grade 4 female); Sri Lanka_a refers to Aturupane *et al.* 2014 (Grade 8); Sri Lanka_b refers to Aturupane *et al.* 2014 (Grade 4); Senegal_g refers to Carneiro et al. 2015 (Grade 4 male); Senegal_c refers to Carneiro *et al.* 2015 (Grade 5 male at second follow-up); Senegal_j refers to Carneiro *et al.* 2015 (Grade 6 female); Senegal_i refers to Carneiro *et al.* 2015 (Grade 6 male); Senegal_e refers to Carneiro *et al.* 2015 (Grade 2 male); Mexico3_a refers to Santibañez *et al.* 2014 (Grade 3); Philippines2 refers to Yamauchi 2014 (Main sample); Philippines1 refers to Khattri *et al.* 2010 (Main sample)

Figure 8.1 h: Maths Test Scores[122]



8.1.4. Summary of findings and discussion

We identified fifteen studies of SBM across six different countries in Latin America, East Asia and Sub- Saharan Africa. We were able to examine effects on teacher attendance, enrolment, completion, dropout, maths, language arts and composite test scores using meta-analysis. The overall average effects range from -0.01, for language art test scores (95% CI [-0.07, 0.05]), for composite test scores (95% CI [-0.10, 0.08]) and teacher attendance (95% CI [-0.26, 0.25] to 0.05, 95% CI [0.00, 0.09], 0.15] for completion. For some outcomes there is a large amount of between study variability. The large amount of heterogeneity is not surprising given the variability in the intervention design, context and populations in our included studies.

[122] Senegal_b refers to Carneiro *et al.* 2015 (Grade 4 female at second follow-up); Senegal_b refers to Carneiro *et al.* 2015 (Grade 3 female at second follow-up); Mexico1 refers to Bando 2010 (Main sample); Senegal_a refers to Carneiro *et al.* 2015 (Grade 3 male at second follow-up); Sri Lanka_a refers to Aturupane *et al.* 2014 (Grade 8); Senegal_f refers to Carneiro et al. 2015 (Grade 2 female at second follow-up); Senegal_i refers to Carneiro *et al.* 2015 (Grade 6 male at second follow-up); Senegal_e refers to Carneiro *et al.* 2015 (Grade 2 male at second follow-up); Senegal_j refers to Carneiro *et al.* 2015 (Grade 6 female at second follow-up); Mexico3_b refers to Santibañez *et al.* 2014 (Grade 6); Senegal_g refers to Carneiro *et al.* 2015 (Grade 4 male at second follow-up); Senegal_d refers to Carneiro *et al.* 2015 (Grade 5 female at second follow-up); Senegal_c refers to Carneiro *et al.* 2015 (Grade 5 male at second follow-up); Mexico3_a refers to Santibañez *et al.* 2014 (Grade 3); Sri Lanka_b refers to Aturupane *et al.* 2014 (Grade 4); Philippines1 refers to Khattri *et al.* 2010 (Main sample); Philippines2 refers to Yamauchi 2014 (Main sample).

Section A8.3 in appendix J provides the results of the full qualitative synthesis, tables 8.1c and 8.1d provide the summary findings from this synthesis, together with details about the context for which they apply. Most of the findings are programme specific and therefore we have not described them all in detail here (the full write up of qualitative findings can be found in Appendix J). The characteristics of the available data is limited both in terms of volume and quality, and the findings are descriptive and often context specific. Therefore we were unable to identify any generalisable, stand-alone findings addressing our secondary research questions.

However, there are a few outliers that may present some additional insights. One example relates to two programmes that were implemented in the Philippines: Across all learning outcomes (math, language arts and composite test scores) the Third Elementary Education Project (Khattri *et al.*, 2012, corresponding to Philippines 1 in the forest plot) which was then mainstreamed into the system wide BESRA programme (Yamauchi *et al.*, 2014, corresponding to Philippines 2 in the forest plot) consistently showed positive effects. For the maths and composite scores outcomes these two programmes showed the largest positive effects ranging from 0.12, 95% CI [0.05, 0.19] for maths in Phillipines1 (Khattri *et al.*, 2012) to 0.16, 95% CI [0.08, 0.23] for language arts in Phillipines2 (Yamauchi *et al.*, 2014). Although we could not come to a conclusion as to why this effect was observed, information from the study suggested that schools receiving the intervention had higher baseline human and social capital as they had to design good school improvement plans in order to receive the SBM grants (Khattri *et al.*, 2006; Yamauchi *et al.*, 2014).

On the other hand, several of the studies implemented in Africa consistently showed zero or negative effects across several outcomes: For example, the COGES programme implemented in Niger (Beasley *et al.*, 2014) had the largest negative effect on teacher attendance, completion, language arts and maths test scores out of all included programmes ranging from -0.12, 95% CI [-0.24, 0.00] for teacher attendance to -0.05 95% CI [-0.15, 0.05] for language arts test scores. Qualitative evidence from both the WSD and COGES programmes implemented in Niger and Gambia both make reference to low baseline human and social capital of school stakeholders, which may present a barrier to success of SBM programmes implemented in these contexts.

In summary, the available evidence suggests a beneficial effect on students in schools that were part of SBM programmes compared to those that were not. However, average effects are relatively small in magnitude and there is a large amount of between study variability. Many of the average effects include examples where programmes have had large and substantively important (positive and negative) effects on children's access to education and learning. Therefore the average effects should be interpreted with caution.

Table 8.1 c: Descriptive findings: Process and implementation

Descriptive findings: Process and implementation	Context	Citation (Info type)
Uneven implementation between states, with divergence from targeting criteria and the timing and amount of school grants	Mexico	Skoufias *et al.* 2006; Santibanez *et al.* 2014 (Impact Evaluation)
Grants may not have been disbursed as intended, with delays to completion of grant disbursement reported in both Mexico and Niger	Mexico, Niger	Blimpo *et al.* 2015 (IE), Skoufias *et al.* 2006 (Impact Evaluation)
High compliance rates for grant distribution	Gambia	Blimpo *et al.* 2015 (Impact Evaluation)
Some schools dropped out due to changes in leadership, conflict (among administrators, teacher and parents) and the work load that accounting for the grant money imposed	Mexico	Murnane *et al.* 2006 (Impact Evaluation)
Treatment schools showed a higher rate of adoption of the SBM concept compared to the control group including higher rates of establishment of school committees	Niger, Sri Lanka	Blimpo *et al.* 2015, Aturupane *et al.* 2014 (Impact Evaluation)
SBM did not lead to an increase in parents' engagement with schools	Mexico, Indonesia	Bando 2010, Pradhan *et al.* 2014 (Impact Evaluation)
There was an increase in parents' engagement with schools	Gambia, Mexico	Beasley *et al.* 2014,, Skoufias *et al.* 2006, Gertler *et al.* 2012, Santibanez *et al.* (2014) (Impact Evaluation
Parents were willing to try to improve school quality by participating in school committees, but they were not able to enforce rules and do so	Niger	Beasley *et al.* 2014 (Impact Evaluation)
Few school councils functioned as a collaborative planning or shared decision-making tool and were limited to signing off on decisions made by the principal.	Mexico, Sri Lanka	Santibanez *et al.* 2014 Aturupane *et al.* 2014 (Impact Evaluation, Additional Document)

School councils fulfilled an important monitoring function that encouraged transparency and ensured that resources would actually be spent in schools.	Mexico	Santibanez *et al.* 2014 (Impact Evaluation)
Teachers may feel resistant to SBM as they perceive it as undermining their authority	Niger	Beasley *et al.* 2014 (Impact Evaluation)
Treating teachers as allies and investing in teachers working conditions may reverse a negative reaction to SBM	Niger	Beasley *et al.* 2014 (Impact Evaluation)
The school principal plays a key role in motivating stakeholders to participate in school governance	Philippines	SanAntonio 2008 (Impact Evaluation)
While there were no effects on student achievement, SBM led to higher levels of self- empowerment, commitment to work for school improvement and trust in school authorities among stakeholders compared to the control group	Philippines	SanAntonio, 2008 (Impact Evaluation)
Grants were focused on construction and other material inputs, rather than books, learning materials or teacher training	Mexico, Niger	Bando 2010, Beasley *et al.* 2014, Skoufias *et al.* 2006 (Impact Evaluation)
The major share of funds went to teaching materials and teacher training	Brazil	Carnoy *et al.* 2008 (Impact Evaluation)
Some teachers and principals found the administrative work and time spent on community engagement burdensome, potentially taking away time spent on pedagogical responsibilities	Mexico, Philippines, Gambia	Murnane *et al.* 2006, Khattri *et al.* 2012, Blimpo *et al.* 2015 (Impact Evaluation)

Table 8.1 d: Descriptive findings: Contextual Factors

Descriptive findings: Contextual factors	Context	Citation (info type)
The social capital and education of parents may influence their ability to hold the schools accountable and participate effectively in school management	Niger, Gambia	Beasley *et al.* 2014, Blimpo *et al.* 2015 (Impact Evaluation)
Engagement with existing community institutions in the planning of educational activities may be important in contexts where such institutions are powerful	Indonesia	Pradhan *et al.* 2014 (Impact Evaluation)
Capacity of state departments of education to provide support and training to schools implementing SBM may play an important factor in determining the effectiveness of the SBM	Mexico	Murnane *et al.* 2006 (Impact Evaluation)
Centralised, hierarchical education systems may present a challenge for implementing some SBM programme goals such as increasing the involvement of parents and the community	Niger, Mexico	Beasley *et al.* 2014, Murnane *et al.* 2006 (Impact Evaluation)
Low teacher quality, reduced instructional time due to widespread double- shift schools, and teacher compensation may have been a barrier to the success of SBM	Niger	Blimpo *et al.* 2015 (Impact Evaluation)
Awareness sessions may have been essential in clarifying objectives of the programme and overcoming scepticism by schools that the programme would decrease government support and hand over power to parents (Aturupane et al. 2014)	Sri Lanka	Aturupane *et al.* 2014 (Additional Document)

8.2 Community-based monitoring interventions

It has been widely recognised that the quality of governance is central to development in L&MICs (World Bank. 2004). Resources needed to provide services for citizens are often lost because of widespread corruption and inefficiencies resulting in low quality services and in some cases no services at all often affecting the poor disproportionally (Bruns *et al.*, 2011). In the education sector this may result in inequitable allocation of public education spending, funding leaks or teacher absence. In this context, the concept of accountability has gained increasing importance in development policy and practice.

One specific focus has been on social accountability, a demand-side approach towards building accountability which consists of 'actions and mechanisms that citizens, communities, independent media, and civil society organizations can use to monitor and hold public officials accountable' (Agarwal *et al.*, 2009). Community based monitoring (CBM) initiatives is one example of such programmes, defined as "interventions where the community is given the opportunity to participate in the process of monitoring service delivery, where monitoring means being able to observe and assess providers' performance and provide feedback to providers and politicians" (Molina *et al.,* 2013). Interventions of this type are used in many sectors, including education.

8.2.1 How may CBM affect education outcomes?

CBM initiatives are based on the idea that citizens are in the best position to see what is going on as they regularly interact with providers (World Bank, 2004). They seek to improve the representation of communities and thereby improve the accountability of providers, governments, or other public bodies operate towards the people they serve (Westhorp *et al.*, 2013). Figure 8.2a provides a programme theory with the main programme components and mechanisms through which CBM may contribute to improved education outcomes, drawing on the programme theory developed by Molina *et al.* (2013).

CBM programmes can include several components but are usually centred on an information campaign to either make a particular accountability mechanism known or to provide information about the current performance of education providers. Information provision can also be used as a tool in public resource tracking, with the aim of reducing leakage from corruption (Bruns *et al.*, 2011). In some cases, information is also given on parents' entitlements to resources such as bursaries for students or school feeding or on the benefits to schooling. Different means may be used to reach the community. It may involve active engagement of parents through meetings in schools (Nguyen, 2008) or in the village (Bannerjee *et al.*, 2010), or less direct approaches such as newspaper/ radio or local TV campaigns (Reinikka & Svensson, 2004).

Information campaigns are often coupled with capacity building activities such as the provision of monitoring tools or training on how to monitor services or use tools to assess the status of a child's learning. A commonly used monitoring tool is a report card or score card that provide information on school and/or child performance in test scores or allow parents and communities to rank the performance of their school and teachers (Bruns *et al.*, 2011).

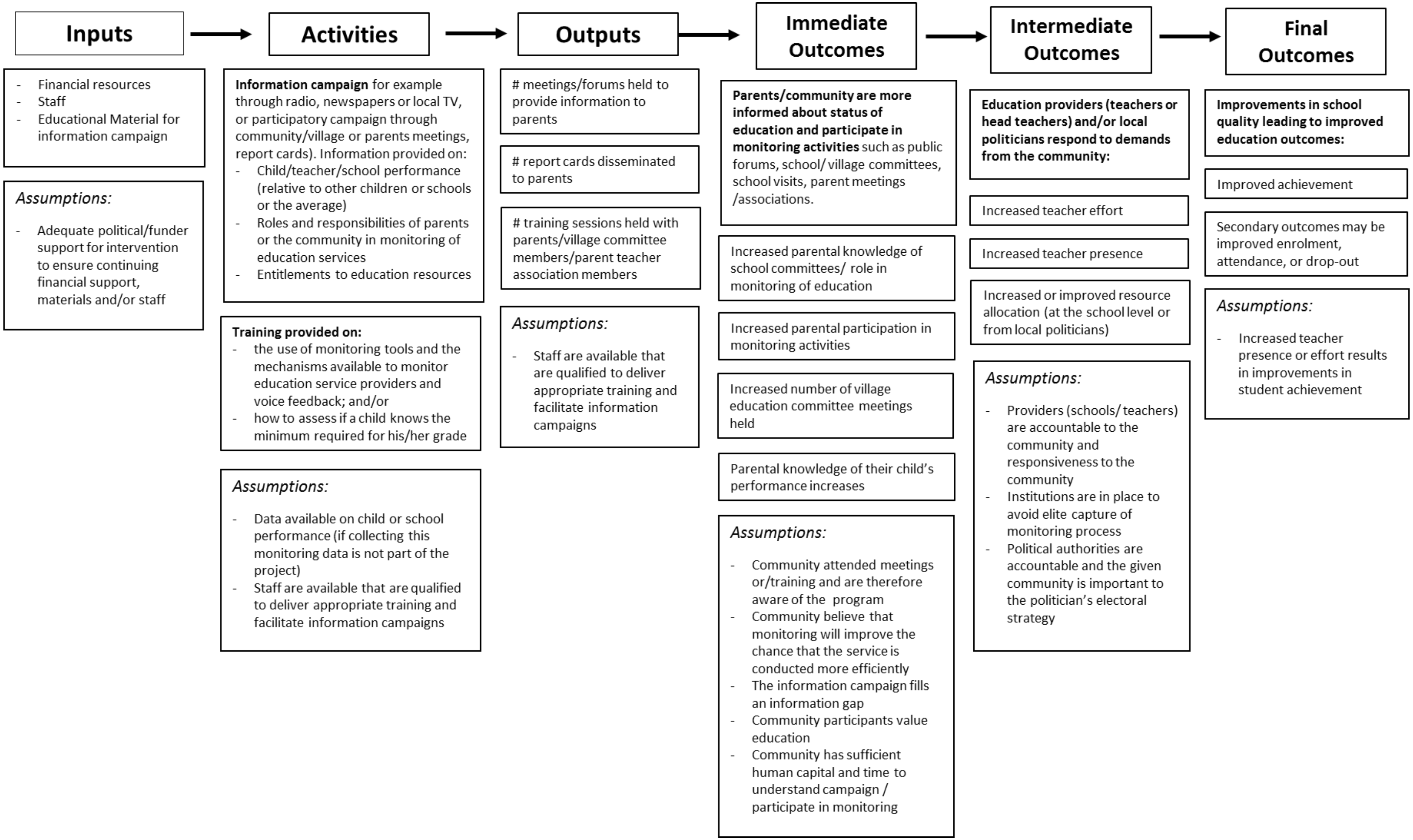
Empowered with these tools as well as knowledge on how to hold providers accountable the theory assumes that citizens decide to take action and participate collectively in monitoring activities. Depending on the type of CBM intervention, these can take the form of public forums, joining school or village committees, school visits, and parents meetings as well as collecting information on problems encountered or filing complaints (Molina *et al.,* 2013). In addition, it is expected that participating citizens will share the information acquired through monitoring activities with fellow citizens who will then join to pressure and lobby providers and politicians.

The programme theory makes a strong link between citizen's empowerment and increased voice through monitoring activities and improved service delivery. By providing communities with comparative information on school or teacher performance, and/ or training and tools to monitor education service providers, it is hypothesised that CBM will increase responsiveness and accountability of providers and politicians (Wild *et al.,* 2011). Such improved responsiveness may manifest itself in different ways, including improved teacher attendance, teaching quality, school management or resource allocation in the education sector. The theory then suggests that the improvement of the quality of educational services may then lead to improved education outcomes.

Several assumptions must hold for CBM programmes to lead to improved education outcomes. For instance, community participants (students, parents, marginalised groups, community leaders) must have interest and incentive to monitor and hold their providers to account for the delivery of services (Molina *et al.,* 2013). In order to do so they need to receive adequate information on the performance of educational services provided and on how to monitor education service providers. They also need to be able to pay the opportunity cost to participate in monitoring activities and have the human capital to do so effectively. Finally, community participants need to believe they can exercise their power and act on this to make evidence- based demands and coordinate action to collectively pressure service providers and the government. Further, politicians and schools then need to have the appropriate social and financial incentives to respond to the demands of parents and citizens.

There is a growing body of evidence to suggest that improving citizen's voice and accountability does not necessarily result in improved service delivery outcomes due to the complex range of contextual factors influencing provider response (Wild and Harris, 2011). For instance, local power dynamics in the education sector may mean that service providers are more strongly focused on responding to demands from political patrons or other government bodies than to demands than from citizens (Wild and Harris, ibid). Effects of interventions may also be mediated by parental human and social capital (Blimpo and Evans, 2011; Beasley and Huillery, 2012). Finally, CBM does not directly address underlying teacher or school quality issues (Bruns *et al.*, 2011), so the existing quality of schooling may influence the extent to which CBM can improve final education outcomes

Figure 8.2 a: CBM programme theory



8.2.2 Description of included studies

We included 11 studies[123] reported in 16 different papers that evaluated the effect of CBM intervention in L&MICs. These referred to nine unique programmes. We have used the term 'study' to refer to a unique output from an author team, which in some cases was reported in several papers. There were also occasions were we included several different studies that use the same dataset but had different author teams undertaking different analysis or reporting additional outcomes. For example, we included two different studies that evaluate the Newspaper campaign in Uganda and two studies that evaluate the AGEMAD initiative in Madagascar. In the following section, we describe the characteristics of these studies in detail. Table 8.2b summarises the characteristics of the included studies.

Population

Five of the nine programmes were targeted at public (government-run) schools only. The information campaigns in Brazil and Chile published data on the performance of both public and private schools, and the report card programme in Pakistan (Andrabi *et al.,* 2013) included data on all schools in each area, covering both public and privately run schools. The majority of the included programmes were targeted at the primary school level only (n=7), covering a range of ages and grades up to the end of primary school, although not all studies reported information on the age group covered. This is with the exception of the two programmes in Brazil (Camargo *et al.*, 2012) and Chile (Mizala & Urquiola, 2013), which published data on the performance of secondary schools[124]. Table 8.2b reports more detailed information about the students in the sample for each study. In most cases, grade and age is reported for children in the sample rather than across the programme itself.

Setting

The included CBM studies covered programmes in a broad range of settings in South Asia, Sub-Saharan Africa and Latin America and the Caribbean. Two of these programmes took place in India (Banerjee *et al.*, 2010; Pandey *et al.*, 2011), two in Uganda (Bjorkman, 2006; Reinikka & Svensson, 2007, Zeitlin *et al.*, 2012), and one in Chile (Mizala & Urquiola, 2013), Brazil (Camargo *et al.*, 2012), Madagascar (Lassibille *et al.*, 2010; Glewwe *et al.*, 2011), Kenya (Duflo *et al.*, 2012), and Pakistan (Andrabi *et al.*, 2013) respectively. Most of the studies took place primarily in rural areas (n=6), as reported by the authors. The cluster RCT reported in Pandey *et al.* (2011) consisted of an information campaign taking place separately in three different Indian states, Madhya Pradesh, Uttar Pradesh and Karnataka. The sample in the study in Brazil (ibid) was limited to the Sao Paulo Metropolitan area and the sample in Chile (ibid) to urban areas only, although both were nationwide schemes. The studies in Uganda looked at the effect in both urban and rural areas (ibid). In the study in Kenya, it was unclear whether the programme took place in urban or rural areas or both.

Intervention

All nine programmes included the key features of CBM interventions. That is, the provision of credible information on school performance, on resources for schooling and/or on the role of parents and communities in oversight of education, to motivate them to demand better education provision and to motivate schools to perform better (Bruns *et al.*, 2011). Table 8.2a summarises the intervention design characteristics.

---

123 Aturupane *et al.* (2014) has been included in the School Based Management chapter as it is evaluating a SBM programme in Sri Lanka. The authors also set out to assess the effect of a CBM programme, but this was never actually implemented and was therefore not included in our analysis for CBM interventions.
124 The programme in Chile also published data on performance of primary schools.

Type of information

Four of the programmes presented information on school and/or child performance. This information ranged from simple reporting of student and school level test scores through a report card (Andrabi *et al.*, 2013) to dissemination of information on outstanding schools, identified as such through the creation of a composite index of a range of school performance indicators and then grouped by 'homogenous schools', as in the SNED programme in Chile (Mizala & Urquiola, 2013). Two programmes provided information on resources available for schools (the AGEMAD initiative in Madagascar and the Newspaper campaign). The newspaper campaign in Uganda can be considered unique in this group of studies in that it announced the release of education grants for schools in order to reduce funding leakages from the education system through corruption. Four initiatives (as evaluated in Banerjee *et al.*, 2010, Zeitlin *et al.*, 2012, Pandey *et al.*, 2011 and Duflo *et al.*, 2012) provided information on roles and responsibilities of parents and school committees.

Campaign delivery

An important variation between programmes is the mechanism of information delivery. In this regard, the included programmes fall into either one of two categories[125]: relatively hands-off information provision at the national level and more direct provision of information at the community level. Three of the programmes were delivered at the national level, through publication and dissemination of data on school performance relative to other schools or school resources in national newspapers, on television and through parent associations (Camargo *et al.*, 2012; Mizala *et al.*, 2013; Reinikka & Svensson, 2007; Bjorkman, 2006).

The remaining six programmes provided information at the community or village level through group meetings or forums. Of these five, the programmes in India, Uganda, Madagascar and Pakistan used report cards or score cards to present information to parents (Banerjee *et al.*, 2010; Zeitlin *et al.*, 2012; Glewwe *et al.*, 2011; Lassibille *et al.*, 2010; Andrabi *et al.*, 2013).

Campaign intensity

Within those interventions that took a more direct approach at the community level, there was some variation in intervention intensity. For example, Banerjee *et al.*'s (2010) intervention involved the one-off delivery of information on roles and responsibilities to the community participants. At the other end of the spectrum is Pandey *et al.*'s (2011) study in India, which undertook three campaign rounds, with two or three meetings in each, presenting the same information to parents and school committee members on their role in oversight of teachers and schooling each time.

Capacity building

An additional element of some of these programmes was a capacity building component that trained communities in the use of monitoring tools (n=4), typically the production of a report or score card (Banerjee *et al.,* 2010, Pandey *et al.* 2011, Zeitlin *et al.*, 2012) or in the case of Kenya, how to monitor the attendance and performance of teachers (Duflo *et al.,* 2012).

[125] Although even within these categories there is significant variation

Table 8.2 a: Intervention Design Features of included studies

	Banerjee India	Mizala Chile	Zeitlin Uganda	Camargo Brazil	Andrabi Pakistan	Reinikka Uganda	Bjorkman Uganda	Glewwe Madagascar	Lassibille Madagascar	Pandey India	Duflo Kenya
Intervention key design features											
Information provision: school/child performance		✓		✓	✓			✓	✓		
Information provision: resources for schools						✓	✓	✓	✓		
Information provision: roles of community in monitoring	✓		✓							✓	✓
Capacity building in monitoring of services	✓		✓							✓	✓
Additional features	✓	✓						✓	✓	✓	✓
Information delivery mechanisms											·
Newspaper/television/ radio campaign		✓		✓		✓	✓				
Use of report cards	✓		✓		✓			✓	✓		
Group meetings	✓		✓		✓			✓	✓	✓	✓
Information delivery characteristics											
Implementation level	Community based	Nation wide	Community based	Nationwide	Community based	Nation wide	Nationwide	Community based	Community based	Community based	Community based
School type covered by intervention	Public	Private, public	Public	Private, public	Private, public	Public	Public	Public	Public	Public	Unclear

	Banerjee India	Mizala Chile	Zeitlin Uganda	Camargo Brazil	Andrabi Pakistan	Reinikka Uganda	Bjorkman Uganda	Glewwe Madagascar	Lassibille Madagascar	Pandey India	Duflo Kenya
Frequency of delivery	At least 1 meeting in each village	Every 2 years	Training at beginning of project. Scorecard visit once a term. One community meeting a term	Once a year	Report cards delivered once	Not clear – possibly monthly publication	Not clear – possibly monthly publication	At least two meetings between schools and community	At least two meetings between schools and community	Three rounds - 2/3 meetings in each round	One off training for school committee members
Content of school performance information	NA	Winner schools, based on an index of school performance	NA	School performance in exams	Raw school and child test scores	NA	NA	Enrolments, resource endowments (compared to other schools)	Enrolments, resource endowments and comparative data relative to other schools in these indicators	NA	NA

Comparisons

All but two of the included studies compared the effect of an intervention to business as usual (that is, a comparison group with no intervention). The two studies that evaluated the Newspaper campaign in Uganda (Reinikka & Svensson, 2007; Bjorkman, 2005) were different in this regard as they exploited variation in newspaper access across the country to artificially create a comparison group with less or no exposure to the campaign.

Outcomes

The included studies reported on a wide range of education outcomes. Ten of the eleven studies reported on some measure of achievement or learning, Mizala *et al.*'s (2013) study in Chile reporting findings on enrolment only. For the most part, achievement was measured through examination test scores in mathematics (n= 3), local language (n=3), official language (n=2) and composite scores (n=8). Two studies evaluated competencies in literacy and numeracy, for example, through tests of a child's ability to read stories, to read paragraphs, to recognise words or letters or less, to undertake addition and subtraction or not, and so on (Banerjee *et al.,* 2010; Pandey *et al.,* 2011). Some of these studies presented results separately for boys and girls (Camargo *et al.,* 2012 and Reinikka & Svensson, 2007 for composite test scores).

In addition to achievement, a number of studies also reported results for enrolment (n=5), attendance (n=3), dropout (n=3), and progression or completion (n=3). One study reported impacts on grade repetition. We also collected data on secondary outcomes of interest, teacher performance and teacher attendance. Three studies (n=4) reported findings for both of these outcomes (Banerjee *et al.*, 2010; Lassibille *et al.*, 2010; Pandey *et al.*, 2011).[126] Zeitlin *et al.*, (2012) reported outcomes for teacher attendance only. The measures of teacher performance were fairly heterogeneous; examples include the probability that teachers were actually teaching on an unannounced visit and whether or not a teacher was executing all tasks deemed essential for good classroom management.

Two studies measured outcomes after 12 months or less of intervention exposure (Banerjee *et al.*, 2010; Camargo *et al.*, 2012). Seven studies collected outcome data between 12 months and 48 months after the start of the intervention. The two studies evaluating the Newspaper Campaign in Uganda were the outliers in this regard; they reported outcome data on achievement approximately 72 months (six years) after the start of the government campaign (Reinikka & Svensson, 2007, Bjorkman, 2006).

Study Design

We identified a mix of experimental and quasi-experimental studies. Seven of the included studies[127] reported on cluster randomised control trials (Banerjee *et al.*, 2010; Pandey *et al.*, 2011; Andrabi *et al.*, 2013; Zeitlin *et al.*, 2012; Lassibille *et al.*, 2010; Glewwe *et al.*, 2011; Duflo *et al.*, 2012), where assignment to the intervention took place at the school, village, village cluster, or district level. Two studies used regression discontinuity design (Camargo *et al.*, 2012; Mizala & Urquiola, 2013). Camargo *et al.*'s study in Brazil used the government policy of only publishing test results for schools that had more than 10 students taking the national examination to create a regression discontinuity and compare achievement of

[126] The study in Kenya (Duflo *et al.*, 2012) presented results for teacher performance, teacher attendance and student attendance however the authors do not present the results compared to the comparison group with no treatment and therefore we were unable to include these outcomes in our meta-analysis.

[127] Assessing six unique programs.

schools just above and below the cut-off point. Mizala *et al.* (2013) exploited the design of Chilean government's SNED programme, which selects 'winner' schools based on an index of school quality indicators and which account for around 25 per cent of enrolment, to evaluate the impact of announcing this information. The final two studies, Reinikka & Svensson (2007) and Bjorkman (2006), made use of a natural experiment to evaluate the impact of a newspaper campaign in Uganda, using distance to a newspaper outlet and newspaper penetration in a district respectively to instrument for exposure to the campaign.

Qualitative studies, process evaluations and project documents

There was very limited qualitative and process information identified for the included programmes outside of the impact evaluations themselves and therefore much of the following synthesis is based on information provided in the impact evaluations. We identified one descriptive quantitative paper for Banerjee *et al.*'s evaluation in India and two project documents for the report card intervention in Pakistan and the scorecard experiment in Uganda respectively. Pandey *et al.*'s (2011) impact evaluation of community based-information campaigns in three states in India also included a significant qualitative component and we have drawn heavily on this paper for our analysis. A reason for this lack of additional process or qualitative information may be that seven of the included studies (covering six of the eight programmes) were cluster RCTs, in several cases evaluating a one-off trial. In general, there is less qualitative and process information available outside of the impact evaluation for experimental studies than for evaluations of government-run programmes.

Table 8.2 b: Characteristics of included studies CBM

Included study	Setting	Description of the programme	Included outcomes	Follow-up	Study design	Sample Size
Camargo (2012)	Brazil (urban). Secondary school Grade: 3 (secondary) Age: average age 18.43 years in public schools and 17.4 years in private schools.	Disclosure of the National Secondary Education Examination (ENEM): This programme of the Brazilian Ministry of Education releases schools' average scores on the ENEM for the previous year to function as an indicator of school quality, with the aim of improving teaching and to identify shortcomings. ENEM scores are standardised and publicised by major newspapers and websites.	Composite test scores	Approx. 12 months	Regression Discontinuity Design	3074 students
Mizala & Urquiola (2013)	Chile (urban). Primary school; Secondary school Grade: 1 and 9 Age: not reported	Subsidized School Performance Evaluation System (SNED): The SNED programme is an initiative introduced by the Chilean government to improve academic performance via an economic incentive for teachers of best performing schools every two years. The government introduced a system that identifies outstanding public or subsidised private schools via a system of homogenous groups. This study evaluates the effect of being identified as a winning/losing school in the education market (rather than the teacher bonus).	Enrolment	Approx. 24 months and 48 months.	Regression Discontinuity Design	5750 schools at 24 month follow up. 4494 schools at 48 month follow up.

Included study	Setting	Description of the programme	Included outcomes	Follow-up	Study design	Sample Size
Banerjee *et al.* (2010)	India (rural). Primary school. Grade: not reported Age: 7-14 years	Encouraging Participation in Sarva Shiksha Abhiyan: This experiment introduced three initiatives to encourage community participation in education: The first provided information about the structure and organisation of Village Education Committee (VEC)s. The second intervention was provided alongside the first in a sub-set of villages and shared information about the status of student's learning in the villages, taught villagers how to generate their own reading report cards, and transferred a specific monitoring tool to the community. The third intervention supplemented the first and second interventions in a further sub-set of villages, introducing local volunteers to a simple technique for teaching children how to read, using pedagogy from Pratham's flagship "Read India" programme.	Enrolment; attendance; maths test scores; language arts test scores	6 - 8 months	Cluster RCT	17,419 students (across all treatment groups). 316 schools
Pandey *et al.* (2011)	India (rural). Primary school Grade: ranged between 2 and 4 across states Age: Uttar Pradesh: average age of 8.75 years. Madhya Pradesh: average age of 8.9 Karnataka: average age of 10 years	Community based-information campaign: The experiment consisted of an information campaign in three Indian states (Madhya Pradesh, Uttar Pradesh, Karnataka) to inform parents and school committees of their oversight roles and responsibilities in education, of the education services that they are entitled to and the minimum levels of language arts and mathematics skills that children were expected to acquire by grade. In addition, there was an additional treatment carried out only in Karnataka state where a film was shown on the economic benefits of schooling.	Maths test scores; language arts test scores	MP and UP: follow up after 14 months and 24 months. Karnataka: 1st after 12 months, and 24 months.	Cluster RCT	Approximately 1195 schools

Included study	Setting	Description of the programme	Included outcomes	Follow-up	Study design	Sample Size
Glewwe & Maiga (2011)	Madagascar (urban, peri-urban, rural). Primary school. Grades: 3-4 Age: not reported	AGEMAD Initiative: The AGEMAD initiative provided actors in the primary education administrative system at the district and sub-district level with tools and training to help manage their schools, as well as village communities with information on their school's performance. Accountability meetings were held for parents to discuss the performance and contribute to school improvement plans. This was a nested experimental design: the first group received only a district level intervention, the second treatment group received a district and sub-district level intervention, and the third group received interventions at all three levels.	Composite test scores	20 months	Cluster RCT	606 schools
Lassibille *et al.* (2010)	Madagascar (urban, peri-urban, rural). Primary school. Grades: 3-5 Age: not reported	AGEMAD Initiative: The AGEMAD initiative provided actors in the primary education administrative system at the district and sub-district level with tools and training to help manage their schools, as well as village communities with information on their school's performance. Accountability meetings were held for parents to discuss the performance and contribute to school improvement plans. This was a nested experimental design: the first group received only a district level intervention, the second treatment group received a district and sub-district level intervention, and the third group received interventions at all three levels.	Maths test scores; language arts test scores; attendance; composite test scores; dropout; Completion	Student outcomes: 20 months Teacher outcomes: approx. 15 - 20 months	Cluster RCT	606 schools
Andrabi *et al.* (2013)	Pakistan (rural). Primary school Grades: 1-5 Age: average age in the sample is 9.7 years	Village level report card intervention: The experiment consisted of the introduction of report cards containing school and child level test scores to households in rural villages that contained both public and private schools. Card 1 reported the score of the child in English, maths and Urdu, including quintile rank across all tested children and average score for the child's village and school. Card 2 reported scores for all schools in the village.	Maths test scores; language arts test scores; composite test scores; enrolment; dropout	Approx. 12 months and 24 months (after 1 school year and 2 school years).	Cluster RCT	9887 students (112 villages)

Included study	Setting	Description of the programme	Included outcomes	Follow-up	Study design	Sample Size
Bjorkman (2006)	Uganda (urban, peri-urban, rural). Primary school Grades: children in grade 7 Age: not reported	Newspaper information campaign for Education: The Government of Uganda initiated an information campaign in 1997 to systematically publish data in national newspapers on the monthly transfers of education grants to districts, on school entitlements and responsibilities, and stories on misuse of education grant funds. The aim was that parents and head teachers could monitor local administration and voice complaints if funds did not reach schools.	Composite test scores	Approx. 72 months	Natural experiment – using DID	53 schools
Reinikka & Svensson (2007)	Uganda (urban, peri-urban, rural). Primary school Grades: 1- 7 Age: not reported	Newspaper information campaign for Education: The Government of Uganda initiated an information campaign in 1997 to systematically publish data in national newspapers on the monthly transfers of education grants to districts, on school entitlements and responsibilities, and stories on misuse of education grant funds. The aim was that parents and head teachers could monitor local administration and voice complaints if funds did not reach schools.	Enrolment; composite test scores	Approx. 72 months	Natural experiment – using IV regression	374 schools
Zeitlin *et al.* (2012)	Uganda (rural). Primary school Grades: 3 at baseline and 5 at end line Age: not reported	School scorecard for community-based monitoring: The experiment introduced the use of score cards to School Management Committees (SMC) so that they were better equipped to monitor performance of schools and students. SMC members were trained in the use of the score card, and then collected data on outcomes in the scorecard on a termly basis to make targets and plans for improvements. Two variations were tested: a standardised scorecard that incorporated best practices for indicators of school performance, and a participatory scorecard, which allowed SMC's to develop their own indicators for school performance.	Completion; attendance; composite test scores; enrolment	14 months	Cluster RCT	100 schools (approximately 3512 students

Included study	Setting	Description of the programme	Included outcomes	Follow-up	Study design	Sample Size
Duflo *et al.* (2012)	Kenya (rural/urban location unclear). Primary school Grades: grade 1 at baseline Age: average age of 7.7 years at baseline	Extra Teacher Programme (ETP): The Extra Teacher Programme was an experiment with four different treatment arms, testing combinations of giving school committees grants to hire an extra contract teacher, providing school-based management training to parents and tracking children by ability. This study refers to the treatment arm that gave grants to schools to hire a contract teacher + provided SBM training (but did not track students by ability). Parent Teacher Association members and parents were encouraged to supervise recruitment of the extra-teacher, taught how to undertake interviews and taught techniques for checking teacher attendance.	Dropout; completion; language arts test scores; maths test scores; composite test scores	Achievement outcomes: 15 months and 27 months Dropout and Completion: 24 months after programme	Cluster RCT	Approximately 1566 students

8.2.3 Synthesis of findings

The results of our synthesis are presented in two sections. First, we have presented the findings of the meta-analysis on the effects of CBM on primary and secondary outcomes, including results reported separately according to different population sub-groups where we identified sufficient data. This is followed by a discussion of the findings incorporating evidence from our descriptive qualitative synthesis to assess factors related to intervention design, implementation and context which might act as barriers or facilitators of the effectiveness of CBM (Questions 2a and 2b).

Effects of CBM interventions on enrolment, attendance, dropout rates, completion and learning outcomes

This section reports the results of the meta-analysis of the effects of CBM interventions, addressing question 1a of the review. We have structured the presentation of results according to the causal chain outlined in the programme theory of change, starting with education access outcomes (enrolment, attendance, dropout) and final outcomes (completion, learning outcomes: composite test scores, language arts test scores, maths test scores). All eleven studies provided data for the meta-analysis, but none of the studies reported on all outcomes. The number of comparisons with effect sizes ranges from three for dropout and completion, up to 12 for enrolment.

As previously described, several of the studies contained multiple treatment arms testing different CBM initiatives. In all meta-analyses, we included the treatment arm that tested an intervention that was most similar to the other included programmes. For Zeitlin *et al.*'s (2012) evaluation in Uganda, we chose to include the treatment arm that introduced a school scorecard covering standard indicators of school effectiveness in the meta-analysis over a scorecard developed in participation with the community, however effect sizes for both are presented. For Banerjee *et al.*'s (2010) study in India, we chose to include the second treatment arm in all relevant meta-analysis that included both information on roles and responsibilities in oversight of education plus teaching for villagers on how to evaluate a child using a simple testing instrument in order to produce report cards. We have presented effect sizes for the other two treatment arms in the narrative discussion where different from the arm included in the meta-analysis and in full in Appendix H. Pandey *et al.*'s (2011) evaluation of information campaigns in three Indian states (Madhya Pradesh, Uttar Pradesh and Karnataka) reports results separately by state, and we were able to include results for all three in the meta-analyses for maths and language arts.

For Glewwe *et al.* (2011) and Lassibille *et al.'s* (2010) evaluations of the AGEMAD initiative in Madagascar, we have only presented information on the third treatment arm as this was the only arm that included community-based monitoring of education. We took results for composite scores from Glewwe *et al.* (2011) as it was the most recent paper, but data for all other outcomes is taken from Lassibille *et al.* (2010) and associated papers. Camargo's (2012) study in Brazil reported effects on composite test scores for both public and private schools. We included both sub-samples in our meta-analysis for this outcome.

Two studies reported outcome data on composite test scores for the Newspaper campaign in Uganda (Bjorkman, 2006; Reinikka & Svensson, 2007). We chose the study that used an analysis method most similar to the other included studies that reported outcome data (Bjorkman, 2006). Reinikka & Svensson (2007) reported impacts of the campaign on enrolment by grade sub-groups only and so we included all seven effect sizes for each group in the meta-analysis for student enrolment as they can be considered independent samples.

In cases where numerous measures of language arts, mathematics or cognitive scores were presented (for example reading and spelling for language arts), we combined the measures to create one synthetic effect size. In the case of studies reporting multiple language arts outcomes, for example French and Malagasy, we chose the official language or the local language spoken by the most number of children in the sample to include in the meta-analysis but also present effect sizes for the other outcomes.

Where multiple follow up periods were reported, we included the follow up that was most similar to other studies in the meta-analysis for that outcome. For example, Mizala & Urquiola (2013) report results on enrolment at approximately 24 months and 48 months after the programme began but we chose to include the 24 month follow up period in the meta-analysis as this was closer to the follow up period in the other six programmes.

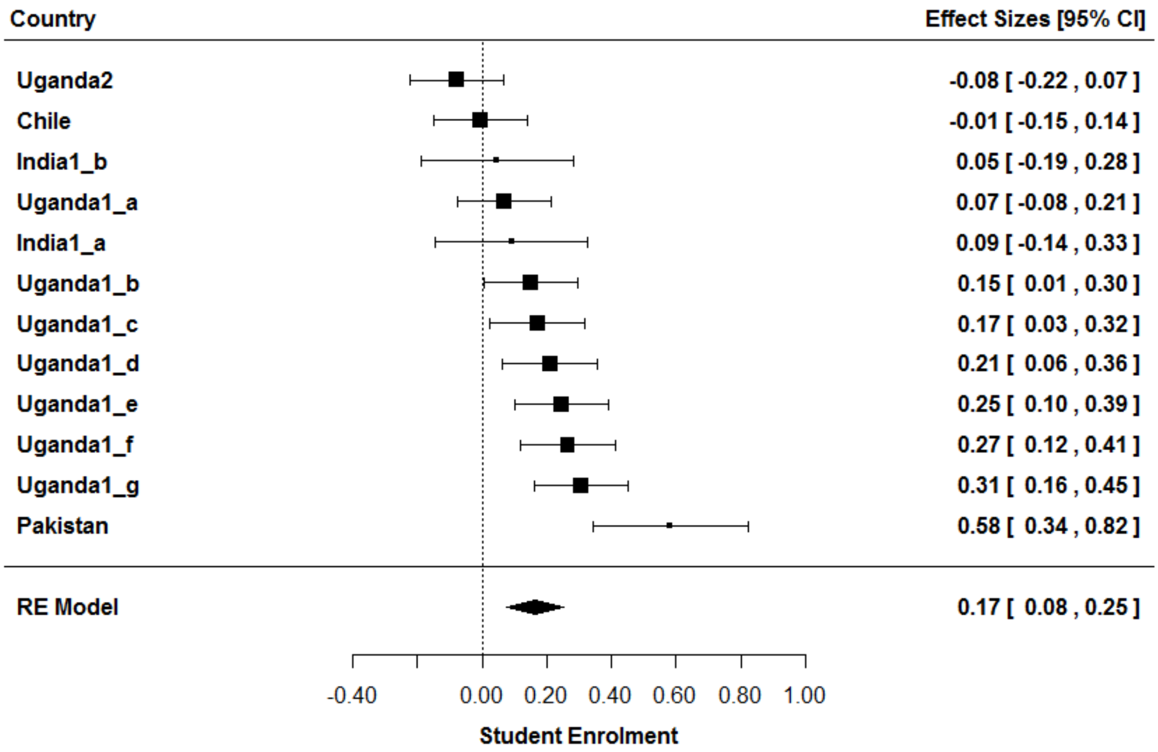
All effect sizes are expressed as standardised mean difference (SMD), interpreted as the magnitude of the number of standard deviation changes in the outcome for the intervention group as compared to students in non-SBM schools. SMD scores are interpreted as the number of standard deviation changes in the outcome.

Enrolment

Figure 8.2b presents the forest plot with the results of the individual community based monitoring (CBM) studies and the pooled point estimate on student enrolment in school. The overall average effect of CBM initiatives on enrolment is 0.17, with a 95% confidence interval (CI) [0.08, 0.25], calculated under a random effects model. This is statistically significant. The assessment of heterogeneity suggests a modest amount of variability between studies (I^2 = 72.52%, τ^2 = 0.0174, Q(df = 11) = 37.6374, p-val < .0001). Sensitivity analysis suggests the overall pooled effect is mostly robust to removal of any studies, with the average effect moving slightly between 0.15 and 0.19 (see Appendix H for results of all sensitivity analyses). The exception is the removal of the observation from Pakistan from the analysis, which causes the point estimate to drop to 0.13 (95% CI, [0.06, 0.21]) and the I^2 to drop to 60.2%. It can be seen from the forest plot that this study is an outlier in terms of the magnitude of the positive effect.

It should be noted that a large number of the observations in this meta-analysis come from the same programme in Uganda (Uganda1), the government newspaper campaign to publish data on education grants. These correspond to the effect on individual primary school grades, ranging from grade 1 (Uganda1_a: 0.07, 95% CI, [-0.08, 0.21]) to grade 7 (Uganda1_g: 0.31, 95% CI [0.16, 0.45]). It can be seen that the positive impact on enrolment increases by grade level.

Figure 8.2 b: Enrolment[128]



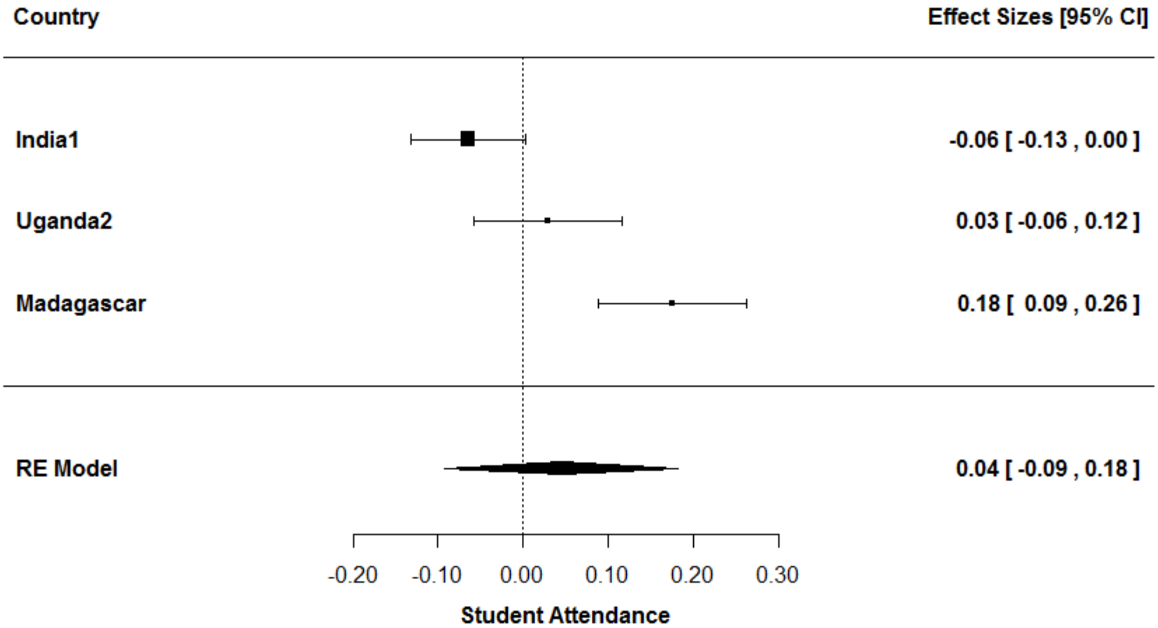
We did not have a sufficient number of studies to conduct a meta-analysis of sub-groups of children, for example by gender, age or grade. However, Appendix H reports the individual effect sizes for sub-groups available in the included studies. The one finding that differs substantively from the results for the main sample is the negative impact on boy's enrolment of -0.50 (95% CI, [-0.73, -0.26]) reported for one of the treatment arms in the study by Banerjee *et al.* (2010). The treatment provided parents with a simple pedagogy tool to teach their children alongside the usual CBM interventions such as report cards and an information campaign for parents.

Attendance

The overall weighted average effect of CBM on student attendance at school was 0.04, (95% CI [-0.09, 0.18]), calculated under a random effects model. Figure 8.2c suggests the presence of large heterogeneity of effects, which is also suggested by the assessment of homogeneity (I^2 value of 88.69%, $\tau^2 = 0.0131$, Q(df = 2) = 18.27, $p<0.0001$). Effect sizes range from -0.06, 95% CI, [-0.13, 0.00] in India (Banerjee et al., 2010) to 0.18 (95% CI [0.09, 0.26]) in Madagascar (Lassibille *et al.,* 2010). The results are sensitive to the removal of each study (see Appendix H for results of all sensitivity analyses). The point estimate increases to 0.10 (95% CI, [-0.05, 0.25]) on the removal of India1, and reduces to -0.02 (95% CI [-0.11, 0.07]) on the removal of the evaluation of the AGEMAD initiative in Madagascar.

[128] Studies labelled Uganda1 refer to results from the Newspaper Campaign in Uganda (Reinikka & Svensson, 2007) and studies labelled Uganda2 to results from the School scorecard for community-based monitoring trial in Uganda (Zeitlin *et al.*, 2012). Uganda1_a refers to effects for grade 1 students only in the sample, Uganda_1b refers to grade 2 students only, Uganda1_c for grade 3 students only, Uganda1_d for grade 4 students only, Uganda1_e to results for grade 5 students only, Uganda1_f to grade 6 students only and Uganda1_g to grade 7 students only. India1_a refers to effects for boys from the Encouraging Participation in Sarva Shiksha Abhiyan trial in India (Banerjee *et al.,* 2010) and India1_b to effects for girls from the same study.

Figure 8.2 c: Student Attendance[129]



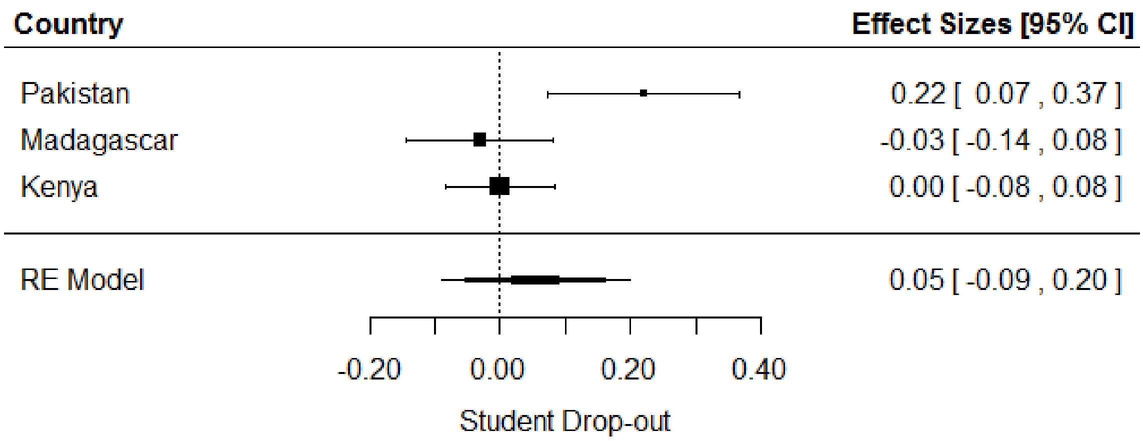
We did not have enough studies to undertake a meta-analysis of any sub-groups. However, Appendix H reports the individual effect sizes for sub-groups reported in the included studies. One finding of note here is that the alternative treatment arm of the study from Uganda (Uganda2) report an effect that is much larger in magnitude than the one we included in our meta-analysis (0.20, 95% CI [0.11, 0.29]). This treatment arm trained community members in the use of a school scorecard that used indicators of school effectiveness developed in participation with the community, compared to the treatment arm in the meta-analysis that introduced a school scorecard covering standard indicators of school effectiveness.

Dropout

The overall average effect of CBM initiatives on drop-out is 0.05, 95% CI [-0.09, 0.20], noting that a positive sign for an effect size for dropout can be interpreted as an increase in drop-out. The assessment of homogeneity suggests a large amount of between-studies variation (I^2 = 79.84%, τ^2 = 0.0129,Q(df = 2) = 8.0685, p-value = 0.0177). Results are sensitive to the inclusion of the study from Pakistan (Andrabi *et al.,* 2013); when we removed this study from the analysis, the point estimate changed to -0.01 (95% CI [-0.08, 0.06]).

[129] India1_a refers to results from the Encouraging Participation in Sarva Shiksha Abhiyan trial in India (Banerjee *et al.,* 2010). Uganda2 to results from the School scorecard for community-based monitoring trial in Uganda (Zeitlin *et al.,* 2012).

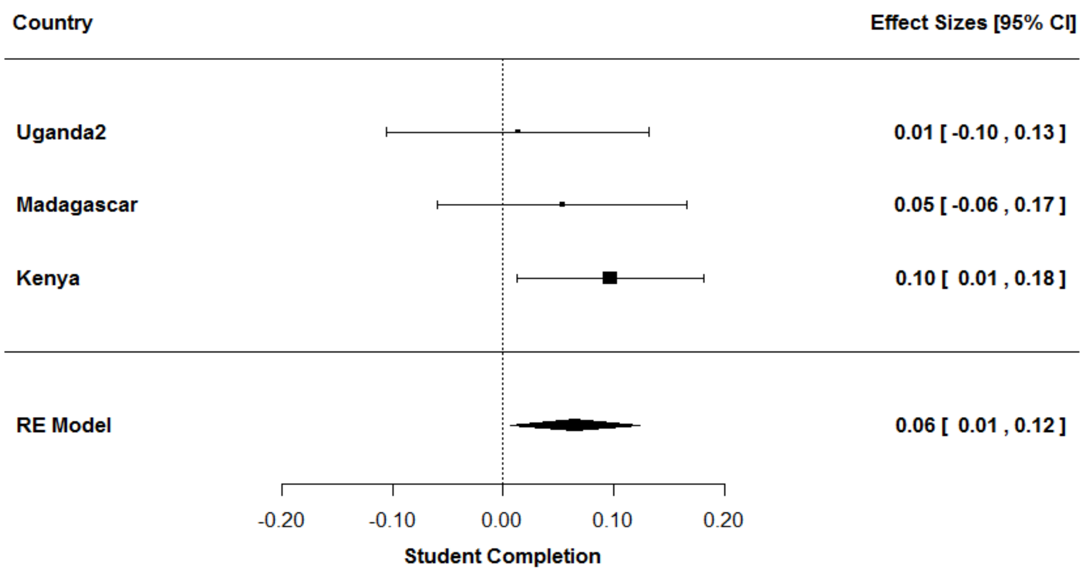
Figure 8.2 d: Student drop-out



Completion

The three included studies for this outcome reported on some measure of progression to the next grade, and in the case of Madagascar, the pass rate on the test for progression from primary school to secondary school. The overall average effect of CBM initiatives on these outcomes is 0.06, 95% CI [0.01, 0.12], calculated under a fixed effects model. The forest plot for completion can be seen in figure 8.2e. The assessment of homogeneity suggests that between-study variability is low (I^2 = 0%, τ^2 = 0.00 Q (df = 2) = Q(df = 2) = 1.3181, p-val = 0.5174). Sensitivity analysis indicates that removing any one of these studies does not make a substantive difference to the overall pooled effect (see Appendix H for results of all sensitivity analyses).

Figure 8.2 e: School Completion[130]



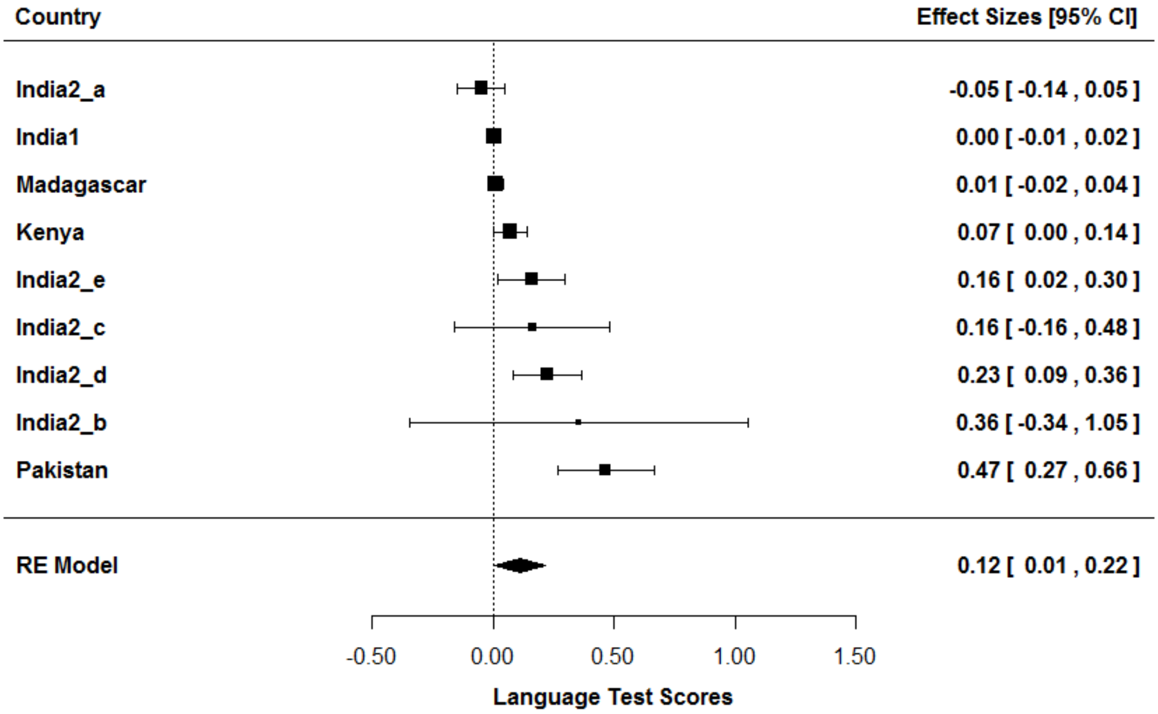
Language arts test scores

The overall average effect of SBM on language arts test scores is 0.12 (SMD= 0.12, 95% CI [0.01, 0.22]). The tests for homogeneity indicate a large amount of between study variability

[130] Uganda2 to results from the School scorecard for community-based monitoring trial in Uganda (Zeitlin et al., 2012).

(I^2 = 95.11%, τ^2 = 0.0183, Q(df = 8) = 39.70, p-val= < .0001), indicating that the results are not consistent across studies. This can be seen visually in the forest plot in figure 8.2f, where effect sizes vary between -0.05 (95% CI [-0.14, 0.05]) for India2_a [131](Pandey *et al.*, 2011) and 0.47 (95% CI [0.27, 0.66]) in Pakistan (Andrabi *et al.*, 2013). Studies labelled India2 evaluate the same community based-information campaign, implemented in different states in India and for different grade samples (reported in Pandey *et al.*, 2011). There is some variation in effects even within the same intervention design, although for most observations the confidence intervals are overlapping. Removing the study from Pakistan, reduces the point estimate to 0.06, 95% CI [-0.0060, 0.1260] (see Appendix H for results of all sensitivity analyses).

Figure 8.2 f: Language Arts Test Scores[132]



We did not have enough studies to undertake meta-analysis for any sub-groups. However, Appendix H reports the individual effect sizes for sub-groups and other treatment arms that we identified in included studies.

Maths test scores

The overall average effect of CBM initiatives on mathematics test scores is 0.09 (95% CI, [-0.02, 0.20]), calculated using a random effects model. As with the results for language arts scores, there is a large amount of variability between studies, as measured by the tests for homogeneity (I^2 = 96.51%, τ^2 = 0.0224, Q(df = 8) = 66.3819, p-val < .0001). This can also be

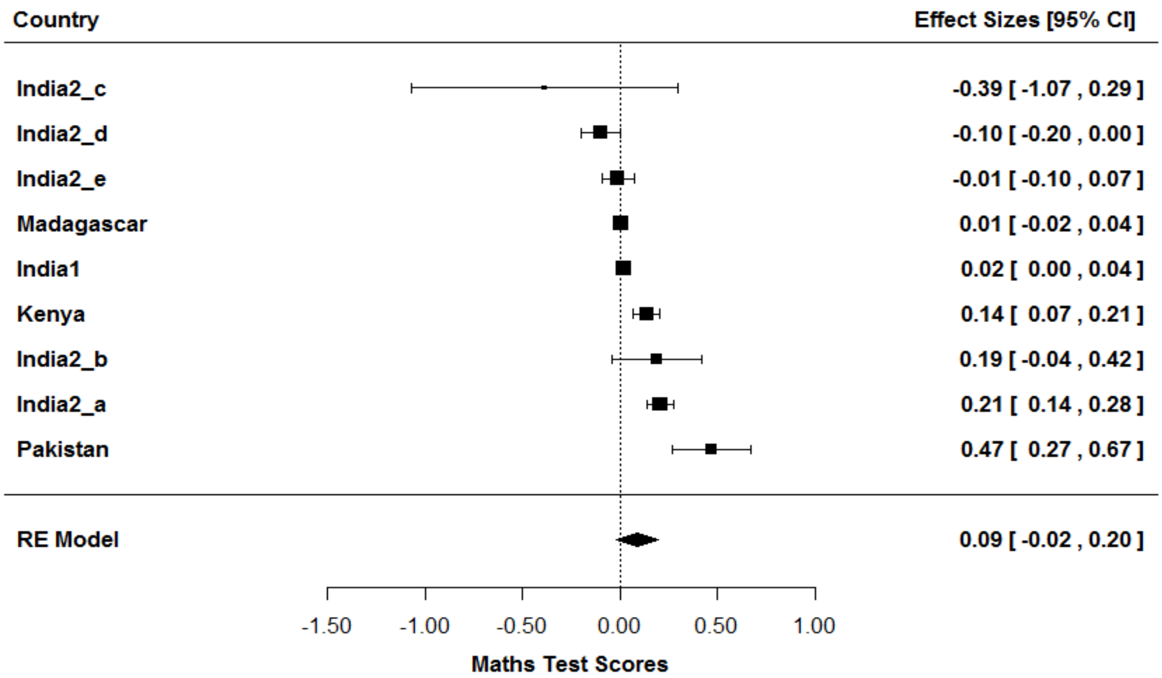
[131] Studies labeled India2 evaluate the same intervention, implemented in three different states in India and for different grade samples (Pandey *et al.,* 2011). India2_a refers to students from the trial in Karnataka, India2_b refers to grade 5 students from Uttar Pradesh, India2_c refers to grade 3 and 4 students from the trial in Uttar Pradesh, India2_d refers to grade 5 students from Madhya Pradesh and India2_e refers to grade 3 and 4 from the trial in Madhya Pradesh.

[132] Studies labeled India1 refer to results from the Encouraging Participation in Sarva Shiksha Abhiyan trial in India (Banerjee *et al.,* 2010). Studies labeled India2 evaluate the same intervention, implemented in three different states in India and for different grade samples (Pandey *et al.,* 2011). India2_a refers to students from the trial in Karnataka, India2_b refers to grade 5 students from Uttar Pradesh, India2_c refers to grade 3 and 4 students from the trial in Uttar Pradesh, India2_d refers to grade 5 students from Madhya Pradesh and India2_e refers to grade 3 and 4 from the trial in Madhya Pradesh.

seen visually in the forest plot in figure 8.2h. Effect sizes range between -0.39 (95% CI [-1.07, 0.29]) for India2_c (Pandey et al, 2011) and 0.47 (95% CI [0.27, 0.67]) in Pakistan (Andrabi *et al.*, 2013).

The sensitivity analysis suggests the results are sensitive to the inclusion of study outliers (see Appendix H for results of all sensitivity analyses). When the study is removed from Pakistan (ibid) from the analysis the point estimate changes to 0.05 (95% CI, [-0.03, 0.12]), although heterogeneity remains high.

Figure 8.2 g: Maths test scores[133]



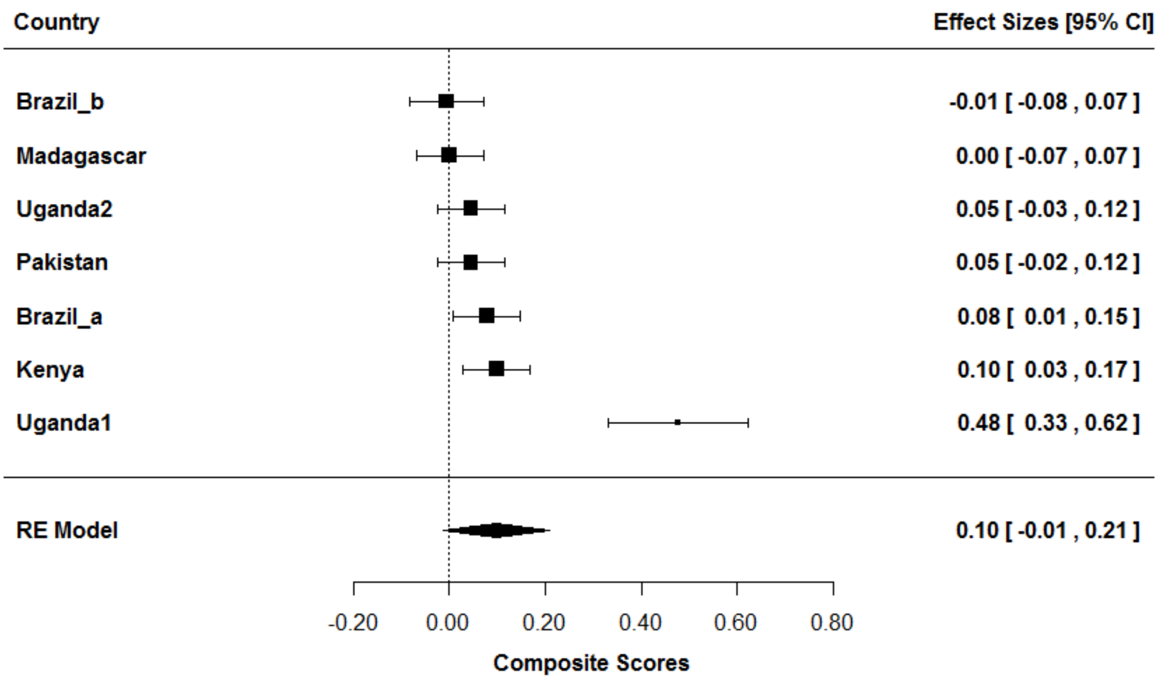
Composite test scores

The overall average effect of CBM on learning outcomes as measured by a composite test score is 0.10, 95% CI [-0.01, 0.21]. The test scores appear from the forest plot in Figure 8.2h to be more homogeneous, with all but one of the studies having overlapping confidence intervals. However, the assessment of homogeneity suggested a large degree of variability across studies (I^2 = 93.13%, τ^2 = 0.0204, Q (df=6)= 38.84, p-val= < .0001).

The study from Uganda appears to be an outlier (Bjorkman, 2005) and the confidence intervals for this study do not overlap with the other six studies in the meta-analysis. This is supported by the sensitivity analysis; when we remove this study from the analysis, the point estimate is reduced to 0.04 (95% [0.01, 0.07]) and the study variability is reduced (I^2=20.59%, $\tau^{2=}$ 0.0204).

[133] Studies labelled India1 refer to results from the Encouraging Participation in Sarva Shiksha Abhiyan trial in India (Banerjee *et al.*, 2010). Studies labelled India2 evaluate the same intervention, implemented in three different states in India and for different grade samples (Pandey *et al.*, 2011). India2_a refers to students from the trial in Karnataka, India2_b refers to grade 5 students from Uttar Pradesh, India2_c refers to grade 3 and 4 students from the trial in Uttar Pradesh, India2_d refers to grade 5 students from Madhya Pradesh and India2_e refers to grade 3 and 4 from the trial in Madhya Pradesh.

Figure 8.2 h: Composite Test Scores[134]



8.2.4 Summary of findings and discussion

We identified eleven studies of CBM initiatives across seven countries in South Asia, Sub-Saharan Africa and Latin America and the Caribbean. We were able to examine effects on enrolment, attendance, drop-out, completion, maths, language arts and composite test scores using meta-analysis.

The overall average pooled effects range from 0.04 for student attendance (95% confidence interval [-0.9, 0.18]) to 0.17 for student enrolment (95% CI [0.08, 0.25]). Thus, the meta-analysis findings suggest that overall the studied CBM initiatives had beneficial effects on school participation, as measured by enrolment, attendance, dropout and completion. However, this is based on a relatively small number of studies and there was significant heterogeneity in results for all participation outcomes. The combined average effect of CBM initiatives on learning outcomes range from 0.09 SMD for maths test scores (95% CI, [-0.02, 0.2]) to 0.10 SMD (95% CI [-0.01, 0.21]) for composite test scores and 0.12 (95% CI [0.01, 0.22]) for language arts test scores. There is however significant heterogeneity within the outcomes across programmes, and particularly for maths test scores.

We conducted a qualitative synthesis of evidence on process and implementation to try to explore reasons for heterogeneity. The findings from this synthesis that may explain the observed effects, and in particular heterogeneity of effects across studies are discussed below (section A8.2 in appendix J provides the results of the full qualitative synthesis, tables 8.2c and 8.2d provide the summary findings, together with details about the context for which they apply).

[134] Studies labelled Uganda1 in this meta-analysis refer to results from the Newspaper Campaign in Uganda (Bjorkman, 2006). Studies labelled Brazil_a in this meta-analysis refer to results for composite test scores for private schools and studies labelled Brazil_b refer to results for public schools.

Intervention design

The report card experiment in Pakistan (Andrabi *et al.*, 2013) is a consistent outlier in terms of the relatively large positive effects reported for enrolment and learning outcomes. The main difference between the intervention assessed in this study and the other studies is the content of the information provided: the programme presented raw test scores of children and schools in a defined 'school market' - schools that would be considered as viable options for a parent living in a given village. The authors note they chose this simpler information as it was more difficult to explain 'value-added' scores to both parents and schools. It included information about private schools in this school market as well as public primary schools. The authors found large improvements in student test scores for children in initially poorly performing private schools, with only very small improvements in public schools (Andrabi, *ibid.*), and present data to suggest that private schools faced stronger incentives to change their investments.

Bjorkman's (2005) evaluation of the newspaper campaign to reduce corruption in Uganda is another study where we observed relatively large effects on composite test scores as compared to the other studies in the meta-analysis. In this case the intervention design is very different to the other included programmes and this may explain the observed heterogeneity. After a public expenditure tracking survey undertaken in 1995 had revealed that on average only 24 per cent of the education grant from the central government was reaching intended schools, the government initiated a newspaper campaign in 1997 to publish data in national newspapers on the monthly transfers of education grants to districts and on school entitlements. In 2001, the median received by schools was 82 per cent of the entitlement (Reinikka & Svensson, 2007). Such a huge increase in the amount of money reaching schools may help to explain such a large positive effect on achievement. The follow up period for this study is also longer than the other included studies; approximately 6 years compared to many of the other included studies that cover a one to two year period or less.

Intermediate outcomes

The three studies that reported on uptake of the intervention found that community meetings to disseminate information were well attended by the community (Banerjee *et al.*, 2010; Pandey *et al.*, 2011; Lassibille *et al.*, 2011). A key intermediate stage in the theory of change is that once parents are empowered with the knowledge and tools to hold providers accountable, citizens decide to take action and participate collectively in monitoring activities, for example joining school or village education committees, or participating in school visits or parents meetings. The most commonly reported finding in our analysis of process and implementation is that parents' participation in school management or monitoring did not increase following the CBM initiative, although only four of the eleven studies reported on this outcome. The four studies that reported this finding on participation all assessed similar intervention designs using similar information delivery mechanisms; namely group meetings with the community to present information on the role of parents in education or oversight or on the performance of their children/schools (Banerjee *et al.*, 2010; Zeitlin *et al.*, 2012; Nguyen and Lassibille (2008); Pandey *et al.*, 2011). The two CBM studies in India also found that there was limited or no change in parents' knowledge of community education institutions following the information campaign. It is perhaps not surprising then that several of these studies found limited effects on education outcomes and in particular test scores.

Contextual factors

Our descriptive qualitative analysis suggest community human capital, teacher responsiveness and extent of power of school committees may be contextual factors influencing intervention effectiveness across different contexts for the included studies, as reported in table 8.2d. High rates of illiteracy among parents was a factor reported in several of the included studies, however the potential role in influencing intervention effects does not appear to be consistent across contexts. Pandey *et al.*,'s evaluation of an information campaign in three states in India found greater effects on student achievement in villages with low literacy rates. They explain this as being because villages with more illiterate parents should have a greater demand for schooling. Conversely, Lassibille *et al.* (2013) suggests that wealthier and more literate parents are better able to use the information provided by the report cards, and, presumably, better able to monitor school activities.

Pandey *et al.* (2011) and Banerjee *et al.* (2010) suggest the structure and power of the organisations involved in monitoring of school activities may contribute to the success (or failure) in improving participation, specifically committee size and length of term. In the state of Uttar Pradesh, where Parent- Teacher Associations (PTAs) are small village level bodies elected for 5 years, there were greater improvement in participation and awareness than in Madhya Pradesh, where committees are at the school-level, relatively large in size and elected for a year only (Pandey *et al.,* 2011). Additionally, in Madhya Pradesh school committees had to verify teacher presence for a teacher to receive her salary, while in Uttar Pradesh committees only control the tenure of contract teachers, while in Karnataka committees have neither of these powers. The authors suggest this may explain the improvement in teacher effort indicators in Madhya Pradesh but none in Uttar Pradesh and Karnataka.

The only study that asked participants about their opinion on perceived barriers to the effectiveness of the CBM intervention found that a lack of responsiveness of the teachers was a commonly reported barrier (Pandey *et al.,* 2011). A key assumption at the last stage of the theory of change is that education providers are responsive to increased parental demand. Parents reported that if they raised concerns regarding their children's' learning this was frequently met with a negative or angry response. Alongside this qualitative evidence, authors in three of the included impact evaluations (Pandey *et al.*, 2011; Andrabi *et al.*, 2013; Zeitlin *et al.*, 2012) discuss anecdotally poor teacher incentives structures in these contexts that limit their responsiveness to parent demands, and thus the power of accountability mechanisms. This may go some way to explain the lack of improvement in learning outcomes observed in many contexts.

In summary, CBM initiatives appear to have a beneficial effect on all outcomes, but the average effect is relatively small in magnitude for some outcomes, in particular for attendance, dropout and completion. The magnitude of effect on enrolment and learning outcomes is slightly larger. However, there is considerable heterogeneity across studies and the presence of outliers indicates that CBM programmes may have larger effects in some settings. The qualitative discussion suggests community human capital, lack of teacher responsiveness and extent of power of parent teacher associations may be some of the factors that influence the effectiveness of community monitoring initiatives in education.

Table 8.2 c: Descriptive findings: CBM process and implementation

Descriptive findings: Process and implementation	Context	Citation/ info type
Uptake of the interventions		
Strong attendance/interest from the community in CBM meetings, including among minority groups	India (two studies), Madagascar	Banerjee *et al.*; Pandey *et al.* 2011; Lassibille *et al.* 2011. *Impact evaluations*
Good uptake by parents of tool to assess child learning in India	India	Pandey *et al.* 2011. *Impact evaluation/qualitative*
Schools or the community followed up on information provided as part of the intervention	India, Chile	Pandey *et al.* 2011; Mizala & Urquiola, 2013. *Impact evaluations*
Parents' knowledge following information campaigns		
Limited improvement in parental and/or education committee knowledge of monitoring institutions following information campaigns	India (two studies)	Banerjee *et al.*, 2010; Pandey *et al.*, 2011 *Impact evaluation*
Small increase in parents and school committee knowledge of the status of education in their village after the intervention compared to the control	India, Pakistan	Banerjee *et al.* 2010; Andrabi *et al.* 2013. *Impact evaluations*
Parent and school committee participation in school oversight and management		
Parental participation in schools did not increase as a result of the CBM intervention	India (two studies), Uganda, Madagascar	Banerjee *et al.* 2010; Zeitlin *et al.* 2012; Nguyen and Lassibille, 2008; Pandey *et al.* 2011. *Impact evaluations*
Minority groups excluded from using information provided to them as part of CBM initiatives	India	Pandey *et al.* 2011. *Impact evaluation/qualitative*
Parent response to information campaigns: switching schools		
Parents did not switch their children into better quality schools as a result of improved information about school quality.	India, Pakistan, Brazil	Banerjee *et al.,* 2010; Andrabi *et al.,* 2013; Camargo *et al.,* 2012 *Impact evaluations*
Education sector response to information campaigns		
Substantial changes in private school fees as a result of more information in Pakistan	Pakistan	Andrabi *et al.* 2013 *Impact evaluation*

Information campaign reduced leakage of funds from the education system but did not eliminate it in Uganda	Uganda	Reinikka & Svensson, 2007 *Impact evaluation*

Table 8.2 d: Descriptive findings: Contextual factors

Descriptive findings: Contextual factors	Context	Citation (info type)
Teacher incentive structures may limit the effectiveness of CBM initiatives	India, Uganda, Pakistan	Pandey *et al.* 2011 (*impact evaluation/qualitative*); Andrabi *et al.* 2013; Zeitlin *et al.* 2012: *Impact evaluations*
Extent of power of the school committees may play a role in determining the effectiveness of CBM initiatives	India (two studies	Pandey *et al.* 2011; Banerjee *et al.* 2010 *Impact evaluations*
Responsiveness of the education provider to demands for better education may vary by school provider type	Brazil, Pakistan	Camargo *et al.* 2012; Andrabi *et al.* 2013 *Impact evaluations*
Parents' human and/ or social capital may moderate the effectiveness of information campaigns	Pakistan, India, Madagascar	Andrabi *et al.* 2013; Pandey *et al.* 2011; Lassibille *et al.* 2013 *Impact evaluations*
The quality or capacity of teachers is often an underlying issue, but the intervention does not directly address supply issues	Madagascar, India	Lassibille *et al.* 2010; Pandey *et al.* 2011 *Impact evaluations*
Concerns with elite capture did not appear to have materialised in Kenya	Kenya	Duflo *et al.* 2012 *Impact evaluations*

8.3 Public-private partnerships

The involvement of the private sector in education is seen by many as a viable strategy to improve access to and quality of education in L&MICs (Patrinos *et al.*, 2008). Although there are many ways in which the private sector may be involved in the educational sector, including through private investment funds or philanthropic foundations, our focus in this review is on public-private partnerships (PPPs). Fielden and LaRocque (2008, p.8) define PPPs as ‘‘initiatives under which private operators are contracted to manage public schools, voucher programmes, and school infrastructure partnerships’’.

Under PPPs typically the government develops education policy and provide finances, while private actors, either profit, non-profit or faith-based organisations, delivers services to students (Patrinos *et al.*, 2009:1). Some countries subsidise existing private schools to improve their capacity to educate, while other countries bring in private organisations to manage public schools (Patrinos *et al.*, 2009). In some instances, the private sectors solely invest and builds school infrastructure while the government provides funding for schooling. Government subsidies facilitate access to privately provided education for children from poorer households through a range of mechanisms, including school vouchers, direct subsidies and scholarships covering all or parts of the costs of accessing private education.

8.3.1 How might PPPs improve educational outcomes?

The argument about the role of the private sector in providing education suggest two main mechanisms through which private providers may contribute to improving education outcomes. First of all, bringing private sector partners into the national schooling system will lead to an increase the number of school providers and therefore to a rapid expansion of access to education and improved school choice (Patrinos *et al*, 2008:30). Secondly, the involvement of private actors may introduce both providers that deliver better quality services and increase competition due to parental choice, leading to an improvement in the quality of education in the long run.

The *World Development Report 2004 (*World Bank, 2004, xv) notes that the strengthening of relationships of accountability between policymakers, providers and citizens is crucial for improving the delivery of essential services to the poor, including education. The establishment of PPP may increase such accountability between the public sector and private providers, and institutions may be held accountable by parents as a result of the *parental choice,* which results from a PPP (Patrinos *et al.,* 2008).

Figure 8.3a below provides an ideal type programme theory, mapping out the causal chain how PPP may improve education outcomes. The main input is financial resources from the public sector, provided either directly to private providers or to students in the form of vouchers or scholarships. This funding either contributes to an overall increased supply of school places, or to improved access and choice to use private education for students who would otherwise not be able to access this. The theory then suggest that this will lead to improved educational services as a result of the resources and expertise brought into the education sector by private and autonomous providers, as well as the increased market competition. Further, as private and public partners are accountable to each other new benchmarks for educational standards are established; with private partners holding the responsibility for attracting new students and achieving targets, while governments set quality requirements.

The programme theory only holds with a number of assumptions. For instance, the design and management of partnerships between private and public sectors requires that government agencies have the appropriate resources, information and skills to do so (Patrinos *et al.,* 2009). It also assumes the government has set clear objectives and criteria the private sector must meet and that the bidding process is appropriately executed and qualified private partners are available and selected. (Fielden & LaRocque, 2008; Patrinos *et al.*, 2009:5). Sufficient quality assurance (QA) mechanisms and performance measures also need to be in place to ensure private providers are held accountable. It is also assumed that mechanisms are in place for parents to make informed schooling choices. Finally, the educational standard of new providers must be sufficient to attract students and improve learning outcomes.

8.3.2 Description of included studies

We included 13 studies reported in 21 different papers that evaluated the effect of PPPs on education outcomes in L&MICs. These referred to 13 unique programmes and were published between 1998 and 2015. Table 8.3a provides an overview of the characteristics of included studies, described in detail below.

Population

Eight of the included studies looked at the outcomes of these programmes at the primary school level (Barrera Osorio *et al.,* 2011; Alderman *et al.,* 2003; Correa *et al.,* 2014; Lara *et al.,* 2009, 2011; Saavedra Facusse, 2013; Muralidharan and Sundararaman, 2013; Dang *et al.,* 2014; Adelman and Holland 2015), three at the secondary school level (Angrist *et al.,* 2002; Zhang 2009; Barrera-Osorio *et al.,* 2015), and two at both levels (Barrera-Osorio 2006; Barrera-Osorio and Raju, 2011).

Three studies covered both public and private for-profit in their sample (Lara *et al.,* 2009, 2011; Muralidharan and Sundararaman, 2013; Zhang, 2009). Three studies covered only private for-profit schools in their samples (Barrera-Osorio *et al.,* 2011; Barrera-Osorio and Raju, 2011; Barrera-Osorio *et al.,* 2015), one study covered only public schools (Saavedra Facusse, 2013), and two studies covered multiple school types in their sample (for-profit, community and religious schools, and schools run by charitable foundations) (Angrist *et al.,* 2002; Adelman and Holland 2015). Finally, three of the studies evaluated programmes that established new schools. There were therefore no schools as baseline for these studies (Dang *et al.,* 2011; Alderman *et al.,* 2003; Barrera-Osorio, 2006).

Eleven studies reported the grades assessed, which ranged between grade 2 and grade 11 (Correa *et al.,* 2014; Lara *et al.,* 2009, 2011; Saavedra Facusse, 2013; Muralidharan and Sundararaman, 2013; Dang *et al.,* 2014; Barrera-Osorio, 2006; Angrist *et al.,* 2002; Zhang 2009; Barrera-Osorio and Raju, 2011; Barrera-Osorio *et al.,* 2015; Adelman and Holland 2015). Three studies report the ages of students in their sample, which ranged between 4 and 16 years old (Barrera-Osorio *et al.,* 2011; Alderman *et al.,* 2003; Barrera-Osorio *et al.,* 2015).

Setting

The programmes evaluated by the studies cover a range of settings in Latin America, South Asia, East Asia and Sub-Saharan Africa. Three of these programmes were located in in Chile (Lara *et al.,* 2011; Saveedra, 2013; Correa *et al.,* 2014), three in Pakistan (Alderman *et al.,* 2003; Barrera-Osorio *et al.,* 2011; Barrera-Osorio & Raju, 2011); two in Colombia (Angrist *et al.,* 2002; Barrera-Osorio, 2006); one in India (Muralidharan and Sundararaman, 2013); one in Bangladesh (Dang *et al.,* 2011); one in China (Zhang, 2009); one in Haiti (Adelman and Holland 2015); and one in Uganda (Barrera-Osorio *et al.,* 2015). Four studies took place in an urban setting (n=4) (Alderman *et al.,* 2003; Zhang, 2009; Angrist *et al.,* 2002; 2008; Barrera-Osorio, 2006), four studies took place in a rural setting (Muralidharan and Sundararaman, 2013; Dang *et al.,* 2014; Barrera-Osorio *et al.,* 2011; Barrera-Osorio *et al.,* 2015) and one study included both urban and rural settings (Barrera-Osorio and Raju, 2011). Four studies took place at a national scale (Correa *et al.,* 2014; Lara *et al.,* 2011; Saavedra Facusse, 2013; and Adelman and Holland 2015) and we have therefore assumed that these programmes were implemented in rural, urban and peri-urban settings.

Interventions

Table 8.3b provides an overview of the main design features of the PPP programmes in the included studies. All interventions have the key element of PPPs, that is, an initiative under which the public sector funds and contracts private operators to run schools, voucher programmes and school infrastructure partnerships (Fielden and LaRocque, 2008). Although PPPs are highly heterogeneous, the evaluated programmes can generally be divided into two main categories; first of all, nine programmes include the provision of a publicly-funded per-student voucher, or subsidy (Angrist *et al.,* 2002; Barrera-Osorio & Raju, 2011; Correa *et al.,* 2014; Lara *et al.,* 2009, 2011; Muralidharan and Sundararaman, 2013; Saavedra Facusse, 2013; Zhang, 2009; Barrera-Osorio *et al.,* 2015; Adelman and Holland 2015). The voucher or subsidy was either paid directly to schools or to students to use to attend fee charging schools. The Magnet Schools in China (Zhang, 2009) falls within this first category, but it should be noted that they differ from schools in other programmes in that they only receive public funds to cover teacher salaries.

Three programmes include the establishment of publicly-funded schools which are constructed and managed by private providers (Barrera-Osorio, 2006; Barrera-Osorio *et al.,* 2011; Alderman *et al.,* 2003). One programme, the ROSC School programme covers both elements in its two treatment arms (Dang *et al.,* 2011). Six programmes have an additional activity to improve the capacity of educational providers and/or parents in addition to the key features of public private partnerships as described above (Angrist *et al.,* 2002; Barrera-Osorio and Raju, 2011; Dang *et al.,* 2011; Barrera-Osorio *et al.,* 2015; Barrera-Osorio *et al.*, 2011; Alderman *et al.,* 2003).

Table 8.3 a: Intervention design features

	Angrist et al. (various), Bettinger et al. (2008)	Correa et al. (2014)	Lara et al. (2009, 2011)	Saavedra Facusse (2013)	Muralidharan and Sundararaman (2013)	Zhang (2009)	Barrera-Osorio and Raju (2011)	Adelman and Holland 2015	Dang et al. (2011)	Barrera-Osorio et al. (2015)	Barrera-Osorio (2006)	Barrera-Osorio et al. (2011)	Kim et al. (1998), Alderman et al. (2003)
Main components													
Publicly-funded per-student vouchers to private schools or students (to attend private school)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
Construction of publicly-funded privately-run schools									✓		✓	✓	✓
Additional activity to improve the capacity of educational providers	✓						✓		✓	✓		✓	✓

Comparisons

All but one study compare a treatment to a business as usual comparison group. Out of these, two studies (Barrera-Osario *et al.,* 2011; Dang *et al.,* 2011) include multiple treatment arms, but both include a no treatment comparison group, which is the comparison selected for all analyses. One study uses a pipeline (wait-list) design (Barrera-Osorio *et al.,* 2015).

Outcomes

The included studies reported on a variety of outcomes. The majority of the studies (n=6) looked either at some measure of learning outcomes (n=9), or enrolment (n =8). All studies measuring learning outcomes did so through test scores in mathematics (n=8), official language (n=8), local language (n=1) and composite measures (n=4). Enrolment was measured in different ways. Two studies used enrolment rates as reported by schools.[135] (Barrera-Osorio *et al.,* 2011; Barrera-Osorio and Raju, 2011); two studies measured enrolment through a household survey (Dang *et al.,* 2011; Zhang, 2009); one study used enrolment rates as reported by school censuses (Adelman and Holland 2015); one study administered school surveys (Barrera-Osorio *et al.,* 2015); and for one study it was unclear as to how enrolment rates were measured (Angrist *et al.,* 2002).

Three studies measured student completion outcomes. Both Angrist *et al.* (2002), Adelman and Holland (2015), and Barrera-Osorio *et al.* (2015) measured student repetition rates. One study assessed drop-out rates (Barrera-Osorio, 2006), although it is not clear how this was measured.[136] Another study measured attendance rates ('absence last week') (Barrera-Osorio *et al.,* 2015). Finally one study measured a secondary outcome: teacher attendance (Barrera-Osorio *et al.,* 2015).

Some of the studies present results separately for boys and girls (Angrist *et al.,* 2002; Barrera-Osorio *et al.,* 2011; Kim *et al.,* 1998, Alderman *et al.,* 2003; Dang *et al.,* 2011; Barrera-Osorio *et al.,* 2015), for different grades (Muralidharan and Sundararaman, 2013; Barrera-Osorio *et al.,* 2015), for boys and girls by grade (Adelman and Holland 2015) for different ages (Dang *et al.,* 2011) or for different quantiles of baseline achievement (Angrist *et al.,* 2002).

Study Design

Four of the included studies were cluster- randomised control trials where the intervention was assigned at either the school or village level (Alderman *et al.,* 2003; Barrera-Osorio *et*

[135] Enrolment rates used by Barrera-Osorio and Raju (2011) were verified by the implementing agency (PEF) in case enrolment increased by 50 students or more.
[136] Paper reports: number of students who dropped out.

al., 2011; Muralidharan and Sundararaman, 2013; Barrera-Osorio *et al.,* 2015). Three studies were natural experiments where vouchers were allocated by lottery or similar mechanism (Angrist *et al.,* 2002; Barrera-Osorio *et al.,* 2006; Zhang, 2009). Another four studies used a controlled before and after study design (Correa *et al.,* 2014; Dang *et al.,* 2011; Saavedra Facusse, 2013). One study used a randomised control trial (Adelman and Holland 2015), and one study used a regression discontinuity design (Barrera-Osorio & Raju, 2011).

Table 8.3 b: Characteristics of included studies PPP

Included study	Setting	Intervention summary	Included outcomes	Follow- Up	Study design	Sample Size
Angrist *et al.* (2002)	Colombia (urban), Secondary school Age: 12.5 (mean age voucher applicants) Grade: 6 at baseline	Programa de Ampliación de Cobertura de la Educación Secundaria (PACES). The programme provided vouchers to allow low-income students to attend private secondary schools. Vouchers covered students' annual matriculation fee in addition to ten monthly tuition payments. Vouchers were automatically renewed, providing that students maintained the required academic performance.	Enrolment; Completion; Learning;	3 years (+7 year follow up)	Natural experiment	Up to 1223 students (smaller for some outcomes)
Barrera-Osorio (2006)	Colombia (urban), Primary school; secondary school Age: not reported Grade: 1-11	The Concession School Programme. Concession schools are privately-run public schools providing public education for an agreed period of fifteen years. While the government is in charge of school infrastructure (building the schools), financing, and selects student selection, the public provider contracts administrative and teacher staff, and implements its own pedagogic model.	Dropout, learning	3 years	Natural experiment	18630 students
Barrera-Osorio *et al.* (2011)	Pakistan (rural), Primary school Age: 8.5 (mean for sample) Grade: not reported	Promoting Low-Cost Private Schooling in Rural Sindh (PPRS). Entrepreneurs could apply to be granted a per-student cash subsidy to operate coeducational primary schools. The programme was managed by a quasi-governmental organisation who supported and managed the establishment of the schools. Enrolment is tuition-free and open to all children in the village between the ages of 5 and 9, with the entrepreneur receiving an enrolment-based subsidy.	Enrolment	1.5-2 years	Cluster-RCT	15480 students

Alderman *et al.* (2003)	Pakistan (urban), primary school Age: 6 (mean for sample) Grade: Not reported	Urban Girls' Fellowship (UGF) Programme, Quetta. The UGF programme was aimed at delivering educational services to girls in the lower-income neighbourhoods of Quetta by establishing private girls' schools. Parents were invited to form a Parent Education Committee (PEC) to then develop a proposal for the new school. The schools would receive public support for a three-year period and, although they were expected to be largely self-sufficient by the fourth year, school could still apply for grants from The Balochistan Education Foundation.	Enrolment	1 year	Cluster- RCT	Up to 1553 students (smaller for some outcomes)
Correa *et al.* (2014)	Chile (rural, urban, per-urban), Primary school Age: not reported Grade: 4th	Subvencion Escolar Preferencial (SEP). Schools receive a subsidy for each enrolled priority student (low-income students with limited access to high-performing schools) in addition to the flat voucher fee schools receive under the Universal Voucher System. Schools sign the Equality of Opportunity and Educational Excellence Agreement, thereby agreeing to improve the quality of their education and to enrol and retain priority students. The schools agree to exempt priority students from all payments and to retain all students, even those with poor academic performance.	Learning	2 years	Cluster- RCT	17651 students
Lara *et al.* (2011)	Chile (rural, urban, per-urban), primary school Age: not reported Grade: 8 at baseline	Vouchers for private schooling. The Chilean Government provides voucher-type subsidies to public (municipal) and private voucher schools. Subsidies are paid directly to schools willing to accept the vouchers on a per-student basis. An important element of the programme is that, providing that students are not limited by	Learning	2 years	Cluster- RCT	22146 students

		geographical or financial constraints, they can attend the school of their choice.				
Saavedra Facusse (2013)	Chile (rural, urban, peri-urban), primary school Age: not reported Grade: 4 at baseline	Financiamiento Compartido (FICOM). Schools which decide to adhere to FICOM are allowed to ask for funds from families as well as from the state. The amount of support the schools will receive from the states, depends on how much the schools receive from the families. The more the schools charge the families, the less the support they will get from the state. Vouchers are awarded to poor students so that they were exempted (partially or totally) from the fees of FICOM. Each school is free to assign these scholarships using its own criteria.	Learning	4 years	CBA (quasi-experiment with baseline and endline data collection)	Up to 42722 students (smaller for some outcomes)
Muralidharan and Sundararaman (2013)	India (rural), Primary school Age: not reported Grade: 2 and 4	The Andhra Pradesh School Choice Project (under APRest). The project provided vouchers for students attending free public schools to attend a participating private school of their choice for the entire duration of their primary education. Private schools could determine the number of places to be allocated to voucher students, but could not select the students.	Learning	4 years	Cluster-RCT	5316 students
Barrera-Osorio & Raju (2011)	Pakistan (rural, uban), Primary school; Secondary school Age: not reported Grade: 1-11	Foundation Assisted Schools Programme. Per-student subsidies are provided to low-cost private schools under the condition that the school waives tuition and fees for all students and that the school receives a minimum pass rate in the Quality Assurance Test (QAT). Fulfilling these requirements, schools are eligible for additional annual cash benefits: group-based bonuses for teachers in schools that achieve high QAT pass rates and bonuses for schools	Enrolment	1.5 years (phase 4 programme schools); 2 years (phase 3 programme schools)	RDD	192784 students 830 schools

		that rank highest in the QAT in each main programme district.				
Zhang (2009)	China (urban), Secondary school Age: not reported Grade: 7 at baseline	Magnet Schools programme. Students have the additional option of applying to a Magnet School, as opposed to their nearest local public school (adding an element of school choice to the educational system). Magnet schools exist only at middle school level (grade 7-9) and are semi-private; they obtain their funding from the local government and through tuition fees.	Enrolment, Learning	3 years	Cluster- RCT	Up to 11734 students (smaller for some outcomes)
Dang *et al.* (2011)	Bangladesh (Rural), Primary school Age: 11.02 (mean) for grant areas, 11.06 (mean) for grant and allowance areas Grade: not reported	The Reaching Out of School Children (ROSC) project. School grant provided to NGOs for the purpose of establishing a new single-teacher and single-classroom school (ROSC school) and to provide complementary materials, infrastructure, teacher training and teacher salaries, sanitation (inc. safe drinking water), maintenance and repairs. A second treatment arm consisted of a similar but smaller school grant and a student allowance. This allowance provided a stipend for out-of-school children to attend school and was conditional upon their grades and attendance.	Enrolment	Approx 4- 5 years	CBA (quasi-experiment with baseline and endline data collection)	799 students

Barrera-Osorio *et al.* (2015)	Uganda (Rural), Secondary School Age:15.84 (mean for treatment group) Grade: senior 1-senior 3	The Universal Secondary Education Programme. Private schools receive approx. 18.8 USD per eligible student per term to cover eligible students' non-boarding fees. Schools additionally receive material support including textbooks and other teaching materials (dictionaries, CDs, laboratory equipment), and training for teachers. Schools may enrol non-eligible students for a fee, but must institute a board of governors to govern school finances and operations. The grant is to be spent as per the approved budget. There are penalties for non-compliance.	Enrolment, Attendance, Completion, Learning	4-11 months (3 follow-ups)	Cluster-RCT	94 schools
Adelman and Holland (2015)	Haiti, rural, urban and peri-urban private schools	Programme de Subvention (TWP or Tuition Waiver Programme): provides an annual per-student payment to participating non-public schools that agree to not charge any form of tuition fees to students. TWP is subject to several conditions, aimed at improving the learning environment and compliance with grade-for-age. Only children entering grade one for the first time and aged between six and eight are eligible for the subsidy. Schools are required to provide students with at least three textbooks. The TWP allows a range of potential uses of the subsidies, treatment schools could invest in improving existing infrastructure, expansion, staffing, furnishings, learning materials etc.	Completion Enrolment	60 months	RCT	652 schools

8.3.3 Synthesis of findings

The results of our synthesis are presented in two sections. First, we present the findings of the meta-analysis on the effects of PPP on primary and secondary outcomes, and any results available for sub-groups (Questions 1a and 1b). Second, we present a discussion and summary of the findings incorporating evidence from the descriptive qualitative synthesis to assess factors related to intervention design, implementation and context which might act as barriers or facilitators of the effectiveness of PPP (Questions 2a and 2b).

Effects of PPP interventions on primary and secondary outcomes

This section reports the results of the meta-analysis of the effects of PPP, addressing research questions 1a and 1b. We have structured the presentation of results according to the theory of change (Figure 8.3a) outlined above, starting with access outcomes (enrolment, girls' enrolment, and completion), followed by learning outcomes (composite test scores, language arts test scores and maths test scores).

All effect sizes have been expressed as standardised mean difference (SMD), interpreted as the magnitude of the number of standard deviation changes in the outcome for the intervention group as compared to students in non-PPP schools. SMD scores have been interpreted as the number of standard deviation changes in the outcome.

None of the studies reported on all outcomes, and for several outcomes only one study contributes effect size data. We were able to conduct meta-analysis for enrolment, completion, composite test scores, language arts and maths. The number of effect sizes range from two for completion, to seven for maths, language arts and girls' enrolment. In the case of girls' enrolment, six of the effect sizes come from the same study, but represent different grade sub-samples (Barrera-Osorio *et al.,* 2015). We did not have sufficient studies to conduct meta-analysis for any other outcome than enrolment, where we have presented the results for girls only separately. The studies often report multiple outcome measures and follow up periods. We followed the general rules outlined in the methods section and detailed below to select outcomes for inclusion in the meta-analysis. However, we calculated SMDs for all outcomes that met our inclusion criteria and we have commented on these in the results section where they have been substantively different from the estimates included in the meta-analysis (all ESs not included in the meta-analysis are provided in Appendix H).

Angrist et al. (2002) report effects on reading and writing separately and we calculated a synthetic effect size for inclusion in the meta-analysis. Muralidharan and Sundararaman, 2013 report effects on Telugu, Hindi and English. Telugu is the local language of the majority of the children in the sample and we included this in the meta-analysis. Barrera-Osorio *et al.,* 2015 report results from both a researcher administered test and a national test, and we included the latter in the meta-analysis.

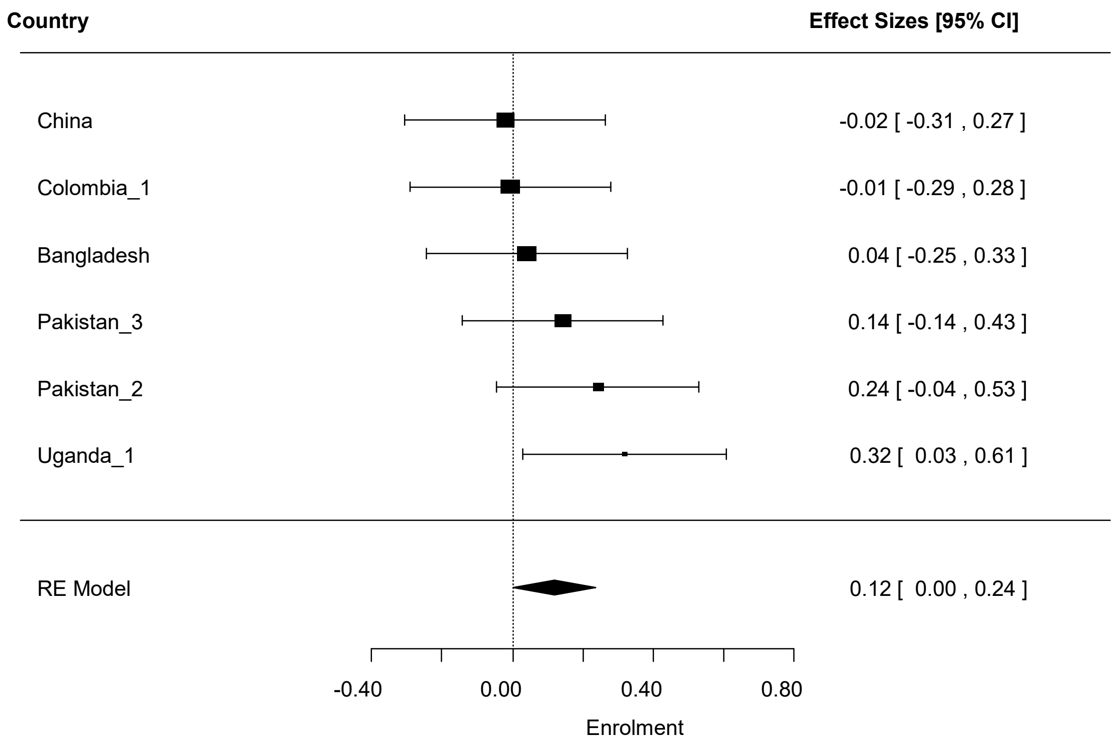
Two studies report on the same outcome measures at two follow up periods (Angrist *et al.,* 2002; Muralidharan and Sundararaman, 2013). The first follow up was most similar to the other studies in both studies (3 years and 2 years respectively), so we included these in the meta-analysis. Barrera-Osorio *et al.* (2015) report results for three follow up periods, all less than one year; we have included the longest follow up (11 months) in the meta-analysis, as this is most similar to the other studies. Finally, we were unable to include one study in the meta-analysis (Alderman *et al.,* 2003) because the study did not report the necessary statistical information and we were unable to obtain this.

Enrolment

The overall average effect of PPP on student enrolment is 0.19, 95% CI [0.01, 0.36], calculated under a random-effects model. The assessment of homogeneity suggests that the effects do not arise from a common population ($I^2 = 87.65\%$ $\tau^2 = 0.05$, Q (df = 6) = 50.39, p-val = < .0001). Figure 8.3b presents the forest plot with the results of the individual studies and the overall estimate. The effect sizes range from -0.02, 95% CI [-0.31, 0.27] in China (Zhang, 2009), to 0.61, 95% CI [0.46, 0.77] in Haiti (Adelman, M. 2015.

The average effect is sensitive to the inclusion of several studies. When the study from Nepal (Adelman, 2015) is removed the magnitude of the effect is reduced to 0.10 (95% CI, [0.00, 0.20], see Appendix H for full results of all sensitivity analyses).

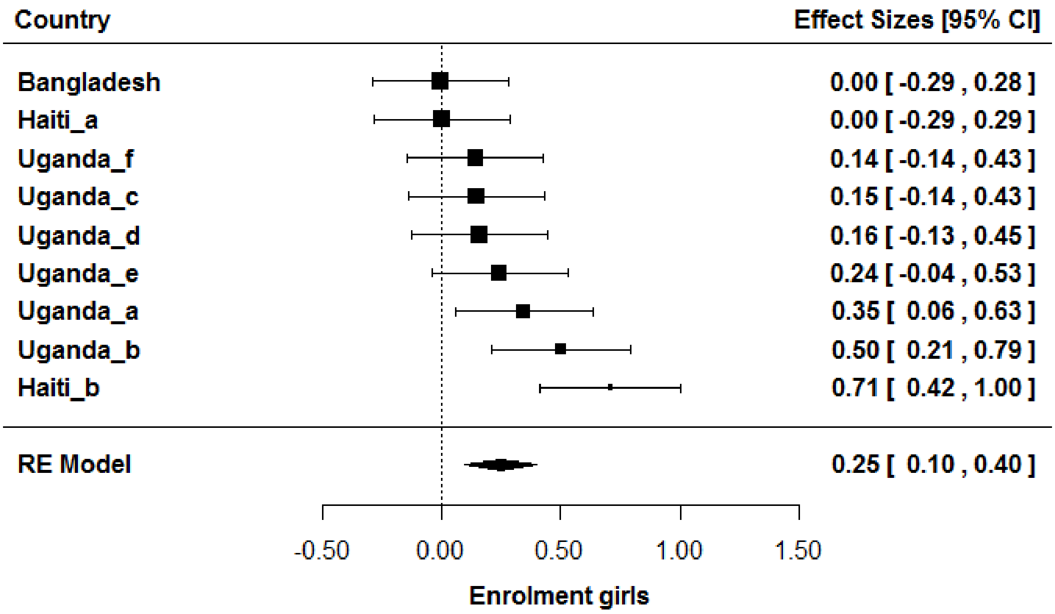
Figure 8.3 b: Enrolment



Girls' Enrolment

The overall average effect of PPP on girls' enrolment is 0.25, 95% CI [0.10, 0.40], calculated under a random-effects model. The homogeneity test suggest a moderate amount of between-study variability (I^2 = 60.28%, τ^2 = 0.033, Q(df = 8) = 20.1, p-val = 0.0101). Figure 8.3c presents the forest plot with the results of the individual studies and the pooled point estimate. The estimates are from three different studies, with one study from Uganda contributing estimates from six independent sub-samples and the study from Haiti contributing two independent sub-samples. The effect sizes range from 0.00, 95% CI [-0.29, 0.28] in Bangladesh to 0.71 SMD, 95% CI [0.42, 1.00] in Haiti (Adelman, 2015). Removing the study from Bangladesh increases the magnitude of the effect slightly and removing the Haiti grade 1- 4 sub- sample, the Uganda grade 1 or grade 2, sub-sample result in a slight reduction in the magnitude of the overall effect, however the effect remains substantively similar.

Figure 8.3 c: Enrolment Girls[137]

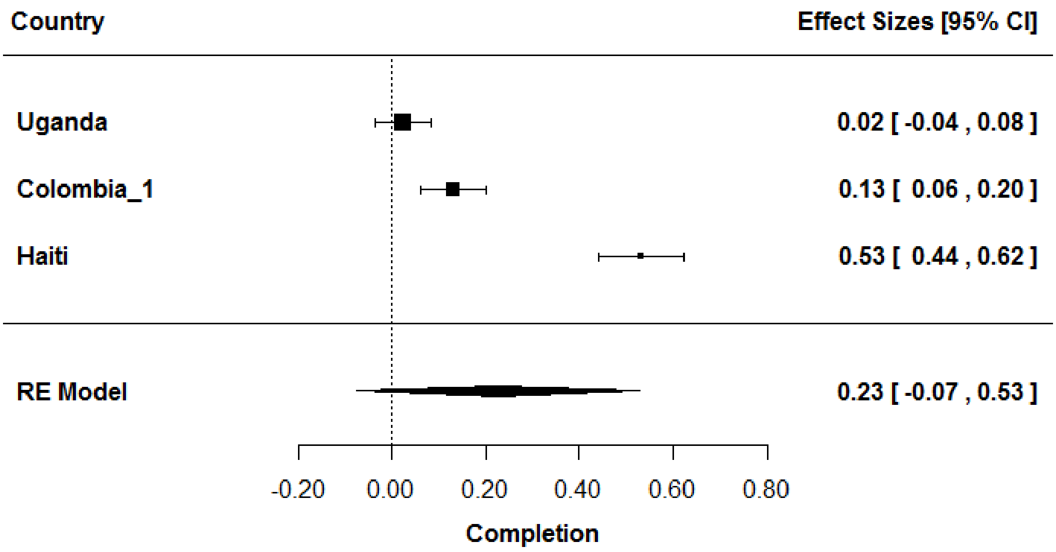


Completion

Three studies report effects on completion and the overall average effect of PPP on student completion is 0.23, 95% CI [-0.07, 0.53], calculated under a random-effects model. However, homogeneity tests (I^2 = 98.09%, τ^2 = 0.0697 , Q(df = 2) = 84.98, p-val < 0.001) indicate a large amount of variability. As can be seen from Figure 8.3d, while there appears to be no effect on completion in Uganda, the programmes in Colombia and Haiti appear to have increased completion rates (SMD=0.13, 95% CI [0.07, 0.19]; SMD= 0.23, 95% CI [0.44, 0.62]).

[137] Uganda_f refers to Barrera-Osorio *et al.* 2015 (Grade 6 Female at third follow-up); Uganda_c refers to Barrera-Osorio *et al.* 2015 (Grade 3 Female at third follow-up); Uganda_d refers to Barrera-Osorio *et al.* 2015 (Grade 4 Female at third follow-up); Uganda_e refers to Barrera-Osorio *et al.* 2015 (Grade 5 Female at third follow-up); Uganda_a refers to Barrera-Osorio *et al.* 2015 (Grade 1 Female at third follow-up); Uganda_b refers to Barrera-Osorio *et al.* 2015 (Grade 2 Female at third follow-up); Haiti_a refers to Adelman 2015 (Grade 5- 6); Haitib_b refers to Adelman 2015 (Grade 1- 4)

Figure 8.3 d: Completion[138]



Other access outcomes

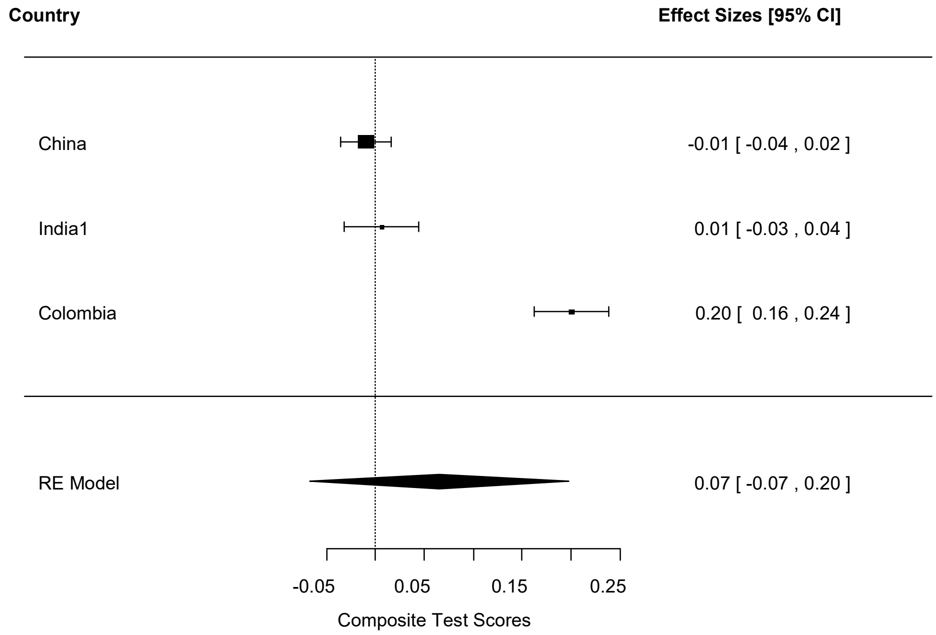
Few studies report on other access outcomes than enrolment and completion. Barrera-Osorio *et al.* (2006) find a reduction in drop-out rates (SMD= - 0.18, 95% CI [-0.24, -0.12]) among children attending concession schools in Colombia as compared to the control.

Composite test scores

Three studies measure outcomes by a composite test score and the average pooled effect under random effects is 0.07, 95% CI [-0.07, 0.20]. The assessment of homogeneity suggest that the effects do not arise from a common population ($I^2 = 97.85\%$ $\tau^2 = 0.0133$, $Q(df = 2) = 85.5202$, p-val = < .0001). Figure 8.3e presents the forest plot with the results of the individual studies and the average pooled estimate. It shows that the two studies from China and India (Zang, 2009; Barrera-Osorio *et al.*, 2011; Muralidharan and Sundararaman, 2013) are clustered around zero and their confidence intervals overlap, whereas the study from Colombia (Angrist *et al.*, 2002) has a substantively larger effect size (SMD=0.20, 95% CI [0.16, 0.24]) without confidence intervals overlapping with the other studies. As expected, the results are sensitive to the removal of this study, reducing the effect to zero (SMD=0.00, 95% CI, [-0.03 0.02]. In addition, reporting the effect of the voucher programme after a longer follow up period of four years, Muralidharan and Sundararaman (2013) find a larger effect than the two year follow up included in the meta-analysis (SMD=0.08, 95% CI [0.04, 0.11]).

[138] Colombia_1 refers to Angrist *et al.* 2002 (Main sample). In both studies included in the meta-analysis as shown in Figure 8.3d 'completion' refers to repetition rates.

Figure 8.3 e: Composite Test Scores

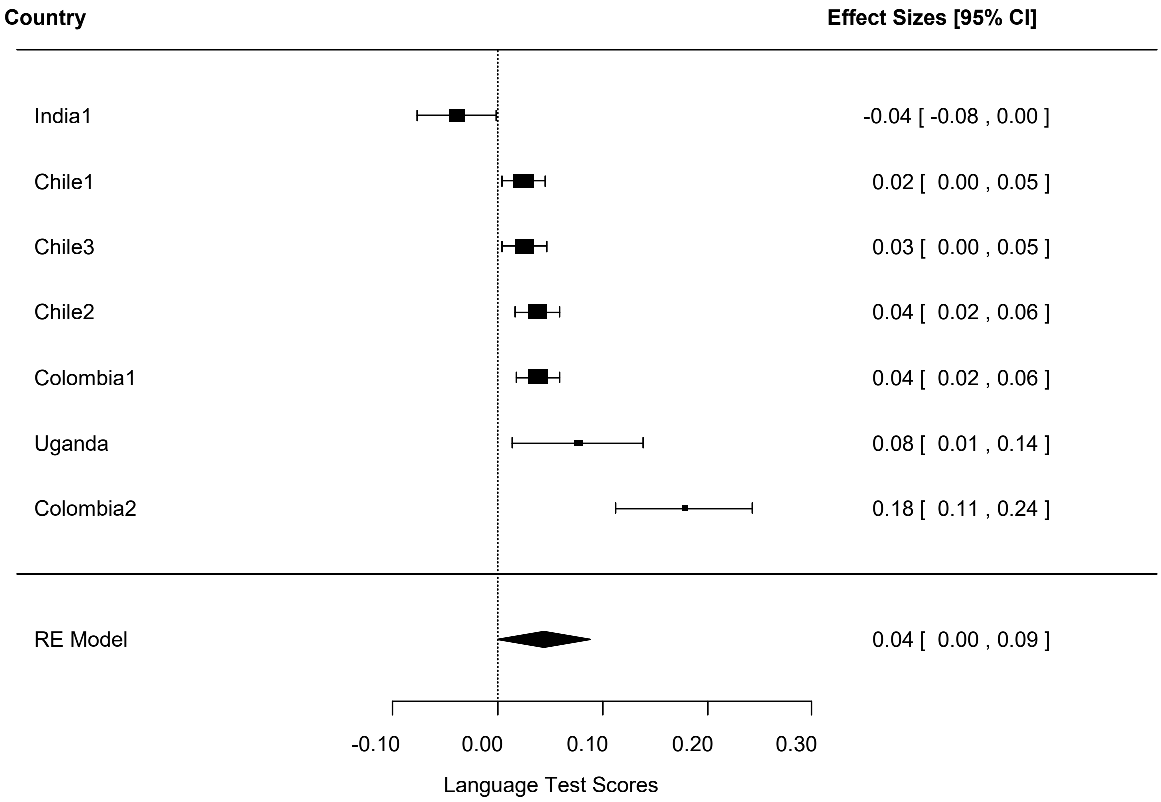


Language arts

The overall average effect of PPP on language arts is 0.04, 95% CI [0.00, 0.09], under random-effects model. The assessment of homogeneity suggest large amount of between-studies variability (I^2 = 94.32%, τ^2 = 0.0031, Q(df = 6) = 36.1790, p-val= < .0001). This is also apparent when inspecting the forest plot in Figure 8.3f. The point estimates range from -0.04 [-0.08, 0.00] in India (Muralidharan and Sundararaman, 2013) to 0.18 [0.11, 0.24] in Colombia (Angrist *et al.,* 2002). Four of the studies show small positive effects, with overlapping confidence intervals (Correa *et al.,* 2014; Saavedra Facusse, 2013; Lara *et al.,* 2011; Barrera-Osorio *et al.,* 2006), one study shows small, non-significant negative effects (Muralidharan and Sundararaman, 2013). Two studies suggest positive effects of larger magnitude (Barrera-Osorio *et al.,* 2015; Angrist *et al.,* 2002). Sensitivity analysis indicates that removing any one study does not make a substantive difference to the overall pooled effect (see Appendix H for results of all sensitivity analyses).

Several of the studies assess the effect of PPP on arts test scores for different sub-groups, but we were unable to combine these in a meta-analysis as they are measuring different groups (table 8.3a in appendix H include the SMD for all of these outcomes). In addition to the test scores for Telugu at two year follow up included in the meta-analysis Muralidharan and Sundararaman (2013) also measure Telegu at four year follow up, English (2 and 4 year follow up) and Hindi (4 year follow up). The latter is the only effect which is substantively different from the outcome included in the meta-analysis, with a positive effect of 0.21, 95% CI [0.17, 0.25]. Reporting the effect of the voucher programme for an earlier follow-up period, Barrera-Osorio *et al.* (2015) find a smaller effect compared to the follow-up included in the meta-analysis, which was conducted a year later (SMD= 0.02, 95% CI [-0.04, 0.09]). Saavedra Facusse (2013) conducted sub-group analysis for different age sub-groups (5-9 years and 14-16 years), but the effects are not substantively different from the effect for the full sample. Finally, Angrist *et al.* (2002) also provided sub-group analysis by gender, and find a larger effect on boys language arts test scores (SMD=0.24, 95% CI, [0.19, 0.29]) than for girls (SMD=0.12, 95% CI, [0.07, 0.18]).

Figure 8.3 f: Language Arts Test Scores[139]



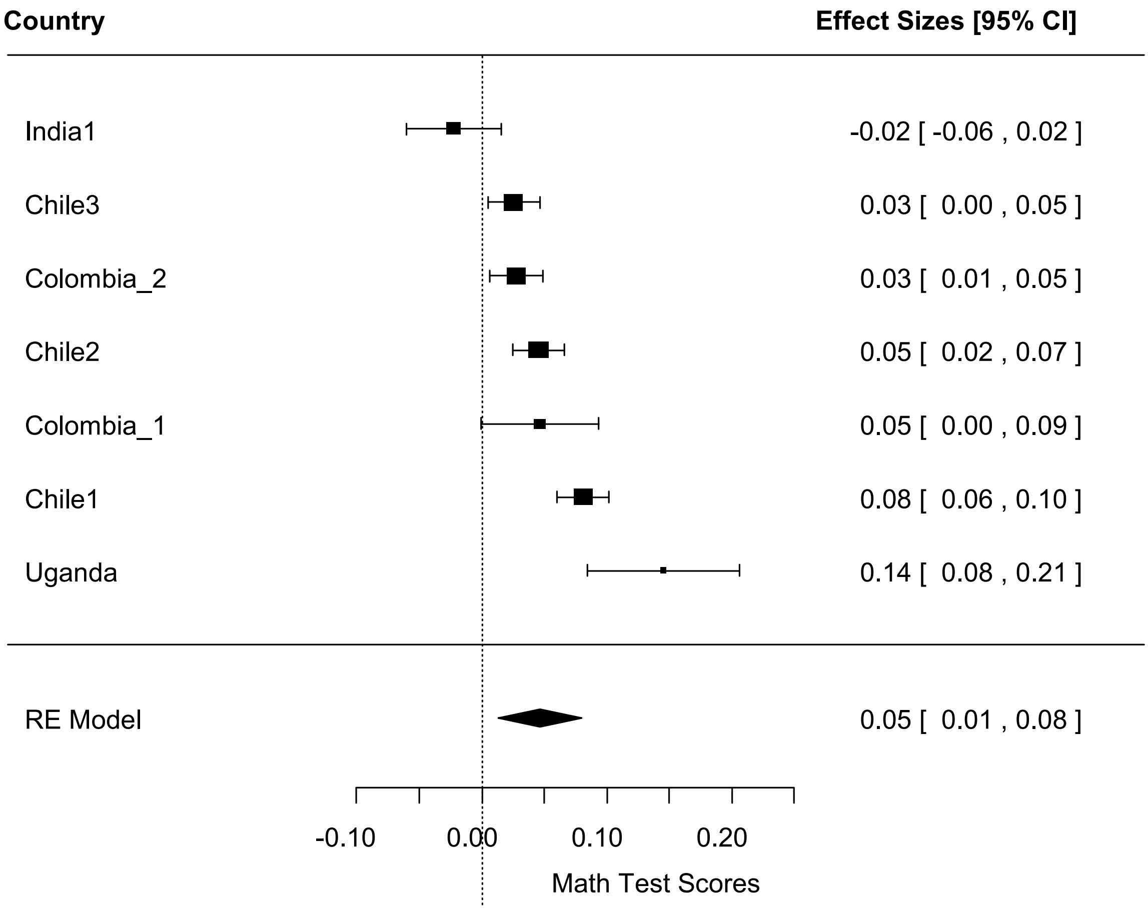
Maths

The overall average effect of PPP on maths is 0.05, 95% CI [0.01, 0.08], under random-effects model. The homogeneity tests (I^2 = 90.67%, τ^2 = 0.0018, Q(df = 6) = 39.8840, p-val < .0001) indicate that the effects did not arise from the same population. Figure 8.3g provides a forest plot with the results of the individual studies and the pooled estimate. The point estimates range from -0.02 [-0.06, 0.02] in India (Muralidharan and Sundararaman, 2013) to 0.14 [0.0, 0.21] in Uganda (Barrera-Osorio *et al.*, 2015).

The results do not change substantively when removing any single study from the analysis. Several of the studies assess the effect of PPP on maths scores for different follow-up periods and sub-groups, which we were unable to combine these in a meta-analysis. Muralidharan and Sundararaman (2013) provide a measure of maths at four-year follow up, but this is not substantively different than the outcome included in the meta-analysis. Reporting the effect of the voucher programme for an earlier follow-up period, Barrera-Osorio *et al.* (2015) find a smaller effect compared to the follow-up included in the meta-analysis, which was conducted a year later (SMD= 0.05, 95% CI [-0.01, 0.10]). Saavedra Facusse (2013) provide sub-group analysis for different age sub-groups (5-9 years, 10-13 years and 14-16 years), but the effects are not substantively different from the effect for the full sample. Finally, Angrist *et al.* (2002) provide sub-group analysis by gender, and find substantively larger effects for girls (SMD=0.41, 95% CI [0.08, 0.73]) than for boys (SMD=0.00, 95% CI [-0.33, 0.34]).

[139] India1 refers to Muralidharan and Sundararaman 2013 (Year 2 Telugu scores); Chile1 refers to Correa *et al.* 2012 (full sample); Chile3 refers to Saavedra Facusse 2013 (Full sample); Chile2 refers to Lara *et al.* 2009 (Full sample); Colombia1 refers to Barrera-Osorio 2006 (Full sample); Colombia2 refers to Angrist *et al.* 2002 (Full sample)

Figure 8.3 g: Maths Test Scores[140]



8.3.4. Summary of findings and discussion

We identified twelve studies of PPP programmes across seven different countries in South Asia, Latin America and Sub-Saharan Africa. We were able to examine effects on enrolment, completion, maths, language arts and composite test scores using meta-analysis. The overall average effect range from 0.04, 95% CI [0.00, 0.09] for language arts to 0.22, 95% CI [0.10, 0.34] for girls' enrolment. The results indicate that overall outcomes were better for children attending PPP schools as compared to those that do not. But the average effects are relatively small in magnitude for most outcomes and there is a large amount of between study variability for all outcomes. The results should therefore be interpreted with caution. The large amount of heterogeneity is not surprising given the variability in the intervention design, context and populations in the included studies, however we are unable to explain why PPPs appear to have worked in some contexts but not in others.

[140] India1 refers to Muralidharan and Sundararaman 2013 (Year 2); Chile3 refers to Lara *et al.* 2009 (Full sample); Colombia2 refers to Barrera-Osorio 2006 (Full sample); Chile2 refers to Saavedra Facusse 2013 (Full sample); Colmbia_1 refers to Angrist *et al.* 2006 (Full sample); Chile1 refers to Correa *et al.* 2012 (Full sample).

Table 8.3 c: Descriptive findings: Process, implementation and context

Descriptive finding: Process and Implementation		
Finding	**Context**	**Citation (info type)**
Inefficient administration led to an underutilisation of WB loan	Colombia	King *et al.* 1997 (Process Evaluation)
Delays to payment of vouchers put strain on programme implementation and school quality	Colombia,	King *et al.* 1997 (Process Evaluation);
Schools faced financial difficulties due to unsatisfactory fee collection and a lack of subsidies	Pakistan	World Bank 2001 (Project Document)
Programme appeared to have successfully targeted low-income students	Colombia, Pakistan,	Calderon 1996 (Process Evaluation); Barrera-Osorio and Raju 2011 (Impact Evaluation);
Not all Upazilas followed the targeting criterion of focusing on out-of-school children	Bangladesh	Ministry of Planning & GoB 2014 (Mixed Methods)
PPP programmes well-received by stakeholders	Pakistan, Colombia,	Orazem 2000 (Project Document); Villa & Duarte 2000 (Project Document)
Stakeholder participation central to some programmes	Chile, Pakistan	Irarrázaval *et al.* 2012: 32 (Mixed Method); Kim *et al.* (1998) (Impact Evaluation)
Lack of organisational capacity a barrier to effective implementation and monitoring, with reports of ghost' voucher awardees	Bangladesh, Colombia, Pakistan	Ministry of Planning & GoB 2014 (Mixed Methods); World Bank 2013 (Project Document); King *et al.* 1999 (Process Evaluation); SCSPEB n.d (Project Document)
Center Management Committees (who were in charge of the day-to-day management of schools) met infrequently	Bangladesh	Ahmed 2004 (Impact Evaluation)
High turnover of government or implementation staff	Colombia, Bangladesh, Pakistan	King *et al.* 1997 (Process Evaluation); World Bank 2013; CfBT Education Trust 2010 (Project Documents)
Schools did not comply with at least one of the programme condition and the ministry of	Haiti	Adelman and Holland 2015 (Impact Evaluation)

education did not attempt to enforce compliance by encouragement or sanctions		
Descriptive finding: Contextual Barriers and Facilitators		
Inflation reduced the value of the voucher, increasing co-payment	Colombia	Bettinger *et al.* 2008 (Impact Evaluation)
There was opposition from teacher unions to privatisation of education	Pakistan	Orazem 2000 (Project Document)
Requirement that community contributed land made it difficult to assign land for schools.	Pakistan	Orazem 2000 (Project Document)
Poor weather conditions were a common reason for not going to school	Uganda	Barrera-Osorio *et al.* 2015 (Impact Evaluation)

Table 8.3 d: Descriptive findings: Intermediate outcomes

Descriptive findings: Intermediate outcomes	Context	Citation (info type)
Students enrolling had higher scores than the district average	China	Zhang 2009 (Impact Evaluation)
There was limited change in children's time use at home and household spending patterns	India	Muralidharan and Sundararaman 2013 (Impact Evaluation)
Children in private school spent more time at school due to longer days and school year	India	Muralidharan and Sundararaman 2013 (Impact Evaluation)
Private schools spent less time on Maths and Telugu, but more on English, Science, Social Studies and Hindi	India	Muralidharan and Sundararaman 2013 (Impact Evaluation)
Teachers in private schools less educated, younger and paid lower salaries	India,	Muralidharan and Sundararaman 2013 (Impact Evaluation)
Teachers in private schools better qualified and higher per student spending	China	Zhang 2009 (Impact Evaluation)
Private schools outperform government schools on measures of classroom practices, teacher absence and teacher performance	India	Muralidharan and Sundararaman 2013 (Impact Evaluation)
Private schools had better infrastructure, equipment and supplies than public schools	Colombia, India	Barrera-Osorio 2006; Muralidharan and Sundararaman 2013 (Impact Evaluations)
Increase in availability of teachers, classrooms and blackboards, but not in number of toilets or student-teacher and student-classroom ratios	Pakistan	Barrera-Osorio and Raju 2011 (Impact Evaluation)
No significant changes in availability of school inputs after programme	Uganda	Barrera-Osorio *et al.* 2015 (Impact Evaluation)
Significant changes in teacher presence after programme	Uganda	Barrera-Osorio *et al.* 2015 (Impact Evaluation)
Lack of and low quality of school facilities	Bangladesh	Ahmed 2004 (Impact Evaluation)
Voucher winners more likely to access private education	Colombia	Angrist *et al.* 2011 (Impact Evaluation)
Participating schools have significantly higher proportion of students from educationally-favourite backgrounds	Uganda	Barrera-Osorio *et al.* 2015 (Impact Evaluation)

9. Multilevel interventions

The previous chapters have assessed the effects of programmes falling into distinct intervention categories. Many of these included programmes are relatively complex in their design with several components. For instance, structured pedagogy programmes typically include new content and materials, together with some training for teachers to deliver this content. However, what most of these programmes have in common is that they are focused on addressing barriers at the child, household, school, teacher or systems levels.

While reviewing the literature we also identified studies of programmes that did not fall clearly into one of these categories as they included a number of different interventions to address two or more barriers to improved education outcomes. In such cases, it was difficult to determine how to best classify these programmes. We therefore grouped these studies as 'multilevel interventions' and analysed them separately. We present the results of this analysis below.

9.1 Description of included studies

We included 12 studies reported in 14 different papers that evaluated the effect of multilevel interventions on education outcomes in L&MICs. These referred to 10 unique programmes and were published between 2002 and 2013. Table 9a provides an overview of the characteristics of included studies, described in detail below.

Population

Nine of the included studies looked at the outcomes of these programmes at the primary school level (Cerdan-Infantes and Vermeersch 2007; Paqueo and Lopez-Acevedo, 2003; Kremer *et al*., 2003; Lockheed *et al.*, 2010; Kazianga *et al.,* 2013; de Hoop and Rosati, 2012; Chay *et al.*, 2005; Tokman 2002; Bellei, 2013), none at just the secondary school level, and three at both levels (Min *et al.,* 2012; Rodriguez and Sanchez, 2010; Rosati and Rossi, 2007).

All but two studies covered public schools in their sample. The two studies assessing the BRIGHT programme did not have any schools at baseline as this programme included the construction of schools as one of their components (Kazianga *et al.*, 2013; de Hoop and Rosati, 2012). Eight studies reported the grades assessed, which ranged from grade 1 to grade 9 (Cerdan-Infantes and Vermeersch 2007; Paqueo and Lopez-Acevedo, 2003; Kremer *et al.,* 2003; Lockheed *et al.*, 2010; Kazianga *et al.*, 2013; Chay *et al.*, 2005; Min *et al.*, 2012; Rodriguez and Sanchez, 2010). Grades 3, 4 and 6 were those most frequently occurring in study samples. Only four studies report the ages of students in their sample, which ranged from 8 to 16 years old (Min *et al.,* 2012; de Hoop and Rosati, 2012; Kazianga *et al.*, 2013; Paqueo and Lopez-Acevedo, 2003).

Setting

The programmes evaluated by the studies took place in a range of settings in Latin America and the Carribean, East Asia and Sub-Saharan Africa. Three of these programmes were located in Chile (Bellei, 2013; Chay *et al.,* 2005; Tokman 2002); two in Mexico (Paqueo and Lopez-Acevedo, 2003; Rosati and Rossi, 2007); two in Burkina Faso (de Hoop and Rosati, 2012; Kazianga *et al*., 2013); one in Colombia (Rodriguez and Sanchez, 2010); one in Jamaica (Lockheed *et al.,* 2010); one in China (Min *et al*., 2012); one in Kenya (Kremer *et al.*, 2003); and one in Uruguay (Cerdan-Infantes and Vermeersch, 2007).

Five studies took place in a rural setting (de Hoop and Rosati, 2012; Kazianga *et al.*, 2013; Kremer *et al.,* 2003; Lockheed *et al.,* 2010; Rodriguez and Sanchez, 2010) and one took place in an urban setting (Cerdan-Infantes and Vermeersch, 2007). One covered both rural and urban areas (Bellei, 2013), whilst three studies evaluated programmes that were national in scale (Chay *et al*., 2005; Tokman, 2002; Ugarte, 2011; Paqueo and Lopez-Acevedo, 2003). We therefore assume that these four programmes were implemented in rural, urban and peri-urban settings.

Interventions

Table 9a provides an overview of the different intervention types incorporated in the programmes in the included studies. We categorised the different elements of these programmes in the same way we classified intervention types across the review. As can be seen in Table 9a, the included studies incorporate between four and 10 different interventions, most of them covering elements from our school and teacher levels. Within these categories, all but two programmes (Min *et al.*, 2012; Bellei, 2013) provided materials; all but three programmes (Min *et al*., 2012; Bellei, 2013; Lockheed *et al.,* 2010) constructed new or rehabilitated existing schools and infrastructure; and all but three programmes (Kremer *et al.*, 2003; Kazianga *et al.*, 2013; de Hoop & Rosati, 2012) trained teachers.

Other components that featured in the multilevel interventions programmes included remedial education (Cerdan-Infantes and Vermeersch, 2007; Chay *et al*., 2005; Tokman, 2002; Bellei, 2013); school feeding (Lockheed *et al*., 2010; Rosati and Rossi, 2007; Kazianga *et al*., 2013; de Hoop & Rosati, 2012), new pedagogical strategies (incl. extra time in the school day) (Cerdan-Infantes and Vermeersch, 2007; Lockheed *et al*., 2010; Bellei, 2013), the hiring of additional teachers (Kazianga *et al.*, 2013; de Hoop & Rosati, 2012), incentives for teachers (Paqueo & Lopez-Acevedo, 2003; Rosati and Rossi, 2007), the provision of information to parents (Kazianga *et al.*, 2013; de Hoop & Rosati, 2012), reduction of school fees (Kremer *et al*., 2003), diagnostic feedback strategies (Bellei, 2013), school-based management (Rosati & Rossi, 2007), and a school-based health component (Cerdan-Infantes and Vermeersch, 2007).

Programmes also included additional intervention components that did not fit neatly into any of our existing intervention categories. We have categorised these as ‘other’ interventions. The most common of these was the institutional strengthening of schools through some type of governance, leadership and management training and was incorporated by all but four programmes (Rodríguez & Sánchez, 2010; Min *et al*., 2012; Kazianga *et al*., 2013; de Hoop & Rosati, 2012). Other interventions in this category included the management of information systems (Lockheed *et al*., 2010; Min et al., 2012; Rosati & Rossi, 2007), local government capacity-building (Min *et al.*, 2012; Kazianga *et al.,* 2013; de Hoop & Rosati, 2012), parent engagement (Cerdan-Infantes and Vermeersch, 2007; Lockheed *et al*., 2010), adult literacy training for parents (Kazianga *et al*., 2013; de Hoop & Rosati, 2012), gender sensitivity training for parents (Kazianga *et al*., 2013; de Hoop & Rosati, 2012), classroom size reduction (Cerdan-Infantes and Vermeersch, 2007); decentralisation (Rodríguez & Sánchez, 2010), new classroom activities (Cerdan-Infantes and Vermeersch, 2007) and a Christmas party (Kremer *et al*., 2003).

Comparisons

All but one study compared a treatment to a business as usual comparison group. One study used a pipeline (wait-list) design (Rosati and Rossi, 2007).

Outcomes

The included studies reported on a variety of outcomes. All but one study (Rosati & Rossi, 2007) looked at some measure of learning. The studies reporting on learning outcomes did so through test scores in language arts (n=9), mathematics (n=8) and composite measures (n=3). Five studies reported on access outcomes. Five also reported on completion (Rodríguez & Sánchez, 2010; Kremer *et al*., 2003; Kazianga *et al*., 2013; de Hoop & Rosati, 2012; Bellei, 2013), three measured enrolment (Rodríguez & Sánchez, 2010; Kazianga *et al*., 2013; de Hoop & Rosati, 2012), two measured attendance (Rosati & Rossi, 2007; de Hoop & Rosati, 2012) and another two measured dropout rates (Rodríguez & Sánchez, 2010; Bellei, 2013). Some of the studies presented results separately for boys and girls (de Hoop & Rosati, 2012; Min *et al.,* 2012), for different age groups (Kazianga *et al.,* 2013), for different grades (Kremer *et al*., 2003) and by rural/urban setting (Paqueo and Lopez-Acevedo, 2003).

Study Design

Five of the included studies were controlled before-after studies (quasi-experiment) with baseline and endline data collection (Cerdan-Infantes and Vermeersch, 2007; Lockheed *et al*., 2010; Paqueo and Lopez-Acevedo, 2003; Bellei, 2013; Rodríguez & Sánchez 2010). Two studies were natural experiments (Min *et al*., 2012; Rosati and Rossi, 2007), two studies used a regression discontinuity design (RDD) (De Hoop and Rosati, 2012; Kazianga *et al*., 2013) and one study was a cluster randomised control trial (Kremer, 2003).

Table 9 a: Intervention design features

Study	Rodríguez & Sánchez, 2010	Lockheed *et al.*, 2010	Min *et al.*, 2012	Kremer *et al.*, 2003		Bellei, 2013	Chay *et al.*, 2005		Paqueo & Lopez-Acevedo, 2003	Rosati & Rossi, 2007	Kazianga *et al.*, 2013	de Hoop & Rosati, 2012
Intervention Component												
School-level: Materials	✓	✓		✓	✓		✓	✓	✓	✓	✓	✓
School-level: NSI	✓			✓	✓		✓	✓	✓	✓	✓	✓
School-level: Pedagogy		✓				✓						
School-level: Extra time					✓							
School level: Remedial education					✓	✓	✓	✓				
Teacher-level: Training	✓	✓	✓		✓	✓	✓	✓	✓	✓		
Teacher-level: Hiring											✓	✓
Teacher-level: Diagnostic feedback						✓						
Teacher-level: Incentives									✓	✓		
Child-level: School feeding		✓								✓	✓	✓
Child-level: SBH					✓							
Household-level: Reducing and eliminating user fees				✓								
Household-level: Providing information to parents											✓	✓
System-level: SBM										✓		
Other: class size reduction					✓							
Other: parent engagement		✓			✓							
Other: governance and management training		✓	✓		✓	✓	✓	✓	✓	✓		
Other: management information systems		✓	✓							✓		
Other: (local) government capacity building			✓								✓	✓
Other: decentralisation	✓											
Other: literacy for parents											✓	✓
Other: gender training parents											✓	✓

Table 9 b: Characteristics of Included Studies- Multilevel Interventions

Included study	Population	Intervention Summary	Included outcomes	Follow- up	Study design	Sample Size
Rodríguez & Sánchez, 2010	Colombia (rural) Primary and secondary school Grade: 1-6 Age: not reported	The Rural Education Project (PER):(1) decentralisation of programme implementation to allow Municipal Operating Units (UOMs) consisting of local officials and members of the education sector to choose among nine different flexible education models to obtain the most needed interventions for their schools. These educational models generally consisted of three key intervention components: (2) a set of materials that could include educational guides, laboratory equipment, desks, chairs, VHS players and videos; (3) a new school library (if demand); (4) training for teachers in all schools on how to implement their school's educational model.	Learning; Enrolment; Dropout; Completion	2-2.5 years	CBA (quasi-experiment with baseline and endline data collection)	3,003 schools
Lockheed *et al.,* 2010	Jamaica (rural) Primary school Grade: 3, 4, and 6 Age: not reported	The New Horizons for Primary Schools (NHP) Programme: School Improvement Plans (SIPs) were developed and interventions were selected for each school according to its needs. These interventions included: (1) improvement of pedagogical strategies in mathematics and literacy; (2) training for teachers in reading, mathematics and the use of technology; (3) supplementary materials and computers; (4) breakfasts to needy children; (5) improved management information systems; (6) parent engagement and parent capacity building; (7) governance, leadership and management training for schools.	Learning	5-7 years	CBA (quasi-experiment with baseline and endline data collection)	Up to 126 schools
Min *et al.,* 2012	China (not clear if rural, urban, peri-urban) Primary and secondary school Grade: 3, 5, 7 and 9 Age: Mean age	Southwest Basic Education Project (SBEP): (1) training for approximately 77,000 teachers is a participatory training approach. It also supported teachers' professional development to improve the quality of education;(2) school-based management (SBM) supporting 1,400 schools in poor townships to carry out school development planning (SDP) to improve the leadership and management capacity of head teachers, the school inspection system and foster parent	Learning	Approx. 5 years	Natural experiment	Up to 12,486 students

	in years Grades: 3: 10.3 (project) 9.7 (non-project) 5: 12.3 (project) 12.1 (non-project) 7: 13.6 (project) 13.8 (non-project) 9: 15.8 (project) 15.9 (non-project)	engagement (the activities implemented and the responsibilities given to various stakeholders in order to achieve this SBM component are not clear); (3) education management information systems through an integrated students-data base; (4) capacity building for government institutions at the national, provincial and county levels.				
Kremer *et al.*, 2003	Kenya (rural) Primary school Grade: 1-7 Age: not reported	Child Sponsorship Programme (CSP): (1) funds for additional textbooks; (2) built ten additional classrooms in each programme school; (3) provided uniforms to all children in treatment schools for the first three years of the programme. In the fourth and fifth years of the programme, half of the grades were provided uniforms in each year. The programme additionally provided a Christmas party to all treatment schools in the beginning of year 3.	Learning; Completion	Approx. 5 years	Cluster RCT	22,991 students
Cerdan-Infantes and Vermeersch, 2007	Uruguay (urban) Primary school Grade: 6th grade Age: not reported	Full Time School (FTS) Programme: (1) new classrooms and equipment; (2) a set of materials including maps, books and dictionaries; (3) lengthened the school day by doubling the amount of hours spent in class; (4) remedial education by allocating an extra 3 hours per week to students with special needs; (5) nutritional and health care support; (6) reduced class size; (7) complementary classroom activities, and (8) encouraged parent involvement; (9) governance, leadership and management training for schools by introducing teacher committees; (10) trained teachers in implementing the FTS pedagogical model and provided other teacher training courses.	Learning	Approx. 6 years	CBA (quasi-experiment with baseline and endline data collection)	Up to 11,115 students and 152 schools

Bellei, 2013	Chile (rural and urban) Primary school Age: Not reported Grade: Not reported	Technical Support to Failing Schools (TSFS) Program: Under this programme the Ministry of Education recruited external agencies to improve the quality of education by focusing on five common intervention components: (1) improved pedagogical practices; (2) diagnostic feedback where teachers reviewed students' progress on a monthly basis in order to help them plan better; (3) workshops and individual feedback for teachers focused on lesson planning and technical assistance in the classroom; (4) Remedial support to weaker students in both normal classes and additional sessions; (5) governance, leadership and management training for schools by encouraging innovation in school management, defining the roles and functions of staff, and encouraging principals to be results focused.	Learning; Completion	Approx. 3 years	CBA (quasi-experiment with baseline and endline data collection)	Up to 10,584 students
Chay *et al.*, 2005	Chile (rural, urban and peri-urban) Primary school Grade: 1-4 Age: not reported	P900 Programme: (1) infrastructural improvements, such as building repairs; (2) instructional materials including textbooks, small classroom libraries, cassette recorders and copiers; (3) training workshops for teachers focused on improving pedagogy in the teaching of language arts and maths; (4) remedial education through after-school tutoring workshops for students who were not performing at grade level; (5) governance, leadership and management training for schools by introducing management teams and annual improvement plans to schools. The former included teachers and principals and were designed to promote goal setting, motivation and accountability.	Learning	Approx. 2 years	RDD	Up to 2,644 students
Tokman, 2002	Chile (rural, urban and peri-urban) Primary school Age: not reported Grade: not reported	P900 Programme: (1) infrastructural improvements, such as building repairs; (2) instructional materials including textbooks, small classroom libraries, cassette recorders and copiers; (3) training workshops for teachers focused on improving pedagogy in the teaching of language arts and maths; (4) remedial education through after-school tutoring workshops for students who were not performing at grade level; (5) governance, leadership and management training	Learning	Approx. 6 years	CBA (quasi-experiment with baseline and endline data collection)	900 schools

		for schools by introducing management teams and annual improvement plans to schools. The former included teachers and principals and were designed to promote goal setting, motivation and accountability.				
Paqueo & Lopez-Acevedo, 2003	Mexico (rural, urban, peri-urban) Primary school Grade: 6th grade Age: 9.4 years for full sample (mean)	Programa para Abatir el Regazo Educativo (PARE): (1) provided books and didactic materials; (2) offered new infrastructure and distance educational technologies; (3) trained teachers and principals; (4) incentives for teachers; (5) institutional strengthening and monitoring. As it is not clear what exactly this entails, we have grouped this under the category 'other: governance, leadership and management training for schools' (see table 9a).	Learning	Approx. 3 years	CBA (quasi-experiment with baseline and endline data collection)	Up to 1,480 students
Rosati & Rossi, 2007	Mexico (not clear) Primary school Secondary school Grade: not clear Age: note clear	The Compensatory Education Programme (CONAFE): (1) the provision of school infrastructure, audio-visual technology and other equipment; (2) provision of learning materials to each student; (3) training and professional development of all teachers; (4) introduction of monetary incentives to teachers and principals; (5) school-based management. This latter component is a sub-component of the CONAFE programme but is also known as the 'Apoyo a la Gestion Escolar' (AGE) programme. It focuses on institutional strengthening through the provision of grants to parents and leaders, and through providing training to parent associations on how to spend the grant.	Attendance; Completion	Approx. 10 years	Natural experiment	Up to 54,431 students

Kazianga *et al.*, 2013	Burkina Faso (rural) Primary school Grade: 1-6 Age: For full sample, average age was 8.76 years.	The Burkinabé Response to Improve Girls' Chances to Succeed (BRIGHT) Programme: (1) constructed 132 primary schools; (2) provision of schools kits and textbooks to all students, as well as desks, chairs and book shelves; (3) school-feeding component consisted of schools canteens with daily meals for all and take-home rations for girls who had a 90 per cent attendance rate; (4) hired additional female teachers; (5) an information campaign about the benefits of education; (6) capacity building for local officials; (7) gender sensitivity training for teachers; (8) literacy training for parents.	Learning ; Enrolment; Completion	Approx. 4 years	RDD	17,970 students
de Hoop & Rosati, 2012	Burkina Faso (rural) Primary school Grade: not reported Age: 8.76 years for full sample (average)	The Burkinabé Response to Improve Girls' Chances to Succeed (BRIGHT) Programme: (1) constructed 132 primary schools; (2) provision of schools kits and textbooks to all students, as well as desks, chairs and book shelves; (3) school-feeding component consisted of schools canteens with daily meals for all and take-home rations for girls who had a 90 per cent attendance rate; (4) hired additional female teachers; (5) an information campaign about the benefits of education; (6) capacity building for local officials; (7) gender sensitivity training for teachers; (8) literacy training for parents.	Learning; Attendance; Enrolment	Approx. 4 years	RDD	Up to 248 villages or 18370 students

Note: We were unable to include any subgroup analysis for Min *et al.* (2012) in the analysis as standard errors for these outcomes were not reported in the study and we were unable to obtain them from the authors.

9.2 Synthesis of findings

The results of our synthesis are presented in two sections. First, we present the findings of the meta-analysis of the effects of multilevel interventions on primary and secondary outcomes. Second, we provide a summary of the findings and discussion, incorporating findings of the qualitative synthesis of process, implementation and contextual factors that may explain effects of education outcomes.

Effects of Multilevel interventions on access and learning outcomes

This section reports the results of the meta-analysis of the effects of multilevel interventions, addressing question 1a. We structure the presentation of results according to the causal chain, starting with school participation through enrolment and attendance followed by dropout, completion, composite scores, maths and language arts.

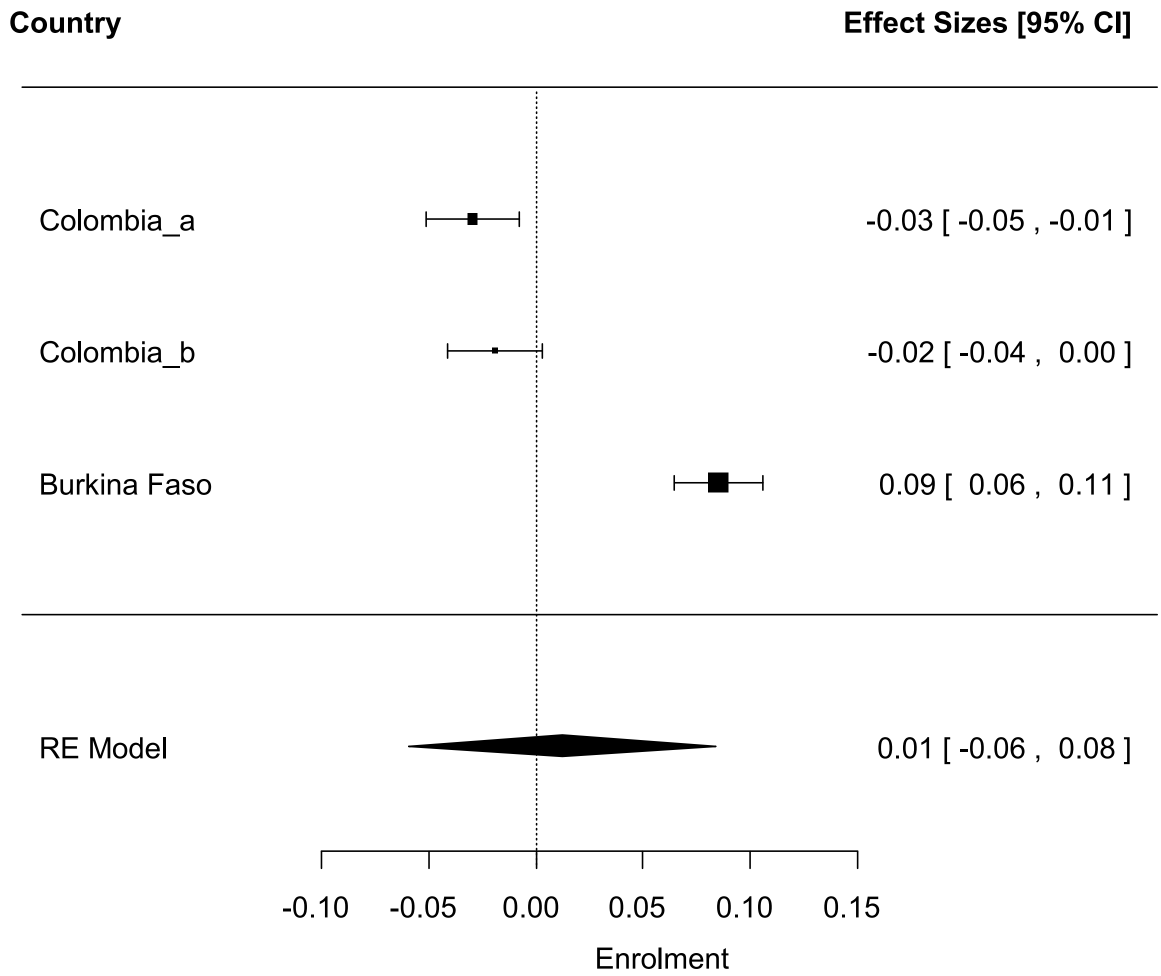
All twelve studies provided data for meta-analysis, but none of the studies measured all outcomes. The number of comparisons with effect sizes range from three for enrolment and attendance, to fourteen for maths outcomes. We included two studies of the BRIGHT school programme in the review (de Hoop & Rosati, 2012; Kazianga *et al.*, 2013). These studies are based on the same dataset, but largely reported on different outcomes and samples. However, both studies provide estimates for the effect on enrolment for the full sample, with de Hoop & Rosati (2012) also reporting effects on different subgroups. We assessed Kazianga *et al.* (2013) as having the lowest risk of bias and therefore included the estimates from that study in the meta-analysis.

All effect sizes are expressed as standardised mean difference (SMD), interpreted as the magnitude of the number of standard deviation changes in the outcome for the intervention group as compared to students in the comparison groups schools. SMD scores are interpreted as the number of standard deviation changes in the outcome.

Enrolment

Two studies measured enrolment, with one study representing independent samples from the same study (Rodriguez & Sanchez, 2010). The overall average effect is 0.01, 95% CI [-0.06, 0.08]. There is a large amount of variability between the studies, as indicated both by the homogeneity tests (heterogeneity (I^2 = 97.02%, τ^2 = 0.039, Q(df = 2) = 69.4115, p= < 0.0001) and by inspecting the forest plot in Figure 9a. The confidence intervals of the study from Burkina Faso do not overlap with the other two estimates, and as expected when removing this study from the analysis the overall results change to a small negative effect (SMD=-0.02, 95% CI [-0.04, -0.01]).

Figure 9 a: Enrolment[141]

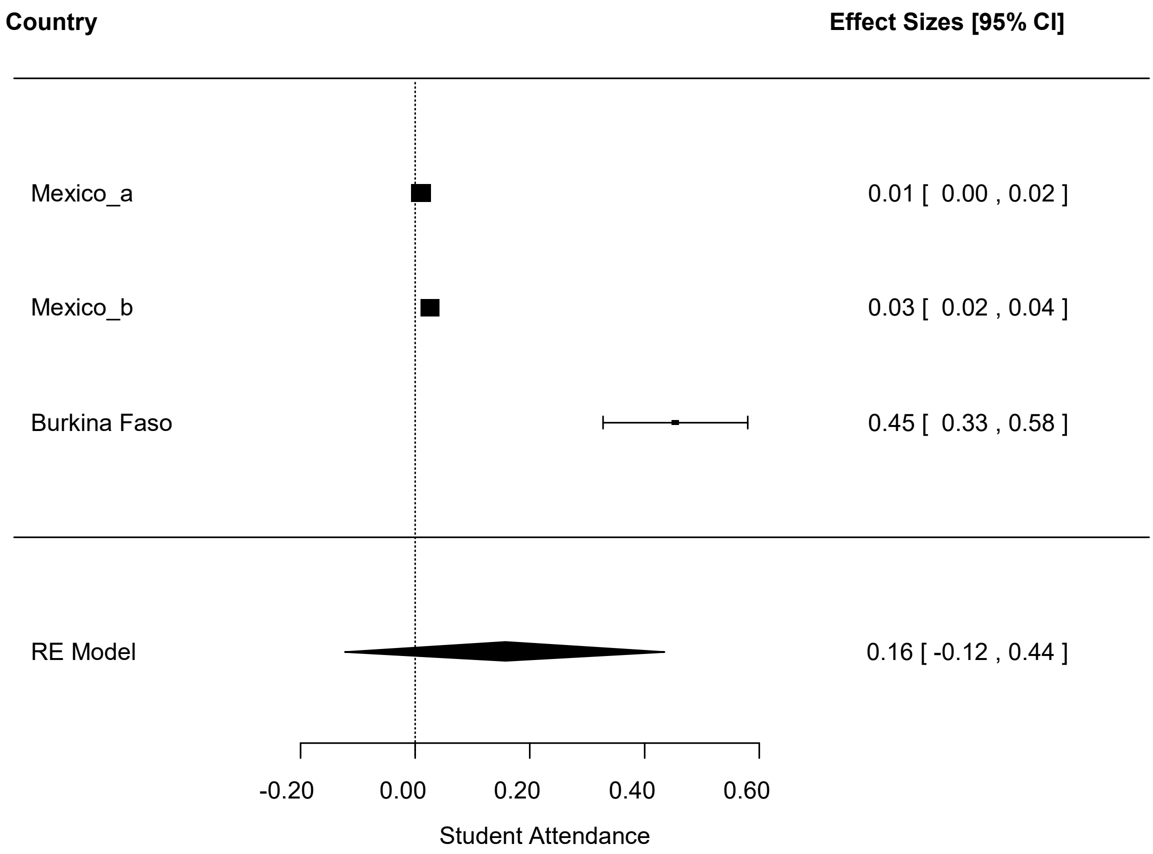


Attendance

The forest plot in Figure 9b below shows the results of our meta-analysis of multilevel interventions on attendance. Two of the three samples in this meta-analysis are from Rosati and Rossi (2007) which were separated by age group (8 – 11 years, and 12-16 years respectively). The overall average effect is 0.16 (95% CI [-0.12, 0.44]). However, there is a high degree of heterogeneity (I^2 = 99%, τ^2 = 0.059, Q(df = 2) = 49.74, p < 0.0001). There is a large difference in the magnitude of effects observed for the two sub-samples in Mexico where the study assessed the effect of the Compensatory Education Programme (CONAFE), and the Burkinabé Response to Improve Girls' Chances to Succeed (BRIGHT) Programme. While it appears the former programme had very small effects on attendance rates, the observed effect for the BRIGHT programme in Burkina Faso represents a relatively large improvement in attendance rates.

[141] In forest plot the labels Colombia_a and Colombia_b refer to the primary and secondary school samples in Rodriguez & Sanchez (2010) respectively. Burkina Faso refers to Kazianga et al (2013).

Figure 9 b: Attendance[142]

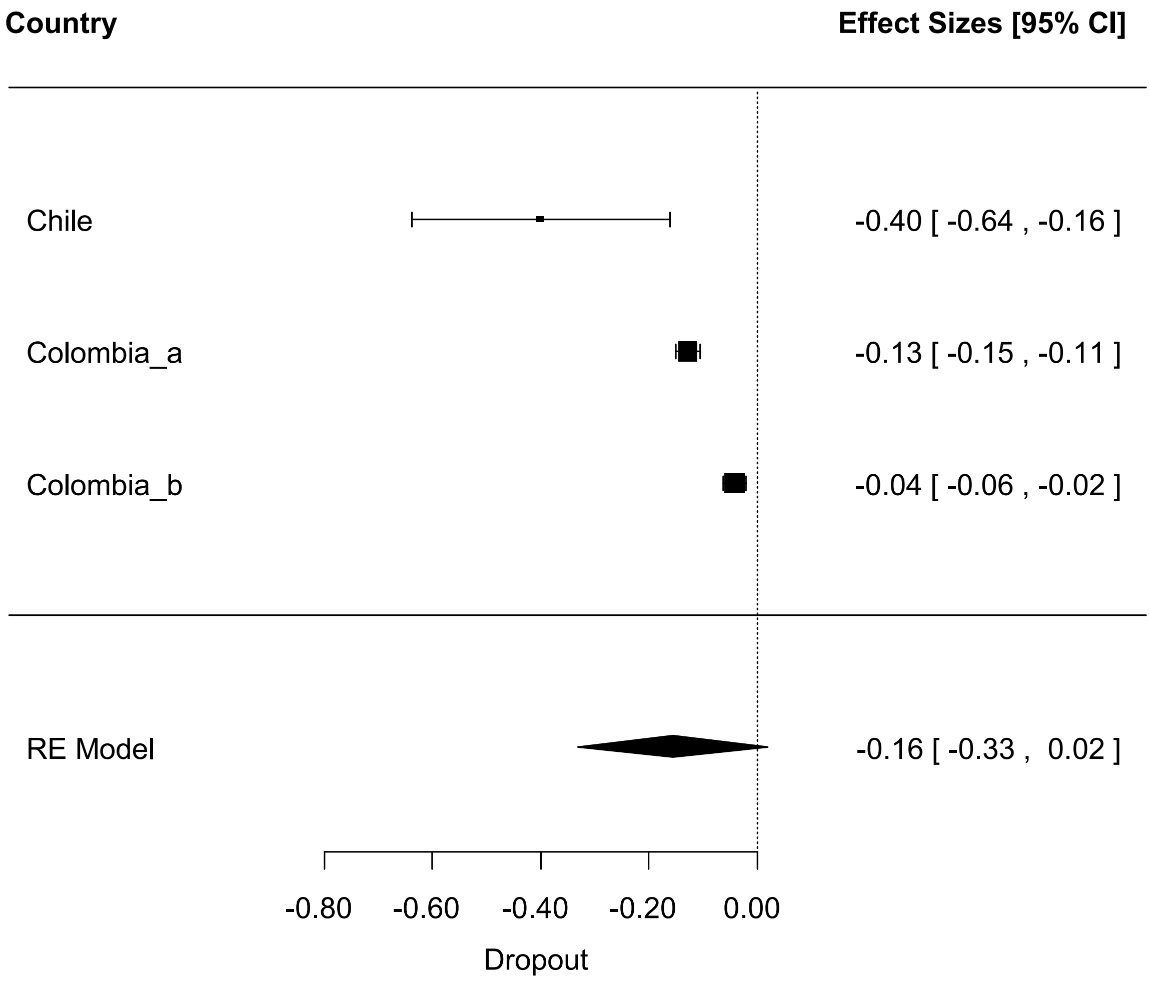


Dropout

Figure 9c shows the forest plot for the meta-analysis of the effects on drop-out rates (SMD=-0.16, 95% CI [-0.33, 0.02]). Two samples in the meta-analysis came from the same study, but represent two independent sub-samples (Rodriguez & Sanchez, 2010), providing estimates of dropout rates for primary and secondary levels separately. Both the forest plot and the tests of homogeneity suggest a large amount of variability between the two studies (I^2 = 99%, τ^2 = 0.0202, Q(2 df) = 36.33, p < 0.0001). While we observe a reduction in dropout rates overall and across all three estimates, the magnitude of the effect appears to be substantially larger for the Technical Support to Failing Schools (TSFS) Programme in Chile, as compared to The Rural Education Project (PER) in Colombia. As expected, the average effect is sensitive to the removal of the TSFS programme (Bellei, 2013). See Appendix H for results of all sensitivity analyses.

[142] In forest plot 9a, Mexico_a and Mexico_b refer to two different samples (primary and secondary schools, respectively) from Rosati & Rossi (2007).

Figure 9 c: Dropout[143]

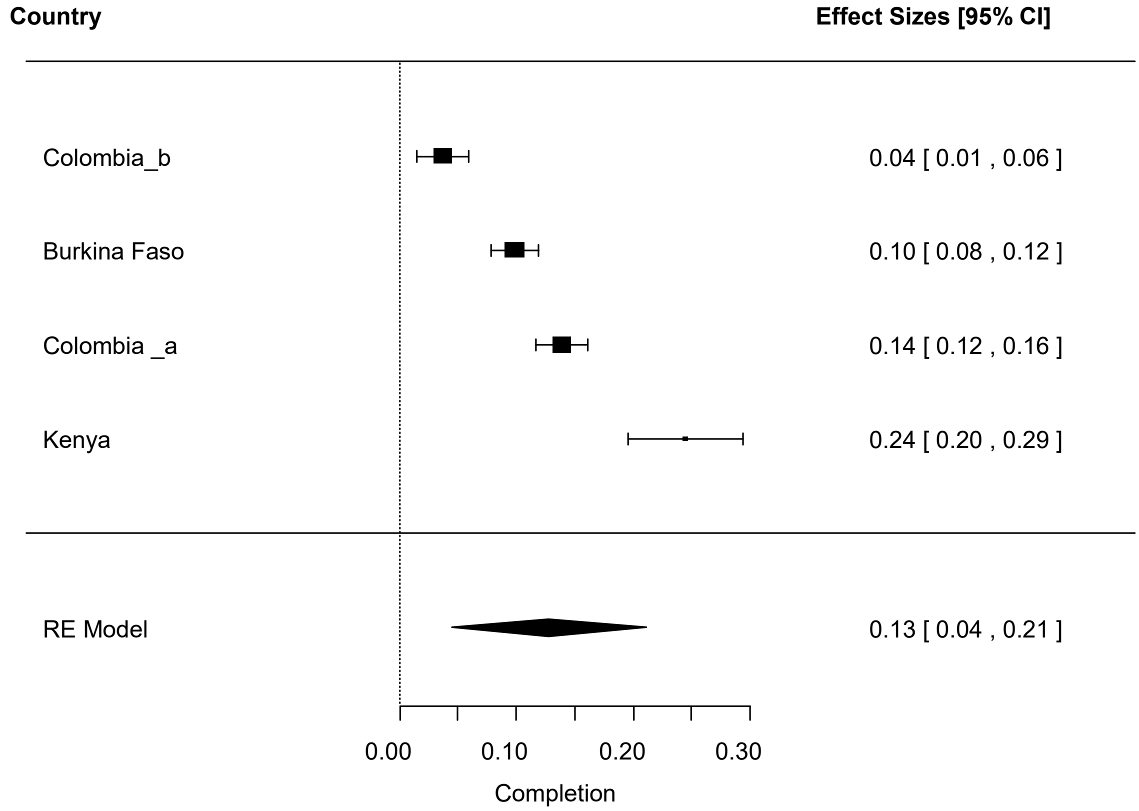


Completion

Two studies measured completion outcomes, and as above two samples represent independent samples from the same study (Rodriguez & Sanchez, 2010). The overall average effect of multilevel interventions on student completion is 0.13, 95% CI [0.04, 0.21]. The assessment of homogeneity suggests that the results do not arise from the same population (I^2 = 97.77%, τ^2 = 0.0070, Q(df = 3) = 77.1847, p-val = < .0001). As can be seen from the forest plot in Figure 9d, none of the confidence intervals overlap. The effect sizes range from 0.04, 95% CI [0.01, 0.06] in Colombia (Rodriguez & Sanchez, 2010, secondary school sample) to 0.24, 95% CI [0.20, 0.29] in Kenya (Kremer *et al.*, 2003). The average effect is not sensitive to the removal of any of the estimates.

[143] In forest plot the labels Colombia_a and Colombia_b refer to the primary and secondary school samples in Rodriguez & Sanchez (2010) respectively.

Figure 9 d: Completion[144]



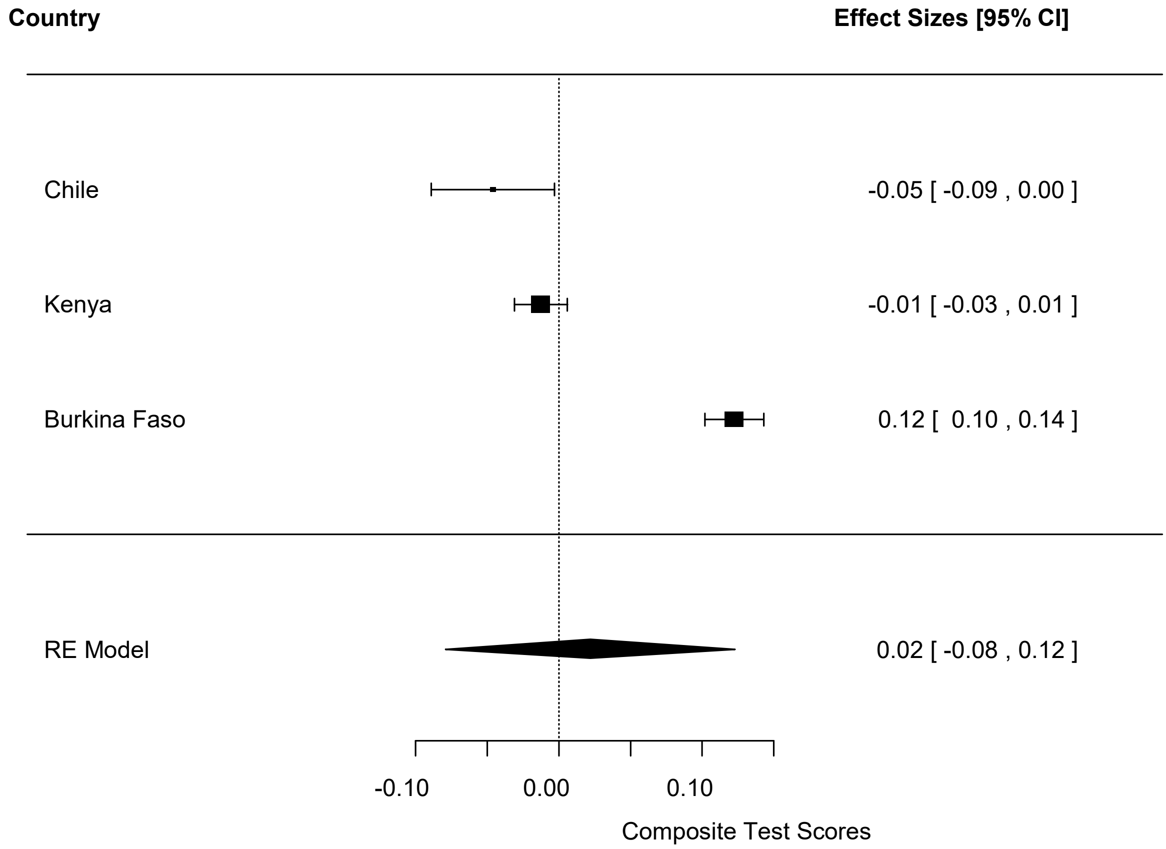
Composite Test Scores

Three studies reported on composite test scores (Kazianga *et al.*, 2013; Kremer *et al.*, 2003; Tokman, 2002) and the overall average effect is 0.02 (95% CI [-0.08, 0.12]). These results do not arise from the same population (I^2 = 98.03%, τ^2 = 0.0077, Q(df = 2) = 107.2985, p-val = < .0001). The effects range from -0.05 (95% CI [-0.09, 0.00]) in Chile to an improvement of 0.12 (95%CI [0.10, 0.14]) for the BRIGHT Programme. As can be seen from Figure 9e the confidence intervals of the latter study do not overlap with the others, and when removing this study the average effect remains small, but negative.

Kazianga *et al.* (2013) also reported effects separately for different age groups. The results suggest some variation in test score gains by age, ranging from 0.12 (SMD=0.12, 95%CI [0.07, 0.17]) for children aged twelve, to 0.29 (95%CI [0.23, 0.34]) for children aged nine. See Appendix H for the results of the meta- analysis on different sub-groups of participants.

[144] Colombia_b refers to Rodriguez & Sanchez, 2010 (Secondary school sample); Colombia_a refers to Rodriguez & Sanchez, 2010 (Primary school sample).

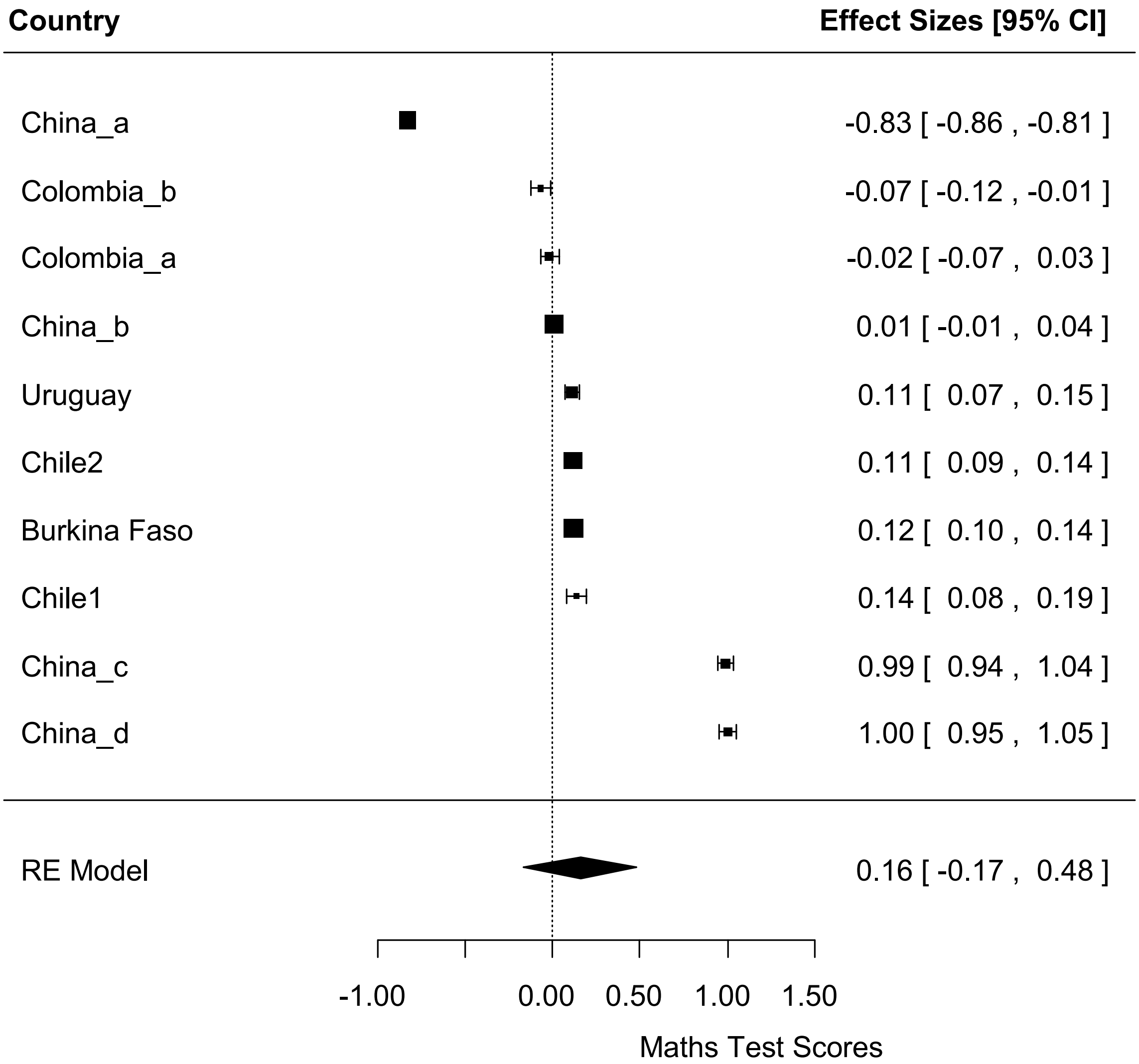
Figure 9 e: Composite Test Scores



Maths Test Scores

The overall average effect of multilevel interventions on maths test scores is 0.16, 95% CI [-0.17, 0.48]). There is a large amount of heterogeneity (I^2 = 99.89%, τ^2 = 0.2766, Q(df = 9) = 7511.6006, p-val < .0001), and as can be seen from the forest plot in Figure 9f there are several clusters of studies without overlapping confidence intervals. The effects range from -0.83, 95% CI [-0.86, -0.81] for the 3rd grade sample in the study from China (Min *et al.*, 2012) to 1.00, 95% CI [0.95, 1.05] 9th grade samples of the same study (Min *et al.,* 2012). Further, results of the sensitivity analysis showed that the results are sensitive to the removal of several of the different sub-samples from the China study. Removing the large negative estimate for the grade 3 sample (China_a Min *et al.,* 2012) increases the estimate to 0.27, 95% CI [-0.01, 0.54], whereas removing either the grade 7 (China_c) or grade 9 (China_d) estimates reduces the overall effect substantially. Considering the large magnitude of these estimates, this is not surprising.

Figure 9 f: Maths scores[145]



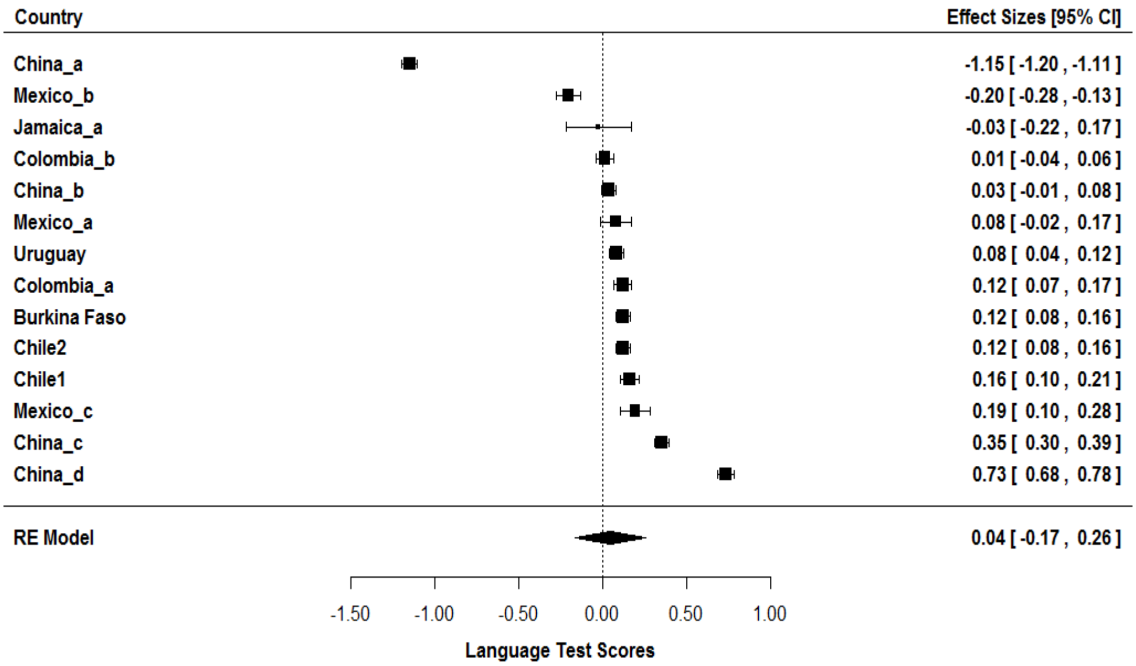
Language arts test scores

Eight studies, contributing fourteen different samples, evaluated the effects of multilevel interventions on language arts test scores. The overall average effect is 0.04, 95% CI [-0.17, 0.26]. As above, the assessment of homogeneity suggests that the results do not arise from a common population (I^2 = 99.56%, τ^2 =0.1629, Q(df = 13) = 3660.4901, p-val < .0001). The forest plot in Figure 9g supports the presence of heterogeneity. As above the outliers are sub-samples from the same study. The effects range from -1.15, 95% CI [-1.18, -1.13] for the 3rd grade sample and to 0.73, 95% CI [0.69, 0.78] for the 9th grade sample of the study by Min *et al.* (2012). Similarly to the meta-analysis of maths test scores, the results are sensitive to the removal of these outliers. Removing the large negative estimate for the grade 3 sample (China_a Min *et al.,* 2012) increases the overall estimate substantially (0.14,

[145] China samples all come from Min *et al*. (2012). China_a refers to grade 3 scores, which China_b, China_c and China_d refer to grade 5, 7, and 9 maths scores respectively. Rodriguez & Sanchez (2010) is the Colombia study, with samples a and b referring to primary and secondary school respectively. Chile1 is from Chay *et al*. (2005) and Chile2 from Bellei (2013).

95% CI [0.02, 0.26]). On the other hand, as above, removing either the grade 7 (China_c) or grade 9 (China_d) estimates reduces the overall effect substantially.

Figure 9 g: Language Arts Test Scores[146]



9.3 Summary of findings and discussion

We identified 12 studies evaluating the effects of multilevel interventions on education outcomes in L&MICs across eight different countries. The main unifying feature of the included programmes is that the interventions included more than one intervention component, with these intervention components falling into more than one of the five levels included in our conceptual framework. We were able to conduct meta-analyses of effects on enrolment, attendance, dropout, completion, composite test scores in maths and language arts.

The overall average effects range from 0.01, 95% CI [-0.06, 0.08] for school enrolment to 0.16, 95% CI [-0.17, 0.48] for maths test scores. The confidence intervals cross the line of no effect for most outcomes, although the average effect for completion is more precise (SMD=0.13, 95% CI [0.04, 0.21]), and it is also the largest in magnitude for that outcome observed across the review. Most of the meta-analyses include only a few studies and we observe a large amount of heterogeneity across all outcomes. Therefore the average effects should be interpreted with caution. However, the overall estimates include several outliers where the authors report substantial benefits for participants. In particular we observe large effects of the BRIGHT programme in Burkina Faso and the Southwest Basic Education Project (SBEP) in China. For instance, the two studies evaluating the BRIGHT programme in Burkina Faso (De Hoop & Rosati, 2012; Kazianga, 2013) suggest large, positive effects for enrolment, attendance, completion, composite scores, maths scores and language arts scores. We were unable to conduct moderator analyses to explore the reasons for the observed heterogeneity because of the low number of studies included in any one meta-

146 China samples all come from Min *et al.* (2012). China_a refers to grade 3 scores, which China_b, China_c and China_d refer to grade 5, 7, and 9 language arts scores respectively; Chile1 is from Chay *et al.* (2005) and Chile2 from Bellei (2013); Mexico samples both come from Paqueo & Lopez-Acevedo (2003). Mexico_a refers to the indigenous sample and Mexico_b refers to the urban sample and Mexico_c refers to ; Rodriguez & Sanchez (2010) is the Colombia study, with samples a and b referring to primary and secondary school respectively.

analysis. The evidence on process and implementation was also limited, but the qualitative synthesis provides some suggestions for why programmes were more successful in some cases.

The programmes classified as multilevel are typically complex to implement, with a range of different components delivered at once. The qualitative evidence suggests that in the case of the BRIGHT programme implementation went more or less as planned. However, for other programmes there were reports of issues with implementation. For instance, in the TSFS programme in Chile negative connotations associated with being labelled an underperforming school resulted in resistance from schools to participate in the programme. Stakeholders also objected to the programme design, which placed emphasis on student results over education in general. These issues resulted in delays in effective implementation in some cases.

The findings on the effects of the BRIGHT programme also contrast with the findings of the study of a programme with similar components, the CONAFE programme in Mexico. In this case the authors observed effects on attendance that are close to zero (SMD=0.01, 95% CI [0.00, 0.02] and 0.03, 95% CI [0.02, 0.04] for primary and secondary schools respectively) (Rosati and Rossi, 2007)[147]. The difference in effects may be explained by the educational contexts of the two different countries in which these programmes took place. While Mexico had a net primary school attendance rate of near 100 per cent between 2008 and 2012, Burkina Faso's net primary school attendance rate was around 50 per cent in the same period (UNICEF, 2013a & b). In other words, much larger gains in attendance rates were to be made in Burkina Faso as compared to Mexico.

Multilevel programmes may provide substantive benefits for children's school participation and learning in some contexts. Future studies should assess whether the large effects of some programmes can be replicated in other contexts and aim to identify the conditions under which multilevel programmes are most effective.

Table 9 d: Descriptive findings- Process and Implementation

Descriptive Findings: Process and Implementation	Context	Citation/ info type
Opinions about the programme were generally positive	Uruguay	Cerdan-Infantes and Vermeersch, 2007 (Impact Evaluation)
The implementation of the extension of the school day, school building and teacher training was practically universal in participating schools	Uruguay	Cerdan-Infantes and Vermeersch, 2007 (Impact Evaluation)
The programme faced financial issues, including delays in budget approvals	Mexico	Paqueo and Lopez-Acevedo, 2003 (Impact Evaluation)

[147] Unfortunately, Rosati and Rossi (2007) do not evaluate the effects of the CONAFE programme on other outcomes.

The programme was not well implemented in the urban schools	Mexico	Paqueo and Lopez-Acevedo, 2003 (Impact Evaluation)
Headmasters increased school fees of CSP schools	Kenya	Kremer *et al.,* 2003 (Impact Evalution)
There was a lack of supervisors, resulting in complaints from schools about inadequate support	Chile	Undurraga, 1994 (Process Evaluation); Guttman, 1993; Carlson, 2000 (Project documents)
Some schools were withdrawn early from the programme	Chile	Undurraga, 1994 (Process Evaluation)
The programme changed year on year	Chile; Jamaica	Tokman, 2002; Chay, 2005; Lockheed, 1999 (Impact Evaluations)
Teacher workshops were often substituted with other activities when schools left the programme	Chile	Ugarte, 2011 (Impact Evalation – not included)
While student support was well developed, the educational management component was not as effective	Chile	Ugarte, 2011 (Impact Evalation – not included)
There was resistance from schools to participate in the TSFS programme in Chile, causing delays in implementation	Chile	Sotomayor, 2006 (project document); Sotomayor and Dupriez, 2007 (qualitative study)
The BRIGHT programme was mainly implemented as intended, although some villages which were selected to receive a BRIGHT school did not because of poor infrastructure	Burkina Faso	Kazianga *et al.,* 2013; de Hoop *et al.,* 2012 (Impact Evaluations)
Despite the largely unproblematic implementation and generally better quality of schools, the long-term progress of BRIGHT schools may have been negatively affected by lack of maintenance	Burkina Faso	Kazianga *et al*., 2013 (Impact Evaluation)
In some cases schools were underused whereas in others schools were oversubscribed	Burkina Faso	Kazianga *et al*., 2013 (Impact Evaluation)

Table 9 e: Descriptive findings- Intermediate outcomes

Descriptive Findings: Intermediate outcomes	**Context**	**Citation/ Info Type**
There reportedly was a successful change in behaviour within schools	Chile	Sotomayor and Dupriez, 2007 (qualitative study)
While the number of classes offered at programme schools increased only modestly, the programme led to substantial increases in class size	Kenya	Kremer *et al.,* 2003
There is some evidence to suggest that the BRIGHT programme changed parents' attitudes towards education	Burkina Faso	Levy *et al*., 2009 (Impact Evaluation)

10. Summary of findings and conclusions

The findings presented in this report summarise the findings of the most comprehensive systematic review of education in L&MICs conducted to date. We summarise the findings of 238 studies evaluating the effects of a range of different education programmes in 52 different L&MICs. This includes 59 studies from Sub-Saharan Africa, 38 studies from East Asia & the Pacific, 87 from Latin America & the Caribbean; 51 from South Asia, two from Middle Eastern & North Africa and one from Europe & CIS. Based on reported sample sizes we estimate the studies include data from almost 20 million children. This chapter provides a summary and discussion of the findings of the review.

10.1 Summary of findings

The average estimates for all included primary outcomes are provided in table 10a below and this section summarises the results by intervention level.

10.1.1 Child level interventions

School feeding

School feeding programmes typically provide children with a meal or snack at school or to take home, and aim to increase school enrolment and attendance, alleviate children's short-term hunger and improve their nutrition and health. We identified 16 studies that evaluated the effect of a school feeding programme implemented in 14 different countries.

The available evidence suggests a beneficial effect of school feeding on the children in our sample for a range of different outcomes. The overall average effects range from 0.01 for completion (95% CI [-0.03, 0.01) to 0.14 for enrolment (95% CI [-0.05, 0.33]) and 0.14 for composite test scores (95% CI [-0.04, 0.33]). Results also suggest positive and significant effects for student attendance (0.09, 95% CI [0.03, 0.16]), language test scores (0.09 (95% CI [0.01, 0.17]) and maths test scores (0.10 [95% CI, 0, 0.19]). However, there is a large amount of variability for most overall estimates and the results should therefore be interpreted with caution. For most outcomes the magnitude of the effect is sensitive to the inclusion the study of the Hinterland Community-Based SFP in Guyana, where the observed effects are substantially larger than all other estimates (Ismail et al., 2012).

The qualitative synthesis suggest a possible reason for some of the observed heterogeneity is that school feeding programmes may be more effective in improving education outcomes in contexts such as Guyana with high food insecurity and low existing school participation. Conversely, in contexts such as Chile, where most extreme child malnutrition has been eliminated and enrolment rates are already high effects may be smaller in magnitude (McEwan, 2013; Altman, 2013; He, 2010).

School-based health

We included 16 studies that evaluated the effect of a school-based health programme across eight different countries. The studies evaluate a range of different interventions, including de-worming, malaria prevention and control, micronutrient supplementation, provision of eye glasses and provision of incentives for anaemia reduction.

Overall, the findings of our analysis suggest malaria prevention and control programmes may have some positive effects on education outcomes, although these are small in magnitude and estimates are imprecise. The overall average effects range from 0.01, 95% CI [-0.05, 0.12] for cognitive test scores to 0.16, 95% CI [-0.08, 0.25]) for maths test scores. Similarly, the results for micronutrient interventions also suggest there may be a beneficial effect on education outcomes with the overall average effect ranging from 0.01 SMD, 95% CI [-0.03, 0.05] for cognitive test scores to 0.06 (95% CI [0.02, 0.10]) for maths test scores. The meta-analyses of studies assessing de-worming programmes suggest small, if any observable benefits for children receiving such programmes. There may be small positive effects on attendance in some contexts, but the average effects are small and imprecise for all outcomes, apart from nutrition, where we observed a negative effect (SMD=-0.26 SMD, 95% CI [-0.43, -0.10]).

The results indicate that overall SBH interventions may be beneficial, but the average effects are small in magnitude, and for some outcomes we observed negative or no effects. Few estimates are statistically significant and are mostly based on only a few studies. We identified very little additional process or qualitative evidence to help explain these findings. However, the meta-analysis of nutrition outcomes (designated a secondary outcome for SBH interventions) suggest that for the programmes evaluated in many of the included micronutrient and de-worming studies there was limited improvement of nutrition outcomes, and in some cases the effect was even negative.

Merit-based scholarships

Merit-based scholarships aim to improve learning outcomes by rewarding high performing students with scholarships to continue their study or by providing one-off cash payments (McEwan, 2013; Berry, 2013). We included 11 studies covering nine unique programmes that evaluated the effects of such interventions.

Overall, the analysis suggest merit-based scholarship programmes can have positive effects on education outcomes, in particular for learning outcomes. No studies evaluate effects on enrolment, and the results for attendance and drop-out are small and only based on a few studies. Results for learning outcomes are more promising, with overall positive and effects for maths test scores (SMD=0.11, (95% CI [0.03, 0.20]) and composite test scores (SMD=0.10 SMD, 95% CI [0.03, 0.17]). The magnitude of effect is smaller and less precise for language arts outcomes (SMD=0.04, 95% CI [-0.07, 0.15]). There was a high degree of heterogeneity for most of the average estimates. In some contexts merit-based scholarships led to improvements that were considerably larger in magnitude than the average effect, and in some cases the direction of the effect was negative.

Each meta-analysis is based on a small number of studies, however the evidence suggests that children receiving merit-based scholarships benefit from an improvement in test scores on average, with potential for improvements that are relatively large in magnitude, as observed in the peer incentives programme in China (Li et al., 2014) and the merit-based scholarship programme in Cambodia in particular (Barrera-Osorio & Filmer, 2013).

Providing information

Providing information to children and parents about the potential future benefits of education in terms of income, employment, and social status is thought to increase school participation, enrolment and continuation (Nguyen, 2008). We included four studies that evaluated the effects of providing information to children and/or or their parents, covering four countries in East Asia, Latin America and the Caribbean and Sub-Saharan Africa.

We were unable to conduct meta-analyses for any of these as no two studies reported on the same outcome measure. The observed effects are mostly small and in a few cases negative, although the study of a programme providing information about returns to education to both children and parents in Madagascar found relatively large effects on school attendance (Nguyen, 2008).

10.1.2 Household level interventions

User fee elimination

Programmes reducing or eliminating school user fees aim to improve access to education by removing all or some of the direct costs of schooling, for instance by providing school uniforms for free or through the elimination of tuition fees, as has been done in many African countries over recent decades (Bentaouet-Kattan, 2006). We included ten studies that evaluated the effects of this type of programme in ten different countries. Four programmes removed tuition fees for certain groups of students, four programmes aimed to reduce the cost of tuition and two studies report on programmes that provided school uniforms.

The findings suggest the average effect of user fee reduction programmes on school participation is small, but with a large amount of heterogeneity. We found an overall positive effect of user fee reduction (SMD=0.09, 95% CI [-0.04, 0.21]) on enrolment but the confidence intervals cross the line of no effects. The results are sensitive to the removal of Adelman’s (2015) evaluation of the tuition Waiver Programme in Haiti which had a very large positive effect on enrolment (SMD= 0.61, 95% CI [0.50, 0.73]). The analysis of attendance suggest no effect (SMD=0.01, 95% CI [-0.13, 0.15], but this is based on only two studies and heterogeneity is high. Effects on dropout and completion are larger in magnitude, confidence intervals cross the line of no effects (SMD=-0.10, 95% CI [-0.23, 0.02]; SMD=0.15, 95% CI [-0.11,0.42]).

Cash transfers

Cash transfers are social safety-net programmes that provide a direct transfer of cash to mothers, households or children. Cash transfer programmes are typically classified into two main categories. Conditional Cash Transfers (CCTs), transfer money to households or children conditional on certain behaviour, such as school enrolment and attendance above a certain rate. Unconditional Cash Transfers (UCT), also include a money transfer, but do not come with any explicit conditions (Baird et al., 2010). We included 50 studies that evaluated the effect of 38 unique programmes.

We observe consistently positive effects across school participation outcomes, and the magnitude of the effects are the largest across the review. Effects range from 0.11, 95% CI [0.07, 0.15] for enrolment, where we included 48 effect sizes, to 0.13, 95% CI [0.08, 0.18] for attendance, where we included 38 effect sizes. The average effects for dropout also suggest a reduction in dropout rates (-0.12, 95% CI [-0.16, -0.07]) and we also see an average improvement in school completion rates (0.12, 95% CI [0.01, 0.22]). As expected, cash transfers do not appear to improve student learning outcomes; the effect on maths test scores was 0.01, 95% CI [-0.07, 0.05] and no different from zero for language test scores (0.00, 95% CI [-0.04, 0.04]).

While the results are fairly robust, there is considerable heterogeneity and for all of our outcomes we observe effects that are both substantially larger and smaller than the average pooled effects. Moderator analyses suggest effects on enrolment and attendance increase with intensity of conditions.

10.1.3 School level interventions

Structured pedagogy interventions

‘Structured pedagogy’ interventions typically seek to introduce new content and instructional approaches by developing new curricula and providing teachers with training in delivering this material, often together with materials for both teachers and children. We included twenty-one studies that evaluated this type of intervention, implemented in 12 different countries.

There is relatively strong evidence for the beneficial effects of structured pedagogy interventions on maths and language outcomes. The meta-analysis for language outcomes includes effects from 17 studies, many of them large scale cluster RCTs, and results show an overall effect of 0.23 (95% CI [0.13, 0.34]). The overall effect on maths test scores is slightly smaller in magnitude (SMD=0.14, 95% CI [0.08, 0.20]). This is not surprising as most of the evaluated programmes focus on literacy. Few of the studies in this category evaluate effects on school participation, cognitive outcomes and composite test scores.

Computer-Assisted Learning

Computer assisted learning interventions (CAL) use computers, either in the form of lap-tops or computer labs, to aid or support children's learning. In some cases the main focus is simply on providing children with access to computers while in other cases they are delivered as an integrated package together with new content and instructional approaches, and training for teachers. We identified 18 studies that evaluated the effect of 16 unique programmes implemented in nine countries.

Based on the studies included in the review it is not clear that the overall effect of CAL on children's learning is beneficial. The overall average effect on children receiving CAL interventions range from -0.01 SMD for language test scores (95% CI [-0.08, 0.05]) to 0.07 SMD for maths test scores (95% CI [0.02, 0.11]). There is a large amount of heterogeneity of effects and we observe both substantive positive and negative effects in different contexts. This may be partially explained by evidence from the qualitative synthesis which suggested that programmes in Chile, Colombia, Peru, Uruguay, Mexico and Nepal all faced process and implementation issues, and in some cases there was a lack of integration of CAL into existing approaches. (Barrera-Osorio et al. 2009; Cristia et al. 2012; Cristia et al., 2013; David and Quispe 2013; De Melo et al. n.d; ; Imbrogno 2014; Sharma, 2014).

Remedial education

Remedial education refers to a range of interventions aimed at improving learning outcomes for students who are lagging behind their peers in normative standards of achievement. We included four studies evaluating the effect of programmes providing tailored assistance to a group of students, implemented in three settings in South America and South Asia. The studies did not assess effects on school participation outcomes, so the results are limited to learning outcomes.

We found relatively large average effects of remedial education interventions on test scores, however for all estimates the confidence intervals cross the line of no effect. We observed the largest effect for composite test scores (0.22 SMD, 95% CI [-0.09, 0.53]), with slightly smaller overall effects for language (0.16 SMD [-0.08, 0.41]) and maths (0.19 SMD [-0.05, 0.44]). In each case, the overall effects were sensitive to the removal of one study (Lakshminarayana et al., 2013), an evaluation which reported on the STRIPES trial in India.

The small number of identified studies and the large amount of heterogeneity suggest more research is needed to identify the programme components that may produce such large effects and whether these effects can be replicated in different contexts.

Providing materials

Studies included in this category include evaluations of any intervention providing 'traditional hardware' materials for schools such as books, chalkboards or other classroom equipment. We identified four studies of programmes that provided learning materials across three different countries. All programmes in this category rest upon the general idea that the provision of educational inputs where they are scarce will improve educational outcomes (Glewwe et al., 2007, 2009; Glewwe et al., 2000, 20004; Das et al., 2013; Sabarwal et al., 2014).

Overall we find a lack of evidence of a beneficial effect of additional materials on student outcomes, although this based on few studies and should therefore be interpreted with caution. The overall average effects range from -0.02 for maths test scores (95% CI [-0.06, 0.02]) to 0.01 for composite test scores (95% CI [-0.01, 0.02]).

This may be explained by issues with implementation of these programmes and lack of use of text books once they arrived, as identified in the qualitative synthesis. Other interventions that include provision of materials together with other components, notably structured pedagogy, have led to improved outcomes. Thus one plausible explanation is that, in addition to there being issues with poor implementation, materials is a necessary but not sufficient condition for children's learning.

New schools and infrastructure

New schools and infrastructure (NSI) interventions are programs involving the construction of schools in areas where there were none previously, or improvement or rehabilitation of existing school infrastructure. The interventions evaluated in the studies included for this category are very diverse, and so we grouped them into three sub-categories for the purposes of analysis. Two studies included the provision of latrines and were classified as hygiene infrastructure interventions. Two studies focused on the establishment of new, community-based schools in underserved areas, and finally three reported on the improvement or replacement of school infrastructure, including the provision of a library and rehabilitation of existing infrastructure.

Overall, the small pool of studies on new schools and infrastructure suggested a beneficial effect on school participation, as measured by enrolment and attendance, but these results should be interpreted with caution as they come from a small number of studies. Construction of new schools may improve enrolment and attendance in some contexts, with large improvements observed in a context of low school participation in Afghanistan, particularly for girls (enrolment: SMD=0.38, 95% CI [-0.29, 1.04]; attendance: SMD=0.08, 95% CI [-0.04, 0.19]), with effects of smaller magnitude observed in Niger. There may also be beneficial effects of construction of new schools on maths and language arts (maths: SMD=0.19, 95% CI [-0.15, 0.53]; language arts: SMD=0.02, 95% CI [-0.01, 0.05]).

The results of the two hygiene infrastructure studies in India and Kenya suggested that these interventions can have a positive effect on school participation (enrolment, 0.11 SMD, 95% CI [0.01, 0.20] and student attendance 0.14 SMD, 95% CI [0.05, 0.24]). The effect of improving existing infrastructure is not clear from the two included studies for which we were able to use data.

Effects of interventions grouping students by ability

We identified one study that evaluated the effect of grade retention and two studies evaluating the effect of tracking interventions. Though these studies do have some characteristics in common, they are still very different and therefore we present the evidence for each of these types of interventions separately.

The single grade retention study reported on a program in China (Chen et al., 2010) and provided outcomes data for maths and language test scores. The effect size for both reported outcomes indicate a negative and statistically significant effect for grade retention on children's learning. The effect on Chinese test scores was -0.10, 95% CI [-0.17, -0.02], while for maths test scores it was also negative and larger in magnitude (SMD = -0.19, 95% CI [-0.26, -0.11]).

The evidence on tracking interventions is also limited, with only two included studies from Kenya and India. The average effects are not statistically significant and range from 0.02 for maths to 0.12 for language arts.

Extra time in school

Extra time programmes aim to provide a longer school day with increased learning time for students. Typically, these programmes abolish 'shift' schooling whereby two separate cohorts attend the same school in a given day, one in the morning and one in the afternoon, and expand existing infrastructure so that all children can attend a full school day. We included three studies that evaluated the effect of increasing time spent in school, from Chile and Ethiopia.

The studies did not assess effects on school participation outcomes, so the results are limited to learning outcomes. The overall pooled effect is 0.19, (95% CI [0.15, 0.24]) for language arts outcomes and 0.09, 95% CI [-0.04, 0.22]) for maths, although this is based on only two studies. Nevertheless, these interventions suggest an effect on learning outcomes which is large in magnitude compared to the other intervention types.

10.1.4 Teacher level interventions

Teacher incentives

Teacher incentive interventions seek to improve the working conditions in schools so that teachers are motivated to come to work and improve their performance. Such interventions take many forms, such as providing direct payments to teachers based on their attendance or on the achievement of their students, and teacher surveillance and monitoring (Glewwe et al., 2008; Cueto et al. 2008). We included ten studies that evaluated the effect of a teacher incentive intervention covering programmes in Latin America and the Caribbean, Sub-Saharan Africa, South Asia and East Asia and the Pacific.

Overall, the results suggest teacher incentive may produce some positive outcomes, many fail to improve outcomes of interest, while in some cases effects are negative. The overall average effects range from -0.04, 95% CI [-0.09, -0.02] for teacher performance to 0.08, 95% CI [0.02, 0.15] for maths test scores. Our analysis suggests teacher incentive programmes did not on average improve teacher attendance, student attendance or drop-out and only observe small effects on enrolment and completion. The findings on learning outcomes are mixed, with an average improvement in maths test scores (SMD=0.08, 95% CI [0.02, 0.13]), but no effects on language arts (SMD = 0.00, 95% CI [-0.13, 0.12]). The qualitative synthesis suggest limited teacher accountability structures and lack of information about the programme may explain why the interventions had limited effects on teacher effort and student learning, in particular in the first year of implementation.

Teacher hiring interventions

Teacher hiring interventions are designed to increase the number and quality of teachers in schools. We included eight studies evaluating eight unique programmes in countries in South Asia, Sub-Saharan Africa and Latin America and the Caribbean. The interventions that we identified in this category are varied, using a variety of mechanisms to achieve their goals.

Interventions hiring additional contract teachers may have beneficial effects on student outcomes. The overall average effects comparing additional additional contract teachers to business as usual range from 0.04, 95% CI [0.01, 0.08] for completion, to 0.06, 95% CI (-0.01, 0.12) for composite test scores, 0.06, 95% CI (0.03, 0.10) for language arts test scores and 0.10, 95% CI (0.00, 0.20) for maths test scores. These results should be interpreted with caution as they are based on just a few studies and are imprecise.

Three studies compared contract teachers to civil-service teachers, either directly or indirectly. Of these, two studies reported that overall performance of contract teachers was superior to that of civil-service peers. Two studies examined competitive teacher recruitment interventions and reported on drop-out and maths and language test scores (Estrada, 2013; Ome, 2012) and these findings were mixed.

We found a large amount of between study variability for most outcomes. This degree of heterogeneity is not surprising given the variability in the intervention design, context and populations in our included studies. The qualitative synthesis suggest there may be challenges in the implementation of teacher hiring interventions as they may threaten existing jobs or provoke opposition because they may mean lower pay, fewer privileges and less job security.

Teacher training

Teacher training is designed to develop teachers' knowledge and skills with the goal of improving the standard of teaching and ultimately of also improving student performance. We included a single study that reported on a teacher training programme, that is, teacher training not combined with any other major intervention components.

The single study reported the effect of the Learning to Read in a Healing Classroom (LRHC) teacher training programme in the Democratic Republic of Congo on learning outcomes. All the effects are positive, but relatively small in magnitude (Early Grade Maths Assessment (EGMA): SMD = 0.04, 95% CI [-0.02, 0.11]; Geometry test: SMD = 0.07, 95% CI [0.01, 0.14]); Language: SMD = 0.08, 95% CI [0.02, 0.15]).

Diagnostic feedback

Diagnostic feedback interventions use 'low-stakes' student tests to provide teachers with information on student achievement that will enable them to target their efforts in the classroom more effectively. 'Low stakes' tests have been described as 'assessments for learning' and can be compared to 'high stakes tests' which can be described as 'tests of learning'.

We included two studies that evaluated the effect of diagnostic feedback to teachers, both taking place in India. The meta-analyses for language arts (SMD = 0.01, 95% CI [-0.01, 0.05]) and maths (SMD = 0.01, 95% CI [-0.01, 0.03) suggest effects of a small magnitude and both estimates are imprecise.

10.1.5 System level interventions

School-based management

School-based management (SBM) interventions include initiatives which decentralise authority to the school level to improve the quality of school administration and leadership. They may involve handing decision-making on budget, staffing and curriculum development over to teachers, parents, students or other community members (Barrera-Osorio, 2009). We identified fifteen studies of SBM from six different countries in Latin America, East Asia and Sub- Saharan Africa.

The overall average effects range from -0.01, for language test scores and composite test scores (95% CI [-0.07, 0.05]; 95% CI [-0.10, 0.08]) to 0.05, 95% CI [0.00, 0.09] for completion. For some outcomes there is a large amount of between study variability. Many of the average effects include examples where programmes have had large and substantively important (positive and negative) effects on children's access to education and learning. For instance, the studies of two programs implemented in the Philippines (the Third Elementary Education Project and BESRA program) suggest consistently positive effects (ranging from 0.12, 95% CI [0.05, 0.19] for math to 0.16, 95% CI [0.08, 0.23]), with confidence intervals not overlapping with the other studies. On the other hand, the studies implemented in different countries in Sub-Saharan Africa consistently showed zero or negative effects across most outcomes.

It is not clear why we observe this difference in effects. The qualitative synthesis suggest existing levels of social and human capital among school stakeholders may influence the extent to which SBM interventions lead to improved education outcomes.

Community-based monitoring

Community based monitoring (CBM) interventions seek to provide information and improve the representation of communities in which service providers, governments, or other public bodies operate (Westhorp et al., 2013). We identified eleven studies of CBM programmes across seven different countries.

CBM initiatives may have a beneficial effect children's education outcomes, but the effects are relatively small in magnitude, in particular for attendance, drop-out and completion. The overall average effect ranges from 0.04 for student attendance (95% CI [-0.1, 0.18]) to 0.17 for student enrolment (95% CI [0.08, 0.25]). However, there is considerable heterogeneity across studies and the presence of outliers indicates that CBM programs may have larger effects in some settings. For example, the report card experiment in Pakistan (Andrabi et al. 2013) reports large positive effects on enrolment and learning outcomes, and Bjorkman's (2005) evaluation of the newspaper campaign to reduce corruption in Uganda found relatively large effects on composite test scores as compared to the other studies in the meta-analysis.

The qualitative synthesis suggests community human capital, lack of teacher responsiveness and extent of power of parent teacher associations may be some of the factors that influence the effectiveness of community monitoring initiatives in education.

Private-public partnerships

Public private partnerships and private provision of schooling (PPP) may seek to increase parents' and students' choice, provide a supply of schooling when there is none, or improve the quality of education provided (Barrera-Osorio et al., 2009). We identified twelve studies of PPP programmes across seven different countries.

Overall, the results suggest PPP may improve participation and learning outcomes. However, for most outcomes the effects are relatively small in magnitude and imprecise. The overall average effect range from 0.04, 95% CI [0.00, 0.09] for language arts to 0.12, 95% CI [0.00, 0.24] for enrolment. There is also a large amount of between study variability for all outcomes and the presence of outliers indicates that programs involving the private sector in providing education may have larger effects in some settings.

10.1.6 Multilevel interventions

We identified a number of studies that evaluate programmes that incorporate a number of different intervention components. We included 12 studies reported in 14 different papers that evaluated the effect of multilevel interventions. These are a diverse group of interventions in terms of design, and the only real unifying feature of the included programmes is therefore that the interventions come as part of a package whose individual components tackle different barriers to education.

We were able to conduct meta-analyses of effects on enrolment, attendance, dropout, completion, composite scores, math scores and language arts. The overall average effects range from 0.01, 95% CI [-0.06, 0.08] for school enrolment to 0.16, 95 % CI [-0.17, 0.48] for maths test scores. The average effects are relatively large in magnitude, as compared to other intervention areas in this review. However, apart from completion rates the confidence intervals of all pooled effects cross the line of no effect.

Most of the meta-analyses include only a few studies and we observe a large amount of heterogeneity across all analyses. Apart from the analysis of completion, all average estimates are sensitive to the removal of studies with particularly large effects. The average effects should therefore be interpreted with caution.

10.2 Overall completeness of the evidence

While we identified a large number of studies for inclusion we also reviewed a broad substantive area. Thus, for most of the interventions covered in the review the findings are based on a relatively small number of studies. Cash transfers, structured pedagogy and to some extent computer assisted learning are notable exceptions. The existing evidence is relatively extensive for these types of interventions. In most other areas the evidence base is not sufficiently extensive to draw strong conclusions. Further studies would help improve confidence in findings about effects, and also help researchers identify which intervention design, implementation and contextual factors are important for intervention success and failure.

The evidence is however particularly limited in some areas. There appears to be a major gap in the evidence on effects of teacher training programmes. Some of the school level interventions include an element of teacher training, but this typically take the form of a few days training focused on delivering specific content as part of a structured pedagogy package or computer assisted learning. We were not able to identify any studies that evaluate programmes aiming to train new teachers. Other areas where the evidence base is particularly limited include remedial education, school based health programmes (malaria, de-worming, micronutrients), different approaches to teacher hiring, new schools and infrastructure, tracking by ability, diagnostic feedback, providing information to children and parents about the returns to education, and finally extending the school day.

Moreover, the geographical coverage of studies is uneven. We identified a large number of studies from India and China in particular. Other countries with several studies include

Kenya, Brazil, Mexico, Chile, South Africa and Uganda. For most countries in Sub-Saharan Africa we identified few or no studies. There is also a lack of studies from the Middle East & North Africa and East and Central Asia. Finally, the evidence from several countries with large populations, such as Indonesia, Nigeria and Bangladesh is limited or non-existent.

Most studies report average effects on all children, without providing sub-group analysis based on samples stratified by gender and grade for instance. We conducted meta-analyses for different sub-groups when this was feasible, but these analyses are based on an unrepresentative sample of studies and it is therefore not clear that these findings can be generalised beyond the specific study context. Studies with sub-group analyses were more likely to be those that find no or negative effects on average.

Finally, we conducted extensive targeted searches to identify qualitative studies, process evaluations and project documents associated with included experimental and quasi-experimental studies to help us address review questions 2a and 2b. While we identified a number of relevant documents, the volume and quality of evidence limited the extent to which we are able to provide any generalisable findings at the intervention or review level.

10.3 Quality of the evidence

The review included studies that used randomisation or other rigorous quasi-experimental study designs to answer our main review question. Over half of the included studies randomly allocated treatments to either individuals or clusters (typically schools or villages) and the remaining studies used other techniques such as allocation according to a known rule, propensity score matching or fixed effects regressions to attempts to reduce selection bias and confounding. The results of the risk of bias assessment also suggest that the risk of bias in the included studies was relatively low for most of the six domains of bias included in our appraisal. The highest risk of bias was found for performance bias, where 41 per cent of studies were assessed as having high risk of bias. However, in a relatively large share of studies there is considerable uncertainty about the risk of bias as study reporting often made it difficult to make a judgement. For example, in 41 per cent of the studies it was unclear whether selection bias and confounding had been adequately addressed and over a third of the included studies (38%) did not report sufficient information to make a judgement about performance bias. In 55 percent of the studies, it was unclear if the authors had adequately addressed spill-overs, cross-overs and contamination.

Another issue which is related to study quality, but does not necessarily effect confidence in estimates, is the poor quality of reporting in many studies. Details on context, intervention design and implementation is often missing, so it can be difficult to assess what was delivered, by whom and at what cost. Clear reporting of details on sample characteristics, as well as exact sample size, standard errors, standard deviations and ICCs was also lacking in many studies. This resulted in a large burden on reviewers in trying to get this data, and failing that, the exclusion of studies from the review. The lack of clear reporting limited the usefulness of many studies for the review, and in particular for addressing questions 2a and 2b.

10.4 Agreement and disagreement with other studies

Overall, the findings of our review are fairly well aligned with the evidence provided by previous reviews. The largest overall effects are for structured pedagogy interventions, something echoed by other reviews. We note that our findings are similar to many of those reported in Conn's (2014) systematic review. However, the effects found by Conn are typically of a larger magnitude than those reported by McEwan (2013), Petrosino et al., (2012) and what we find. One possible explanation for this may lie in differential inclusion criteria. Some of our findings contrast somewhat with those of past reviews, for example those for school materials and infrastructure or for contract teachers. This divergence may be partially explained by the small number of studies included for some of these intervention types, with one or two additional studies changing the balance of evidence. Variability in the scope, methodological inclusion criteria or the way interventions are grouped and outcomes are reported in different reviews may also explain different findings.

10.5 Discussion and conclusions

We have conducted the most comprehensive systematic review to date on the effects of education interventions in L&MICs. We reviewed 238 studies evaluating the effects of over twenty different programmes on education outcomes. Our results suggest most interventions have an overall positive effect on beneficiaries as compared to children not receiving these interventions. As expected, depending on which outcome we look at different interventions produce the largest effects. Below we discuss the most promising interventions organised by school participation and learning outcomes.

What works to improve school participation?

When looking across school participation outcomes we find substantial and consistent beneficial effects for cash transfer programmes. Effects range from 0.11, 95% CI [0.07, 0.15] for enrolment, to 0.13, 95% CI [0.08, 0.18] for attendance, with effects on dropout and completion of a similar magnitude. All estimates are based on a relatively large number of studies, with no less than 16 comparisons included in any single meta-analysis. While the results are relatively robust, there is still considerable heterogeneity and we observe effects that are both substantially larger and substantially smaller than the average effects. Cash transfers do not appear to lead to any improvement in learning outcomes however.

We also identify a number of other intervention areas where we observe effects which are relatively larger in magnitude than most other interventions included in the review, but where estimates are less precise and based on fewer studies. Looking at the results of the meta-analyses on participation outcomes we find the effect of community based monitoring on school enrolment is larger in magnitude than what we find for other interventions (SMD=0.17, 95% CI [0.08, 0.25]). But seven of the twelve estimates are from independent sub-samples from the same programme in Uganda and the magnitude of the effect is sensitive to the removal of the study with the largest average effect. Estimates are not consistent across participation outcomes, with smaller and less precise effects on attendance, drop-out and completion. The results of the meta-analyses suggest community based monitoring may also improve learning outcomes in some cases.

The small pool of studies on new schools and infrastructure also suggest there can be beneficial effects on school participation from these programmes. The results indicate construction of new schools may improve enrolment and attendance (enrolment: SMD=0.38, 95% CI [-0.29, 1.04]; attendance: SMD=0.08, 95% CI [-0.04, 0.19]), with large improvements observed in particular in the context of low school participation in Afghanistan. Evidence from India and Kenya suggest construction of latrines may also lead to substantive improvements in school participation outcomes (enrolment, 0.11 SMD, 95% CI [0.01, 0.20]; student attendance 0.14 SMD, 95% CI [0.05, 0.24]). It appears construction of new schools can also improve learning outcomes, but this finding is very context specific (maths: SMD=0.19, 95% CI [-0.15, 0.53]; language arts: SMD=0.02, 95% CI [-0.01, 0.05]).

Finally, school feeding programmes may also improve school participation in some contexts. The effects on enrolment and attendance is of a larger magnitude than most other intervention areas, although the enrolment estimate is relatively imprecise (enrolment:

SMD=0.14, 95% CI [-0.05, 0.33]; attendance: SMD=0.09, 95% CI [0.03, 0.16]), and effects on drop-out and completion are smaller. School feeding also appears promising for improving learning outcomes. Effects range from 0.09, 95% CI [0.01, 17] for language arts to 0.11, 95% CI [0.00, 0.22] for cognitive test scores. While the confidence intervals are wide for all estimates, there is a large amount of heterogeneity and inspecting the forest plots suggest benefits of a relatively large magnitude is some cases. We were not able to explore the reasons for this through moderator analysis. Nevertheless, the findings from the qualitative synthesis suggest baseline levels of food insecurity and school participation may influence effects, with smaller effects in contexts with high rates of school participation and low rates of childhood malnutrition.

What works to improve learning outcomes?

For learning outcomes we find the largest and most consistent positive average effects for structured pedagogy interventions. Typically these programmes include development of new content focused on a particular topic, materials for students and teachers, and short term training courses for teachers in delivering the new content. The meta-analysis for language arts outcomes includes effects from eighteen studies, many of them large scale RCTs, and the meta-analysis shows an overall effect of 0.23 (95% CI [0.13, 0.34]). The effect on maths test scores is slightly smaller in magnitude (SMD=0.14, 95% CI [0.08, 0.20]), but it is still the largest and most consistent effect observed for maths test scores across the review. Most of the programmes evaluated focus on literacy in particular, so the smaller effect on maths test scores is not surprising. Few of the studies in this category evaluate effects on school participation, cognitive outcomes and composite test scores.

We identify a number of other intervention types that appear promising, but where estimates are less precise and based on few studies. We observe substantial benefits on average of merit-based scholarships for both maths and composite test scores (SMD=0.11, 95% CI [0.03, 0.20]; and SMD=0.10, 95% CI [0.03, 0.17]), with smaller effects on language arts (SMD=0.04, 95% CI [-0.07, 0.15]). Again, few of the studies assess other outcomes, such as enrolment, attendance.

Meta-analyses of effects on language arts and maths suggest interventions increasing the time children spend in school by extending the school day may improve learning outcomes. The effects on language arts is larger in magnitude than that observed for most other interventions (SMD= 0.21, 95% CI [0.12, 0.30]), with effects on maths being smaller in magnitude and less precise (SMD=0.09, CI [-0.04, 0.22]). These results are based on only two studies so there is a need for more studies to identify whether these effects can be replicated.

Similarly, the analyses suggest remedial education programmes targeting students who are lagging behind their peers may be beneficial in some contexts. We observe relatively large average effects on test scores. We find the largest effect for composite test scores (SMD=0.22, 95% CI [-0.09, 0.53]), with slightly smaller overall effects for language (SMD=0.16, 95% CI [-0.08, 0.41]) and maths (SMD=0.19, 95% CI [-0.05, 0.44]). However, the results are based on few studies and the confidence intervals cross the line of no effect for all estimates. In each case, the overall effects were sensitive to the removal of a study evaluating the STRIPES trial in India (Lakshminarayana et al., 2013). The small number of identified studies and the large amount of heterogeneity suggest more research is needed to identify whether these effects can be replicated in different contexts.

Finally, the results suggest interventions classified as 'multilevel' have improved learning outcomes substantially in some contexts (maths: SMD=0.16, 95% CI[-0.17, 0.26]; language:

SMD=0.04, 95% CI [-0.17, 0.26]). While the confidence intervals overlap the line of no effect and the average effect for maths is relatively small, the overall estimates include several outliers where the authors report substantial benefits for participants. In particular we observe large effects of the BRIGHT programme in Burkina Faso and the Southwest Basic Education Project (SBEP) in China. The findings are similar for school participation outcomes, although the average effect for completion is more precise (SMD=0.13, 95% CI [0.04, 0.21]), and it is also the largest in magnitude for that outcome observed across the review.

Caveats and potential reasons for heterogeneity

For several of the other intervention areas the effects are relatively small in magnitude, and we also find zero or small negative effects. For instance, the effects of school based management range from -0.01, for language test scores and composite test scores (95% CI [-0.07, 0.05]; 95% CI [-0.10, 0.08]) to 0.05, 95% CI [0.00, 0.09] for completion. Similarly, for de-worming the average effects on education outcomes range from -0.04, 95% CI [-0.11, 0.02] to 0.05 SMD, 95% CI [-0.02, 0.13] for math test scores. We also observe a negative effect on nutrition (SMD= -0.26, 95% CI [-0.43, -0.10]). For interventions providing materials we find limited, if any, difference between children receiving interventions and those that do not. The average effects observed for computer assisted learning also lead us to conclude it is not clear the effect of computer assisted learning on children's learning is beneficial.

One reason for some of the relatively small average effects is the large amount of heterogeneity observed for most of our analyses. Effects that appear small in magnitude often include examples where programmes have had large and substantively important (positive and negative) effects on children's access to education and learning in some contexts. For instance, in the case of school based management programmes we observe effects that are consistently positive and larger in magnitude for the two programmes implemented in the Philippines, as compared to the other programmes (the Third Elementary Education Project and BESRA program). It is not clear why benefits were more substantial in the Phillippines, but the qualitative synthesis provides some potential reasons for this. Many of the programmes with zero or negative effects were primarily providing relatively small grants, with limited decentralisation of decisionmaking powers beyond how to spend the grant. They were also typically implemented in contexts with weak existing education systems and low levels of human capital. On the other hand, programmes in the Phillipines involved more extensive decision making with receipt of the grant being conditional on the quality of school improvement plans (SIPs).

Another reason for some of the relatively small average effects is that the programmes assessed in the included studies have simply not been effective in improving education outcomes. For instance, the overall effects on children receiving CAL interventions range from -0.01 for language test scores (95% CI [-0.08, 0.05]) to 0.07 for maths test scores (95% CI [0.02, 0.11]).The small average effects for many outcomes include several programmes where we observe relatively large negative effects across outcomes, with close to zero effect for most programmes. There are a number of factors that may help explain these results. Evidence from the qualitative synthesis suggest programmes in Chile, Colombia, Peru, Uruguay, Mexico and Nepal all faced process and implementation issues. Firstly, several programmes faced technological issues, including insufficient, damaged and dysfunctional equipment, lack of internet access and software not being compatible with hardware. Secondly, in both Peru and Nepal it was reported that teachers did not receive sufficient training in delivering the CAL programmes. Finally, findings suggest a lack of integration of the CAL technology into existing learning approaches, with the use of laptops and relevant software reported to either be minimal or unrelated to the curriculum.

Finally, it may be that some of the programmes included in the review provide interventions that are necessary but not sufficient to produce substantive improvements across outcomes. For interventions providing materials we also identify a relatively consistent pattern of small or negative effects across learning outcomes. While this finding is based on few studies and should therefore be interpreted with caution, the included studies do provide some possible explanations for this lack of effect. For instance, many schools in both Sierra Leone and India did not receive the text books that were provided as part of the programme, and in Sierra Leone books received by schools were often kept in storage and not distributed to students. There were also suggestions that the increase in supply of materials was less than intended as funds were diverted to use on other items and parents adjusted their contribution in response to the grant programme. Finally, in one case it was suggested the text books, which were in English (the third language of most students) might have been too difficult for most students. Other school level interventions, notably structured pedagogy, do suggest beneficial effects of programmes that include the provision of materials together with other components. So one interpretation may be that in addition the issues with poor implementation, materials such as text books and flipcharts are a necessary but not sufficient for improving children's learning.

10.6 Implications for policy and practice

There is relatively strong evidence that cash transfer programmes and structured pedagogy interventions are particularly effective in improving school participation and learning outcomes in most contexts. Ensuring that all children have access to high quality education and gain the knowledge and skills needed to realise the benefits of education is a complex process. Children are faced with multiple barriers to school participation and learning. It may therefore not be surprising that programmes improving school participation do not necessarily improve learning outcomes and vice versa.

Depending on the barriers facing children in specific contexts, it may be necessary to intervene across more than one sphere to improve the chances of seeing substantive improvements in one or more outcomes. The main findings of the review offer some support for this. With the possible exception of school feeding, programmes that improve school participation do not appear to improve learning outcomes. Improving children's school participation through cash transfer programmes may have a limited effect on learning outcomes if the existing curriulum content, materials and teachers available are not of sufficient quality. Similarly, when we observe substantial improvements of learning outcomes, as in the case of structured pedagogy, we do not typically find similar improvements for school participation outcomes.

The findings for most other intervention areas are based on few studies and we also observe substantial variability of effects. Average effects that appear small in magnitude often include examples where programmes have had large and substantively important (positive and negative) effects on children's access to education and learning in some settings. Therefore the average effects should be interpreted with some caution, considering also the range of effects observed in different studies.

10.7 Implications for research

Through this review, we have identified a range of programme areas where there are few or no studies. This includes teacher training programmes, remedial education, school- based health programmes (malaria, de-worming, micronutrients), diagnostic feedback, providing information to parents, tracking students by ability, extension of the school day and different approaches to teacher training and hiring. Some of these interventions appear particularly promising and it may be worth focusing new studies in these areas.

While the included studies use rigorous designs to assess the effects of interventions, most studies do not address other questions comprehensively, such as those relating to how and why interventions work or not, and at what cost. Future studies should use mixed-methods study designs to assess the effects of interventions as well as process, implementation and contextual factors that influence final outcomes. This will help explain heterogeneity in effects that can in turn help inform improvements of future programmes. Finally, studies should include information about costs to allow cost-effectiveness analysis.

Not all studies provide clear and comprehensive reporting of methods and results. Studies will be more useful if they clearly describe all main study constructs, report methods in detail and clearly report the statistical information necessary to calculate standardised effect sizes, including sample sizes, standard errors, standard deviations and intracluster correlation coefficients (ICCs). Without clear reporting of what was studied and how, resources used on expensive studies are wasted. Research funders and publishers may consider making it a requirement that researchers follow reporting guidelines such as CONSORT to improve the value of new research.

10.8 Limitations and deviations from the protocol

There are several limitations to the review process. Firstly, due to time constraints, the risk of bias and critical appraisal has only been conducted by one researcher. Secondly, we assessed all studies for unit of analysis errors, but due to resource constraints and a lack of the necessary data (number of clusters and participants in treatment and control group and ICCs) we did not correct the standard errors before including studies in the meta-analyses. Another issue relates to the lack of qualitative studies, project documents and process evaluations. We invested significant efforts to identify additional studies associated with the included impact evaluations, but for many programmes we were unable to identify any such studies. Moreover, many of the studies we identified provided very descriptive findings and suffered from methodological weaknesses, limiting their usefulness. Thirdly, for cash transfers we implemented a more limited coding scheme and the data extraction has not been double checked for all the studies in this intervention category. There were also a number of included studies that did not contain the necessary data for us to calculate effect sizes and so were not included in our meta- analysis. We tried to obtain this information by contacting the author team but in most cases we did not receive a response. Due to a lack of sufficient studies we were also not able to conduct meta- regressions to explore reasons for heterogeneity for any single intervention area apart from cash transfers.

We made several changes to the methods outlined in the protocol. We identified a large number of small studies testing specific techniques, such as the use of concept mapping or a specific computer programme in a few groups of students. These studies were not explicitly included nor excluded according to our protocol. They were very different from the rest of the literature, often of low methodological quality and our assessment was that these studies would not add much to the review. We did not think it was a good use of resources to include these studies and therefore developed additional criteria which allowed us to exclude such efficacy studies systematically.

Table 10 a: Summary of findings for primary outcomes

	Enrolment	Attendance	Drop-out	Completion	Cognitive	Maths	Language arts	Composite
Child level								
School feeding	0.14, 95% CI [-0.05, 0.33] 7 studies	0.09. 95% CI [0.03, 0.16] 6 studies	0.06, 95% CI [-0.15, 0.03] 3 studies	0.01, 95% CI [-0.03, 0.01]	0.11, 95% CI [0.00, 0.22] 7 studies	0.10, 95% CI [0.00, 0.19] 10 studies	0.09, 95% CI [0.01, 0.17] 8 studies	0.14, 95% CI [-0.04, 0.33] 3 studies
School-based health- Malaria	No studies	No meta-analysis	No studies	No studies	0.03, 95% CI [-0.05, 0.12] 3 studies	0.16, 95% CI [-0.08, 0.25] 3 studies	0.03, 95% CI [-0.49, 0.55] 3 studies	No studies
School-based health- Micronutrient	No studies	No meta-analysis	No studies	No studies	0.01 , 95% CI [-0.03, 0.05] 2 studies	0.06, 95% CI [0.02, 0.10] 4 studies	No meta-analysis	No studies
School-based health- Deworming	No studies	0.04 (95% CI [-0.13, 0.21] 4 studies	No studies	No studies	0.01, 95% CI [-0.03, 0.05] 3 studies	0.05 (95% CI [-0.02, 0.13] 2 studies	-0.04, 95% CI [-0.11, 0.02] 3 studies	No studies
Merit-based scholarships	No studies	0.01 SMD, 95% CI [-0.06, 0.08] 4 studies	0.04, 95% CI [-0.11, 0.19] 2 studies	0.32, 95% CI [-0.18, 0.46] 2 studies	No meta-analysis	0.11, 95% CI [0.03, 0.20] 10 studies	0.04, 95% CI [-0.07, 0.15] 3 studies	0.10, 95% CI [0.03, 0.17] 7 studies
Providing information	No meta-analysis	No meta-analysis	No meta-analysis	No meta-analysis	No studies	No meta-analysis	No studies	No meta-analysis
Household level								
User fee elimination	0.03, 95% CI [-0.01, 0.06] 8 studies	0.01, 95% CI [-0.13, 0.15] 2 studies	-0.10, 95% CI [-0.23, 0.02] 4 studies	0.02, 95% CI [-0.10, 0.15] 3 studies	No studies	No studies	No studies	No studies
Cash transfers	0.11 [0.07, 0.15] 49 studies	0.13 [0.08, 0.18] 38 studies	-0.12, 95% CI [-0.16, -0.07] 16 studies	0.12, 95% CI [0.01, 0.22] 28 studies	0.07, 95% CI [-0.11, 0.25] 2 studies	0.01, 95% CI [-0.03, 0.04] 14 studies	0.00, 95% CI [-0.04, 0.04] 14 studies	0.01, 95 % CI [-0.01, 0.03] 3 studies
School level								
Computer-assisted learning	-0.04, 95% CI [-0.11, 0.04] 2 studies	0.04, 95% CI [0.00, 0.07] 2 studies	-0.04, 95% CI [-0.12, 0.04] 2 studies	0.07, 95% CI [-0.07, 0.22] 2 studies	No studies	0.07, 95% CI [0.02, 0.11] 19 studies	-0.01, 95% CI [-0.08, 0.05] 13 studies	0.01, 95% CI [-0.04, 0.07] 6 studies

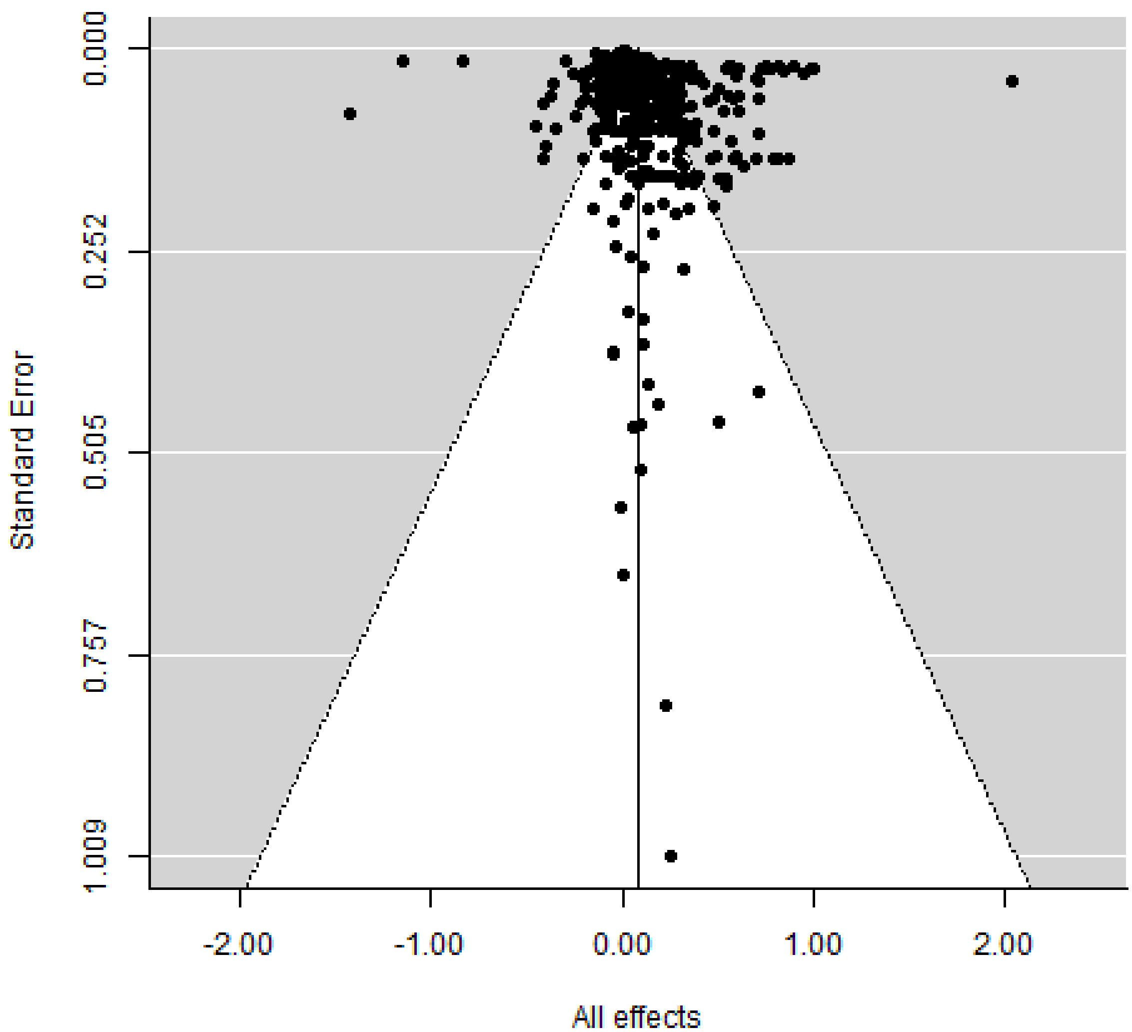
Providing materials	No meta-analysis	No meta-analysis	No meta-analysis	No meta-analysis	No studies	-0.02, 95% CI [-0.06, 0.02] 5 studies	0.00, 95% CI [-0.02, 0.02] 5 studies	0.01, 95% CI [-0.01, 0.02] 5 studies
Remedial Education	No studies	No studies	No studies	No studies	No studies	0.19, 95% CI [-0.05, 0.44] 6 studies	0.16, 95% CI [-0.08, 0.41] 6 studies	0.22, 95%CI [-0.09, 0.53] 5 studies
New Schools and Infrastructure: hygiene infrastructure interventions	0.11 , 95% CI [0.01, 0.20] 4 studies	0.14, 95% CI [0.05, 0.24] 2 studies	No studies	No meta-analysis	No studies	No studies	No studies	No studies
New Schools and Infrastructure: construction of new schools	0.38, 95% CI [-0.29, 1.04] 2 studies	0.08, 95% CI [-0.04, 0.19] 2 studies	No studies	No studies	No studies	0.19, 95% CI [-0.15, 0.53] 2 studies	0.02, 95% CI [-0.01, 0.05] 2 studies	No studies
New Schools and Infrastructure: improvement or construction of new school infrastructure	No meta-analysis	No meta-analysis	No studies	No meta-analysis	No studies	No studies	No meta-analysis	No studies
Pedagogy	No studies	0.02 [0.00, 0.04] 5 studies	No meta-analysis	0.13 [-0.02, 0.28] 2 studies	0.01 [-0.04, 0.07] 2 studies	0.11 [0.03, 0.18] 18 studies	0.24 [0.12, 0.36] 21 studies	0.06 [0.03, 0.08] 3 studies
Extra time in school	No studies	No studies	No studies	No studies	No studies	0.09 [-0.04, 0.22] 2 studies	0.19, 95% CI [0.15, 0.24] 2 studies	No studies
School participating by ability: grade retention	No studies	No studies	No studies	No studies	No studies	No meta-analysis	No meta-analysis	No studies
School participating by ability: tracking	No studies	No studies	No studies	No studies	No studies	0.02, 95% CI [-0.04, 0.08] 2 studies	0.12, 95% CI [-0.03, 0.27] 2 studies	No studies
Teacher level								
Teacher incentives	0.06, 95% CI [-0.05, 0.16] 2 studies	0.01, 95% CI [-0.04, 0.06] 3 studies	0.00, 95% CI [-0.01, 0.01] 4 studies	0.03, 95% CI [0.00, 0.05] 4 studies	No studies	0.08, 95% CI [0.02, 0.13] 11 studies	0.00, 95% CI [-0.13, 0.12] 7 studies	0.02, 95% CI [-0.02, 0.05] 4 studies
Teacher hiring	No studies	No meta-analysis	No meta-analysis	0.04, 95% CI [0.01, 0.08] 3 studies	No studies	0.10, 95% CI [0.00, 0.20] 2 studies	0.06, 95% CI [0.03, 0.10] 2 studies	0.06, 95% CI [-0.01, 0.12] 3 studies

Teacher training	No studies	No studies	No studies	No studies	No studies	No meta-analysis	No meta-analysis	No studies
Diagnostic feedback	No studies	No studies	No studies	No studies	No meta-analysis	0.01, 95% CI [-0.01, 0.03] 3 studies	0.01, 95% CI [-0.01, 0.05] 3 studies	No meta-analysis
System level								
School-Based Management	0.01, 95% CI [-0.04, 0.07] 3 studies	No meta-analysis	-0.02, 95% CI [-0.05, 0.01] 7 studies	0.05, 95% CI [0.00, 0.09] 8 studies	No studies	0.01, 95% CI [-0.02, 0.05] 21 studies	-0.01, 95%CI [-0.07, 0.05] 20 studies	-0.01, 95% CI [-0.10, 0.08] 9 studies
Community Based Monitoring	0.17, 95% CI [0.08, 0.25] 12 studies	0.04, 95% CI [-0.09, 0.18] 3 studies	0.05, 95% CI [-0.09, 0.20] 3 studies	0.06, 95% CI [0.01, 0.12] 3 studies	No studies	0.09, 95% CI, [-0.02, 0.2] 9 studies	0.12, 95% CI [0.01, 0.22] 9 studies	0.10, 95% CI [-0.01, 0.21] 7 studies
Private-public partnerships	0.19, 95% CI [0.01, 0.36] 7 studies	No studies	No meta-analysis	0.23, 95%CI [-0.07, 0.53] 3 studies	No studies	0.05, 95% CI [0.01, 0.08] 7 studies	0.04, 95% CI [0.00, 0.09] 7 studies	0.07, 95% CI [-0.07, 0.20] 4 studies
Multi-component Interventions								
Multi-component Interventions	0.01, 95% CI [-0.06, 0.08], 3 studies	0.16, 95% CI [-0.12, 0.44] 3 studies	0.16, 95% CI [-0.33, 0.02] 3 studies	0.13, 95% CI [0.04, 0.21] 4 studies	No studies	0.16, 95% CI [-0.17, 0.48] 10 studies	0.04, 95% CI [-0.17, 0.26] 14 studies	0.02, 95% CI [-0.08, 0.12] 3 studies

11. Additional analyses

11.1 Publication bias

This section examines the evidence for selective reporting of positive significant findings, or publication bias. Figure 9.2a presents the funnel graph which plots effect size against standard error (inverted scale). Visual inspection of the plot suggests there may be some asymmetry, as there are fewer studies in the region of negative significance than in the region of positive significance. These potentially missing effects are clustered in the region of low statistical power, on the bottom left hand side, which is precisely where theory would suggest they would be (i.e. it is easier for negative study findings to 'disappear' when study sample size is small). Indeed, Egger's test provides statistical support for asymmetry (p<0.000), therefore suggesting studies with smaller sample sizes (and hence larger standard errors) which find negative effects of education programmes may be under-reported.



11.2 Moderator analysis

11.2.1 Meta-regression analysis at the review level: all outcomes

To explore possible sources of heterogeneity we conducted moderator analysis at the review level across four primary outcomes (enrolment, attendance, maths test scores and language test scores). We used the following methodological and substantive moderator variables: study design, risk of bias assessment, World Bank region, country income classification, the primary type of implementing agency (government, NGO, research team or a combination) and length of follow- up of outcome data collection after the start of the programme. For this type of analysis, the moderators are at the level of the study and the dependent variable is the study effect sizes rather than a specific outcome (Borenstein et al., 2009). The results are shown in table 11.2a.

The analysis suggests that studies of programmes taking place in East Asia and Pacific and Latin America and the Caribbean are associated with larger effects across the four education outcomes. This is also the case to a slightly smaller extent for programmes in South Asia and Sub-Saharan Africa. Studies of programs and trials in upper-middle income countries are also associated with smaller effects on education outcomes. Studies implemented by researchers as opposed to NGOs or governments, are associated larger effects. We do not find significant associations between the effect size magnitude and the length of follow up, study design or risk of bias, with the exception of studies assessed as being of low risk of selective analysis reporting, where there is a small negative association with effect size.

Table 11.2 a: Meta-regression analysis at the review level (all outcomes)

	Coefficient estimate (robust standard errors reported in brackets)
Intercept	-0.101
	(0.114)
Follow up period	0.0002
	(0.001)
East Asia and Pacific	0.187***
	(0.071)
Europe and Central Asia	0.059
	(0.096)
Latin America and the Caribbean	0.179*
	(0.093)
South Asia	0.129*
	(0.070)
Sub-Saharan Africa	0.114*
	(0.065)
Low income country	-0.002
	(0.047)
Upper-middle income country	-0.111**
	(0.056)
Government	0.097
	(0.078)
NGO	0.134
	(0.083)
Researchers	0.152*
	(0.080)
Various	0.048
	(0.083)
RCT	0.048
	(0.109)
Cluster RCT	-0.050
	(0.083)
Controlled Before-and-After design	-0.008
	(0.075)
RDD	-0.010
	(0.083)
Other	0.038

	Coefficient estimate (robust standard errors reported in brackets)
	(0.108)
Selection Bias - Low risk	-0.011
	(0.028)
Selection Bias - High risk	0.018
	(0.039)
Spill-overs - Low risk	0.038
	(0.028)
Spill-overs - high risk	0.034
	(0.044)
Outcome bias - Low risk	0.036
	(0.050)
Outcome bias - High risk	0.024
	(0.063)
Analysis bias - Low risk	-0.067**
	(0.029)
Analysis bias - High risk	-0.001
	(0.097)
Performance bias - Low risk	-0.026
	(0.028)
Performance bias - High risk	-0.002
	(0.028)
Enrolment	0.016
	(0.028)
Language	0.007
	(0.025)
Maths	0.006
	(0.027)

Note: *p<0.1, **p<0.05, ***p<0.01

11.2.2 Meta-regression analysis at the review level: by different outcomes

We also explored possible sources of heterogeneity using multivariate meta-regression at the outcome level for enrolment, student attendance, maths test scores and language arts test scores. For this analysis, we assessed the potential associations between effect sizes and a reduced range of moderator variables, namely World Bank region, country income classification, the primary type of implementing agency (government, NGO, research team or a combination) and the length of follow- up of outcome data collection after the start of the programme. Due to the limited number of observations all results should be interpreted with caution. The results are shown in table 11.2b.

Table 11.2 b: Meta-regression analysis at the review level (by outcome)

	Coefficient estimate (robust standard errors reported in brackets)			
	Enrolment	Attendance	Maths	Language
Intercept	-0.148**	0.058	0.048	-0.100
	(0.058)	(0.107)	(0.059)	(0.091)
Follow up period	0.0005	0.001	-0.0002	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
East Asia and Pacific	0.178*	0.094*	0.025	
	(0.104)	(0.052)	(0.049)	
Europe and Central Asia	0.149			
	(0.139)			
Latin America and the Caribbean	0.234	0.093**	0.083	0.018
	(0.141)	(0.044)	(0.089)	(0.092)
South Asia	0.120***	0.057	0.011	-0.059
	(0.042)	(0.036)	(0.051)	(0.079)
Sub-Saharan Africa	0.030	-0.003	0.017	0.047
	(0.123)	(0.039)	(0.070)	(0.125)
Low income	0.025	0.138***	-0.042	-0.102
	(0.115)	(0.046)	(0.068)	(0.080)
Upper-middle-income	-0.235*	-0.078	-0.055	-0.097
	(0.139)	(0.050)	(0.087)	(0.106)
Government	0.178***	-0.083	0.042	0.180**
	(0.046)	(0.107)	(0.072)	(0.073)
NGO	0.219**	-0.124	0.058	0.271***
	(0.097)	(0.112)	(0.066)	(0.081)
Researchers	0.282	-0.055	0.071	0.302***
	(0.184)	(0.120)	(0.052)	(0.086)
Various	0.059	-0.218*	0.007	0.183**
	(0.050)	(0.111)	(0.056)	(0.078)
Observations	99	76	199	210

Note: *p<0.1, **p<0.05, ***p<0.01

There are few consistent significant moderator variables across the four outcomes and no significant moderator variables for the meta-regression of maths outcomes. The analysis suggests that studies of programmes taking place in East Asia and Pacific are associated with larger effects on enrolment and attendance, but not on maths test scores. Studies of programmes and trials taking place in South Asia are also associated with larger effects on enrolment, while studies taking place in Latin America and the Caribbean are associated with larger effects on attendance. Mirroring the all outcomes meta-regression, for enrolment there is a significant negative association between the effect size and status as an upper-middle income country. For student attendance there is also a significant positive association between effect size and status as a low income country.

In terms of the primary type of implementing agency, we see the largest positive significant association between language test scores and programmes implemented by researchers. We also see a significant positive association between government and NGO run programmes and effects on enrolment and language test scores. We do not observe any significant associations between the effect size and the length of follow up of outcome data collection for any of the four outcomes.

References

Included impact evaluations

Abeberese, A. B., Kumler, T. J. and Linden, L., 2011. Improving reading skills by encouraging children to read: A randomized evaluation of the Sa Aklat Siskat reading program in the Philippines.

Abeberese, A., Kumler, T. and Linden, L., 2013. Improving reading skills by encouraging children to read in school: A randomized evaluation of the Sa Aklat Siskat reading program in the Philippines. Cambridge, MA: NBER Working Paper.

Abeberese, A. B., Kumler, T. J. and Linden, L. L., 2014. Improving reading skills by encouraging children to read in school: A randomized evaluation of the Sa Aklat Siskat reading program in the Philippines. *Journal of Human Resources,* 49**,** 611-33.

Adelman Melissa, A. and Holland Peter, A., 2015. Increasing access by waiving tuition: Evidence from Haiti. Washington D.C.: World Bank

Adrogue, C. and Orlicki, M. E., 2013. Do in-school feeding programs have an impact on academic performance? The case of public schools in Argentina. *Education Policy Analysis Archives,* 21**,** 2-20.

Adrogué, C. and Orlicki, M. E., 2011. Do in-school feeding programs have impact on academic performance and dropouts? The case of public argentine schools. Buenos Aires: Universidad de San Andrés.

Adukia, A., 2014. Sanitation and education. Unpublished working paper.

Afridi, F., Barooah, B. and Somanathan, R., 2014. The mixture as before? Student responses to the changing content of school meals in India. Unpublished working paper.

Ahmed, A., Adato, M., Kudat, A., Gilligan, D. and Colasan, R., 2007. Impact evaluation of the conditional cash transfer program in turkey: Final report.

Akresh, R., De Walque, D. and Kazianga, H., 2013. Cash transfers and child schooling: Evidence from a randomized evaluation of the role of conditionality. *The World Bank, Policy Research Working Paper Series: 6340*.

Alam, A. and Baez, J., 2011. Does cash for school influence young women's behavior in the longer term? Evidence from Pakistan. *Working Paper for the World Bank, Policy Research, Series: 5669.* Working paper.

Alviar, C., Davis, B., Handa, S., Hurrell, A., Hussein, A., Musembi, D., Ochieng, S., Pearson, R., Pellerano, L., Visram, A., Ward, P. and Kenya, C. T. O. V. C. E. T., 2012. The impact of Kenya's cash transfer for orphans and vulnerable children on human capital. *Journal of Development Effectiveness,* 4**,** 38-49.

Amarante, V., Ferrando, M. and Vigorito, A., 2012. School attendance, child labor and cash transfers: An impact evaluation of panes. Working Paper.

Andrabi, T., Das, J. and Khwaja, A. I., 2013. Report cards: The impact of providing school and child test scores on educational markets. Unpublished working paper.

Angrist, J., Bettinger, E., Bloom, E., Kremer, M. and King, E., 2002. Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment. *American Economic Review,* 92**,** 1535-58.

Angrist, J., Bettinger, E. and Kremer, M., 2004. Long-term consequences of secondary school vouchers: Evidence from administrative records in Colombia. Cambridge MA: NBER Working Paper No. 10713.

Angrist, J., Bettinger, E. and Kremer, M., 2006. Long-term educational consequences of secondary school vouchers: Evidence from administrative records in Colombia. *American economic review,* 96**,** 847-862.

Angrist, J. D., Bettinger, E., Bloom, E., King, E. and Kremer, M., 2001. Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment. Cambridge, MA: NBER.

Attanasio, O., Fitzsimons, E., Gomez, A., Lopez, D., Meghir, C. and Mesnard, A., 2006. Child education and work choices in the presence of a conditional cash transfer programme in rural Colombia. Working paper.

Attanasio, O. and Gómez, L., 2004. Evaluación del impacto del programa familias en acción - subsidios condicionados de la red de apoyo social. Report.

Aturupane, H., 2011. An impact evaluation of Sri Lanka's policies to promote the academic performance of primary school students through school improvement and report card programs. South Asia Human Development Sector (Report No.35).

Aturupane, H., Glewwe, P., Ravina, R., Sonnadara, U. and Wisniewski, S., 2014. An assessment of the impacts of Sri Lanka's programme for school improvement and school report card programme on students' academic progress. *Journal of Development Studies,* 50**,** 1647-1669.

Baez, J. E. and Camacho, A., 2011. Assessing the long-term effects of conditional cash transfers on human capital: Evidence from Colombia. Washington D.C.

Baird, S., Chirwa, E., Mcintosh, C. and Özler, B., 2010. The short-term impacts of a schooling conditional cash transfer program on the sexual behavior of young women. *Health Economics,* 19**,** 55–68.

Baird, S., Mcintosh, C. and Ozler, B., 2009. Designing cost-effective cash transfer programs to boost schooling among young women in sub-Saharan Africa. Working Paper.

Baird, S., Mcintosh, C. and Ozler, B., 2010. Cash or condition? Evidence from a randomized cash transfer program. Working Paper.

Baird, S., Mcintosh, C. and Özler, B., 2011. Cash or condition? Evidence from a cash transfer experiment. *Quarterly Journal of Economics,* 126**,** 1709-1753.

Bando, R., 2010. *The effect of school based management on parent behavior and the quality of education in Mexico.*

Banerjee, A., Cole, S., Duflo, E. and Linden, L. L., 2005. Remedying education: Evidence from two randomized experiments in India. Cambridge, MA: NBER.

Banerjee, A. V., Banerji, R., Duflo, E., Glennerster, R. and Khemani, S., 2008. Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India. Washington D.C.: The World Bank.

Banerjee, A. V., Banerji, R., Duflo, E., Glennerster, R. and Khemani, S., 2010. Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India. *American Economic Journal: Economic Policy,* 2**,** 1-30.

Banerjee, A. V., Cole, S., Duflo, E. and Linden, L. L., 2007. Remedying education: Evidence from two randomized experiments in India. *The Quarterly Journal of Economics,* 122**,** 1235-1264.

Barrera Osorio, F., Linden, L. L. and Hernández, M., 2008. Programa computadores para educar evaluacion de impacto. Unpublished working paper.

Barrera-Osorio, F., 2006. The impact of private provision of public education: Empirical evidence from Bogotá's concession schools. Washington D.C.: World Bank.

Barrera-Osorio, F., Bertrand, M., Linden, L. and Perez-Calle, F., 2011. Improving the design of conditional transfer programs: Evidence from a randomized education experiment in Colombia. *American Economic Journal-Applied Economics,* 3**,** 167-195.

Barrera-Osorio, F., Blakeslee, D. S., Raju, D. and Linden, L. L., 2011. Expanding educational opportunities in remote parts of the world: Evidence from a RCT of a public-private partnership in Pakistan. Working paper.

Barrera-Osorio, F. and Dhushyanth, R., 2015. Teacher performance pay: Experimental evidence from Pakistan. *Policy Research Working Paper 7307*.

Barrera-Osorio, F., Galbert De, P., Habyarimana, J. and Sabarwal, S., 2015. The impact of public-private partnerships on private school performance: Evidence from a randomized controlled trial in Uganda. Working paper.

Barrera-Osorio, F., Linden Leigh, L. and Urquiola, M., 2007. The effects of user fee reductions on enrollment: Evidence from a quasi-experiment. Unpublished manuscript.

Barrera-Osorio, F. and Linden, L. L., 2009. The use and misuse of computers in education: Evidence from a randomized experiment in Colombia. Washington D.C.: Education Team, The World Bank Human Development Network.

Barrera-Osorio, F. and Raju, D., 2010. Short-run learning dynamics under a test-based accountability system: Evidence from Pakistan. Washington D.C.: World Bank.

Barrera-Osorio, F. and Raju, D., 2011. Evaluating public per-student subsidies to low-cost private schools regression-discontinuity evidence from Pakistan. Washington D.C.: World Bank.

Barrera-Osorio, F. and Raju, D., 2014. Evaluating the impacts of public student subsidies to low-cost private schools in Pakistan. Working paper.

Bau, N. and Das, J., 2011. Buy more, pay less: Measuring the performance of Pakistani contract teachers. Unpublished working paper.

Bau, N. and Das, J., 2014. The misallocation of pay and productivity in the public sector: Evidence from the labor market for teachers.

Baulch, B., 2011. The medium-term impact of the primary education stipend in rural Bangladesh. *Journal of Development Effectiveness,* 3**,** 243-262.

Beasley, E. and Huillery, E., 2012. Empowering parents in schools: What they can (not) do. Unpublished working paper.

Beasley, E. and Huillery, E., 2014. Willing but unable: Short-term experimental evidence on parent empowerment and school quality.

Behrman Jere, R., Sengupta, P. and Todd, P., 2000. Final report: The impact of progresa on achievement test scores in the first year. Washington D.C.

Behrman, J. R., Parker, S., Todd, P. and Wolpin, K., 2012. Aligning learning incentives of students and teachers: Results from a social experiment in Mexican high schools. Unpublished working paper.

Bellei, C., 2009. Does lengthening the school day increase students' academic achievement? Results from a natural experiment in Chile. *Economics of Education Review,* 28**,** 629-640.

Bellei, C., 2013. Supporting instructional improvement in low-performing schools to increase students' academic achievement. *The Journal of Educational Research,* 106 (3).

Benedetti, F., Ibarrarán, P. and Mcewan, P. J., 2015. Do education and health conditions matter in a large cash transfer? Evidence from a Honduran experiment. Working paper.

Benhassine, N., Devoto, F., Duflo, E., Dupas, P. and Pouliquen, V., 2014. Turning a shove into a nudge? A "labeled cash transfer" for education. Working paper.

Berlinski, S. and Busso, M., 2013. Pedagogical change in mathematics teaching: Evidence from a randomized control trial. Unpublished working paper.

Berry, J., 2013. Child control in education decisions: An evaluation of targeted incentives to learn in India. Unpublished working paper.

Bettinger, E., Kremer, M. and Saavedra, J. E. Are educational vouchers only redistributive? Prepared for the CESifo/PEPG joint conference “Economic Incentives: Do They Work in Education” Insights and Findings from Behavioral Research, CESifo Conference Center, May 16-17, 2008, 2008 2008 Munich Germany.

Bjorkman, M., 2006. Does money matter for student performance? Evidence from a grant program in Uganda. Stockholm Sweden: Institute for International Economic Studies (IIES), Stockholm University.

Blimpo, M., 2014. Team incentives for education in developing countries: A randomized field experiment in Benin. *American Economic Journal: Applied Economics,* 6**,** 90-109.

Blimpo, M., Evans, D. and Lahire, N., 2015. Parental human capital and effective school management: Evidence from the Gambia. Washington D.C.: World Bank.

Blimpo, M. P., 2010. Team incentives for education in developing countries a randomized field experiment in Benin. Unpublished working paper.

Blimpo, M. P. and Evans, D. K., 2011. School-based management and educational outcomes: Lessons from a randomized field experiment. Washington D.C.: World Bank.

Bold, T., Kimenyi, M., Mwabu, G., Ng'ang'a, A. and Sandefur, J., 2013. Scaling up what works: Experimental evidence on external validity in Kenyan education. Washington, DC: The Center for Global Development.

Borkum, E., He, F. and Linden, L. L., 2012. School libraries and language skills in Indian primary schools: A randomized evaluation of the akshara library program. Unpublished working paper.

Borkum, E., He, F. and Linden, L. L., 2013. The effects of school libraries on language skills: Evidence from a randomized controlled trial in India. Unpublished working paper.

Brooker, S. and Halliday, K., 2015. Impact of malaria control and enhanced literacy instruction on educational outcomes among school children in Kenya a multi-sectoral, prospective randomised evaluation. *3ie Impact Evaluation Report 18*.

Brooker, S., Inyega, H., Estambale, B., Njagi, K., Juma, E., Jones, C. and Goodman C Jukes, M., 2013. Impact of malaria control and enhanced literacy instruction on educational outcomes among Kenyan school children: A multi-sectoral, prospective, randomized evaluation. Delhi: International Initiative for Impact Evaluation (3ie).

Burde, D. and Linden, L. L., 2009. The effect of proximity on school enrollment: Evidence from a randomized controlled trial in Afghanistan. Unpublished working paper.

Burde, D. and Linden, L. L., 2011. The effect of village-based schools: Evidence from a randomized controlled trial in Afghanistan. Unpublished working paper.

Burde, D. and Linden, L. L., 2013. Bringing education to afghan girls: A randomized controlled trial of village-based schools. *American Economic Journal-Applied Economics,* 5 (3)**,** 27-40.

Buttenheim, A., Alderman, H. and Friedman, J., 2011. Impact evaluation of school feeding programmes in Lao people's democratic republic. Washington D.C.: World Bank.

Cabezas, V., Canessa, M., Cuesta, J. I. and Gallego, F. A., 2011. Evaluación de impacto: Servicio país en educación. Cambridge, Massachusetts: Abdul Latif Jameel Poverty Lab (J-PAL).

Cabezas, V., Cuesta, J. I. and Gallego, F. A., 2011. Effects of short-term tutoring on cognitive and non-cognitive skills: Evidence from a randomized evaluation in Chile. Unpublished working paper.

Camargo, B., Camelo, R., Firpo, S. and Ponczek, V., 2012. Test score disclosure and school performance. Unpublished working paper.

Carillo, P., Onofa, M. and Ponce, J., 2010. Information technology and student achievement: Evidence from a randomized experiment in Ecuador. Washington, D.C.: Inter-American Development Bank.

Carnoy, M., Gove Amber, K., Loeb, S., Marshall Jeffery, H. and Socias, M., 2008. How schools and students respond to school improvement programs: The case of Brazil's PDE. *Economics of Education Review,* 27**,** 22-38.

Cerdan-Infantes, P. and Vermeersch, C., 2007. More time is better: An evaluation of the full time school program in Uruguay. *Policy Research Working Paper Series No 4167.* Washington DC: The World Bank.

Chaudhury, N., Friedman, J. and Onishi, J., 2013. Philippines conditional cash transfer program impact evaluation 2012. Report.

Chaudhury, N. and Parajuli, D., 2006. Conditional cash transfers and female schooling: The impact of the female school stipend program on public school enrollments in Punjab, Pakistan. Washington D.C.

Chaudhury, N. and Parajuli, D., 2010. Conditional cash transfers and female schooling: The impact of the female school stipend programme on public school enrolments in Punjab, Pakistan. *Applied economics,* 42**,** 3565-3583.

Chay, K., Mcewan, P. J. and Urquiola, M., 2005. The central role of noise in evaluating interventions that use test scores to rank schools. *The American Economic Review,* 95 (4)**,** 1237-58.

Chay, K. Y., Mcewan, P. J. and Urquiola, M., 2003. The central role of noise in evaluating interventions that use test scores to rank schools. National Bureau of Economic Research (NBER).

Chen, Yaojiang, S., Hongmei, Y., Linxiu, Z., Di, M., James, C. and Scott, R., 2013. The impact of a senior high school tuition relief program on poor junior high school students in rural china. *China & World Economy,* 21**,** 80–97.

Chen, X. L. C. Z. L. S. Y. R. S., 2010. Does taking one step back get you two steps forward? Grade retention and school performance in poor areas in rural china. *International Journal of Educational Development,* 30**,** 544-559.

Chen Zhejiang, X., Shi, Y., Yi, H., Zhang, L., Mo, D., Chu, J., Loyalka, P. and Rozelle, S., 2013. The impact of a senior high school tuition relief program on poor junior high school students in rural china. Xi'an China: Northwest University.

Cheung, M. and Berlin, M. P., 2014. The impact of a food for education program on schooling in Cambodia. *Asia & the Pacific Policy Studies,* 2**,** 44–57.

Chin, A., 2005. Can redistributing teachers across schools raise educational attainment? Evidence from operation blackboard in India. *Journal of Development Economics,* 78**,** 384-405.

Chyi, H. and Zhou, B., 2014. The effects of tuition reforms on school enrollment in rural china. *Economics of Education Review,* 38**,** 104-23.

Clarke Sian, E., Jukes Matthew, C. H., Njagi, J. K., Khasakhala, L., Cundill, B., Otido, J., Crudder, C., Estambale Benson, B. A. and Brooker, S., 2008. Effect of intermittent preventive treatment of malaria on health and education in schoolchildren: A cluster-randomised, double-blind, placebo-controlled trial. *Lancet,* 372**,** 127-38.

Contreras, D. and Rau, T., 2012. Tournament incentives for teachers: Evidence from a scaled-up intervention in Chile. *Economic Development and Cultural Change,* 61**,** 219-246.

Correa, J. A., Inostroza, D., Parro, F., Reyes, L. and Ugarte, G., 2012. The effects of vouchers on academic achievement: Evidence from the Chilean preferential scholastic subsidy. Santiago (Chile): Working paper.

Correa, J. A., Inostroza, D., Parro, F., Reyes, L. and Ugarte, G., 2014. The effects of vouchers on academic achievement: Evidence from Chile's conditional voucher program. Santiago: Universidad Andrés Bello.

Correa, J. A., Parro, F. and Reyes, L., 2014. The effects of vouchers on school results: Evidence from Chile's targeted voucher program. *Journal of Human Capital,* 8**,** 352-398.

Cristia, J., Czerwonko, A. and Garofalo, P., 2014. Does technology in schools affect repetition, dropout and enrollment? Evidence from Peru. *Journal of Applied Economics,* Vol XVII, No. 1**,** 89-112.

Cristia, J., Ibarraran, P., Cueto, S., Santiago, A. and Severin, E., 2012. Technology and child development: Evidence from the one laptop per child program. Bonn Germany: The Institute for the Study of Labor (IZA).

Cristia, J. P., Czerwonko, A. and Garofalo, P., 2010. Does ICT increase years of education? Evidence from Peru. Washington D.C.: Inter-American Development Bank.

Cueto, S., Torero, M., León, J. and Deustua, J., 2008. Asistencia docente y rendimiento escolar: El caso del programa meta. Lima: Grupo de Análisis para el Desarrollo (GRADE).

Dammert Ana, C., 2008. Heterogeneous impacts of conditional cash transfers: Evidence from Nicaragua. Bonn Germany.

Dang, H. A., Sarr, L. and Asadullah, N., 2011. School access, resources, and learning outcomes: Evidence from a non-formal school program in Bangladesh. Bonn: Institute for the Study of Labor (IZA).

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K. and Sundararaman, V., 2011. School inputs, household substitution, and test scores. Cambridge, MA: National Bureau of Economic Research (NBER).

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K. and Sundararaman, V., 2013. School inputs, household substitution, and test scores. *American Economic Journal-Applied Economics,* 5 (2)**,** 29-57.

David, C. and Quispe, L. Evaluación del impacto del programa olpc sobre el rendimiento académico en el área de matemática. VII Congreso Iberoamericano de Educación Matemática, Montevideo Uruguay. Conference paper.

David, C. and Quispe, L. Evaluación del impacto del programa olpc sobre los procesos de mejoramiento de la educación pública. Nuevas Ideas en Informática Educativa TISE 2013, 2013 2013. Conference paper, Conference paper.

De Brauw, A. and Gilligan, D., 2011. Using the regression discontinuity design with implicit partitions: The impacts of comunidades solidarias rurales on schooling in El Salvador. Discussion Paper.

De Brauw, A., Gilligan, D. O., Hoddinott, J. and Roy, S., 2014. The impact of Bolsa Família on schooling.

De Janvry, A., Finan, F. and Sadoulet, E., 2006. Evaluating Brazil's Bolsa Escola program: Impact on schooling and municipal roles.

De Janvry, A., Finan, F. and Sadoulet, E., 2010. Local electoral incentives and decentralized program performance. Berkeley.

De Janvry, A., Finan, F. and Sadoulet, E., 2012. Local electoral incentives and decentralized program performance. *Review of Economics and Statistics,* 94**,** 672-685.

De Melo, G., Machado, A. and Miranda, A., 2014. The impact of a one laptop per child program on learning: Evidence from Uruguay. Bonn Germany: The Institute for the Study of Labor (IZA).

De Oliveira, P. R. and Kassourf, A. L., 2012. Impact evaluation of the Brazilian non-contributory pension program benefício de prestação continuada (bpc) on family welfare.

Departamento Nacional De, P., 2004. Programa familias en acción: Condiciones iniciales de los beneficiarios e impactos preliminares. Bogota D.C.

Departamento Nacional De, P., 2006. Evaluacion de politicas publicas: Programa de familias en accion: Condiciones initiales de los beneficiarios e impactos preliminares. Washington D.C.

Diagne, A., Lô, M. M., Sokhna, O. and Diallo, F. L., 2014. Evaluation of the impact of school canteen programs on internal efficiency of schools, cognitive acquisitions and learning capacities of students in rural primary schools in Senegal. Nairobi: PEP working paper.

Dinkelman, T. and Martinez, C., 2011. Investing in schooling in Chile: The role of information about financial aid for higher education. CEPS.

Dixon, P., Schagen, I. and Seedhouse, P., 2011. The impact of an intervention on children's reading and spelling ability in low-income schools in India. *School Effectiveness and School Improvement,* 22**,** 461-482.

Dubois, P., De Janvry, A. and Sadoulet, E., 2011. Effects on school enrollment and performance of a conditional cash transfer program in Mexico. *Journal of Labor Economics,* 30**,** 555-589.

Duflo, E., Berry, J., Mukerji, S. and Shotland, M., 2015. A wide angle view of learning: Evaluation of the CCE and LEP programs in Haryana, India. *3ie Impact Evaluation Report 22*.

Duflo, E., Dupas, P. and Kremer, M., 2007. Peer effects, pupil-teacher ratios, and teacher incentives: Evidence from a randomized evaluation in Kenya. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (JPAL).

Duflo, E., Dupas, P. and Kremer, M., 2009. Additional resources versus organizational changes in education: Experimental evidence from Kenya. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (JPAL).

Duflo, E., Dupas, P. and Kremer, M., 2012. School governance, teacher incentives, and pupil-teacher ratios: Experimental evidence from Kenyan primary schools. Cambridge, MA: National Bureau of Economic Research.

Duflo, E. and Dupas, P. K. M., 2015. School governance, teacher incentives, and pupil-teacher ratios: Experimental evidence from Kenyan primary schools. *Journal of Public Economics,* 123**,** 92-110.

Duflo, E., Hanna, R. and Ryan, S., 2008. Monitoring works: Getting teachers to come to school. London: Centre for Economic Policy Research.

Duflo, E., Hanna, R. and Ryan, S. P., 2012. Incentives work: Getting teachers to come to school. *American economic review,* 102**,** 1241-1278.

Duflo, E. D. P. K. M., 2011. Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review,* 101**,** 1739-1774.

Dumitrescu, A., Levy, D., Orfield, C. and Sloan, M., 2011. Impact evaluation of Niger's imagine program. Washington DC: Mathematica Policy Research.

Ebenezer, R., Gunawardena, K., Kumarendran, B., Pathmeswaran, A., Jukes, M. C. H., Drake, L. J. and Silva, N. D., 2013. Cluster-randomised trial of the impact of school-based deworming and iron supplementation on the cognitive abilities of schoolchildren in Sri Lanka's plantation sector. *Tropical Medicine and International Health,* 18**,** 942-951.

Edmonds Eric, V. and Schady, N., 2012. Poverty alleviation and child labor. *American Economic Journal-Economic Policy,* 4**,** 100-124.

Edmonds, E. V. and Schady, N., 2011. Poverty alleviation and child labor. *Working paper, Forthcoming in American Economic Journal: Economic Policy*.

Edmonds, E. V. and Shrestha, M., 2014. You get what you pay for: Schooling incentives and child labor. *Journal of Development Economics,* 111**,** 196-211.

Estrada, R., 2013. Rules rather than discretion: Teacher hiring and rent extraction. Unpublished manuscript.

Evans, D., Hausladen, S., Kosec, K. and Reese, N., 2014. Community-based conditional cash transfers in Tanzania: Results from a randomized trial. Report.

Evans, D. and Kremer, M., 2005. The impact of distributing school uniforms on children's education in Kenya. Santa Monica: RAND Corporation.

Evans, D., Kremer, M. and Ngatia, M., 2009. The impact of distributing school uniforms on children's education in Kenya. Unpublished working paper.

Evans, D., Kremer, M. and Ngatia, M., 2012. School costs, school participation and long-run outcomes: Evidence from Kenya. Unpublished manuscript.

Eyal, K. and Woolard, I., 2014. Cash transfers and teen education: Evidence from South Africa. Report.

Fernando, D., De, S., Damani, Carter, R., Mendis Kamini, N. and Wickremasinghe, R., 2006. A randomized, double-blind, placebo-controlled, clinical trial of the impact of malaria prevention on the educational attainment of school children. *The American journal of tropical medicine and hygiene,* 74**,** 386-93.

Ferrando, M., Machado, A., Perazzo, I. and Vernengo, A., 2010. Una primera evaluación de los efectos del plan ceibal en base a datos de panel. Unpublished working paper.

Ferré, C. and Sharif, I., 2014. Can conditional cash transfers improve education and nutrition outcomes for poor children in Bangladesh? Evidence from a pilot project. Evidence from a pilot project. Working Paper.

Filmer, D. and Schady, N., 2006. Getting girls into school: Evidence from a scholarship program in Cambodia. Report.

Filmer, D. and Schady, N., 2009. School enrollment, selection and test scores. *Policy Research Working Paper Series: 4998 for The World Bank*.

Filmer, D. and Schady, N., 2011. Does more cash in conditional cash transfer programs always lead to larger impacts on school attendance? *Journal of Development Economics,* 96**,** 150-157.

Filmer, D. and Schady, N., 2014. The medium-term effects of scholarships in a low-income country. *Journal of Human Resources,* 49**,** 663-94.

Freeman, M. C., Greene, L. E., Dreibelbis, R., Saboori, S., Muga, R., Brumback, B. and Rheingans, R., 2012. Assessing the impact of a school-based water treatment, hygiene and sanitation programme on pupil absence in Nyanza province, Kenya: A cluster-randomized trial. *Tropical Medicine & International Health,* 17 (3)**,** 380-391.

Fusades, 2009. Evaluación de impacto externa de la red solidaria. Report.

Gajigo, S. Closing the education gender gap: Estimating the impact of girls' scholarship program in the Gambia. 2012 2012 Tunis. African Development Bank.

Galasso, E., 2006. "With their effort and one opportunity": Alleviating extreme poverty in Chile. Working Paper.

Gallego, F., García, P., Lagos, F. and Stekel, Y., 2008. Estudio longitudinal: Evaluación de impacto del sistema nacional de evaluación de desempeño (sned) en la educación subvencionada. Santiago: Pontificia Universidad Católica de Chile.

Gallego, F. A., 2008. Efetos del sned en reultados del proceso educativo. Unpublished working paper.

Garcia, S. and Hill, J., 2009. The impact of conditional cash transfers on children's school achievement: Evidence from Colombia. Bogota.

Garcia, S. and Hill, J., 2009. Impact of conditional cash transfers on children's school achievement: Evidence from Colombia. *Journal of Development Effectiveness,* 2**,** 117-137.

Garlick, R., 2013. How price sensitive is primary and secondary school enrollment? Evidence from nationwide tuition fee reforms in South Africa. Unpublished working paper.

Garn, J. V., Greene, L. E., Dreibelbis, R., Saboori, S., Rheingans, R. D. and Freeman, M. C., 2013. A cluster-randomized trial assessing the impact of school water, sanitation and hygiene improvements on pupil enrolment and gender parity in enrolment. *Journal of Water, Sanitation and Hygiene for Development,* 3 (4)**,** 592-601.

Gertler, P., Patrinos, H. and Rubio-Codina, M., 2008. Empowering parents to improve education: Evidence from rural Mexico. Washington D.C.: The World Bank.

Gertler, P. J., Patrinos, H. A. and Rubio-Codina, M., 2012. Empowering parents to improve education: Evidence from rural Mexico. *Journal of Development Economics,* 99**,** 68–79.

Gitter Seth, R. and Barham Bradford, L., 2008. Women's power, conditional cash transfers, and schooling in Nicaragua. *World Bank economic review,* 22**,** 271-290.

Glewwe, P., Ilias, N. and Kremer, M., 2003. Teacher incentives. Cambridge, MA: National Bureau of Economic Research. NBER Working Paper No. 9671.

Glewwe, P., Ilias, N. and Kremer, M., 2010. Teacher incentives. *American Economic Journal-Applied Economics,* 2**,** 205-227.

Glewwe, P. and Kassouf, A. L., 2012. The impact of the Bolsa Escola/Familia conditional cash transfer program on enrollment, dropout rates and grade promotion in Brazil. *Journal of Development Economics,* 97**,** 505-517.

Glewwe, P., Kremer, M. and Moulin, S., 2007. Many children left behind? Textbooks and test scores in Kenya. Unpublished working paper.

Glewwe, P., Kremer, M. and Moulin, S., 2009. Many children left behind? Textbooks and test scores in Kenya. *American Economic Journal: Applied Economics,* 1(1)**,** 112-135.

Glewwe, P., Kremer, M., Moulin, S. and Zitzewitz, E., 2000. Retrospective vs. Prospective analyses of school inputs: The case of flip charts in Kenya. *NBER Working Paper Series No 8081.* Cambridge, MA: National Bureau of Economic Research (NBER).

Glewwe, P., Kremer, M., Moulin, S. and Zitzewitz, E., 2004. Retrospective vs. Prospective analyses of school inputs: The case of flip charts in Kenya. *Journal of Development Economics,* 74**,** 251 – 268.

Glewwe, P. and Maiga, E. W. H., 2011. The impacts of school management reforms in Madagascar: Do the impacts vary by teacher type? *Journal of Development Effectiveness,* 3**,** 435-469.

Glewwe, P., Park, A. and Zhao, M., 2014. A better vision for development: Eyeglasses and academic performance in rural primary schools in china. Minneapolis: University of Minnesota.

Glewwe, P. O. P., 2004. Evaluating the impact of conditional cash transfers on schooling: An experimental analysis of Honduras' PRAF program. Final report. Washington D.C.

Grigorenko Elena, L., Jarvin, L., Kaani, B., Kapungulya Paula, P., Kwiatkowski, J. and Sternberg Robert, J., 2007. Risk factors and resilience in the developing world: One of many lessons to learn. *Development and Psychopathology,* 19**,** 747-765.

Grogan, L., 2008. Universal primary education and school entry in Uganda. *Journal of African Economies,* 18**,** 183-211.

Gutiérrez, E. and Rodrigo, R., 2014. Closing the achievement gap in mathematics: Evidence from a remedial program in Mexico City. *Latin American Economic Review,* 23:14**,** 1-30.

Halliday, K. E., Okello, G., Turner, E. L., Njagi, K. and Mcharo, C., 2014. Impact of intermittent screening and treatment for malaria among school children in Kenya: A cluster randomised trial. *PLOS Medicine,* 11**,** 1-16.

Halpin, P. F., Torrente, C. and Aber, L., 2014. The early grade reading assessment: Its internal structure and use in impact evaluation. *Working paper*.

He, F., 2010. *Essays on education programs in developing countries.* Columbia University.

He, F., Linden, L. and Macleod, M., 2008. How to teach English in India: Testing the relative productivity of instruction methods within the Pratham English language education program. New York: Columbia University. Unpublished working paper.

He, F., Linden, L. L. and Macleod, M., 2007. Helping teach what teachers don't know: An assessment of the Pratham English language program. New York: Columbia University. Unpublished working paper.

He, F., Linden, L. L. and Macleod, M., 2009. A better way to teach children to read? Evidence from a randomized controlled trial. Unpublished manuscript.

Heinrich, C. J., 2005. Demand and supply-side determinants of conditional cash transfer program effectiveness: Improving the first-generation programs. Working Paper.

Hidalgo, D., Onofa, M., Oosterbeek, H. and Ponce, J., 2013. Can provision of free school uniforms harm attendance? Evidence from Ecuador. *Journal of Development Economics,* 103**,** 43–51.

Hoop De, J. and Rosati, F. C., 2012. Does promoting school attendance reduce child labour? Evidence from Burkina Faso's bright project. Bonn Germany: Institute for the Study of Labor (IZA).

Hoop De, J. and Rosati, F. C., 2014. Does promoting school attendance reduce child labor? Evidence from Burkina Faso's bright project. *Economics of Education Review,* 39**,** 78-96.

Hulett, J. L., Weiss, R. E., Bwibo, N. O., Galal, O. M., Drorbaugh, N. and Neumann, C. G., 2014. Animal source foods have a positive impact on the primary school test scores of Kenyan schoolchildren in a cluster-randomised, controlled feeding intervention trial. *British Journal of Nutrition,* 111**,** 875-86.

Humpage, S., 2013. *Essays on child development in developing countries.* Dissertation Submitted to the Faculty of University of Minnesota in Partial fulfilment of the Requirements for the Degree of Doctor of Philosophy.

Imbrogno, J., 2014. *Essays on the economics of education.* Dissertation Submitted in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy Economics from Carnegie Mellon University.

RTI International, 2014. USAID-Kenya primary math and reading (PRIMR) initiative- final report. RTI International.

RTI International, 2015. The primary math and reading (PRIMR) initiative: DFID/Kenya rural expansion programme. Bungoma and Machakos endline study. RTI International.

Irwing, P., Alya, H., Omar, K. and Richard, L., 2008. Effects of abacus training on the intelligence of Sudanese children. *Personality and Individual Differences,* 45**,** 694-696.

Ismail, S. J., Jarvis, E. A. and Borja-Vega, C., 2012. Guyana's hinterland community-based school feeding program (sfp). Washington D.C.: World Bank.

Jacoby, E., Cueto, S. and Pollitt, E., 1996. Benefits of a school breakfast programme among Andean children in huaraz, Peru. *Food and Nutrition Bulletin,* 17**,** 54-64.

Jayaraman, R. and Simroth, D., 2015. The impact of school lunches on primary school enrollment: Evidence from India's midday meal scheme. *Scandinavian Journal of Economics,* 117**,** 1176–1203.

Jensen, R., 2007. The perceived returns to education and the demand for schooling. Unpublished working paper.

Jensen, R., 2010. The (perceived) returns to education and the demand for schooling. *Quarterly Journal of Economics,* 125 (2)**,** 515-548.

Jukes M., C., Zuilkowski Stephanie, S., Parawan, A. and Lee, S., 2014. Evaluating the cognitive impacts of school health in the Philippines. *International Journal of Child Health and Human Development,* 7**,** 55-66.

Jukes, M. C. H. and Dubeck, M. M., 2015. Teacher professional development and text messages support for improved literacy instruction in Kenya: An experimental evaluation. *Working paper*.

Kazianga, H., De Walque, D. and Alderman, H., 2009. Educational and health impacts of two school feeding schemes: Evidence from a randomized trial in rural Burkina Faso. Washington D.C.: The World Bank.

Kazianga, H., De Walque, D. and Alderman, H., 2012. Educational and child labor impacts of two food for education schemes: Evidence from a randomized trial in rural Burkina Faso. Unpublished working paper.

Kazianga, H., De Walque, D. and Alderman, H., 2012. Educational and child labour impacts of two food-for-education schemes: Evidence from a randomised trial in rural Burkina Faso. *Journal of African Economies,* 21**,** 723-760.

Kazianga, H., Levy, D., Linden, L. L. and Sloan, M., 2012. The effects of "girl-friendly" schools: Evidence from the bright school construction program in Burkina Faso. *NBER Working Paper No. 18115.* Cambridge, MA: National Bureau of Economic Research (NBER).

Kazianga, H., Levy, D., Linden, L. L. and Sloan, M., 2013. The effects of "girl-friendly" schools: Evidence from the bright school construction program in Burkina Faso. *American Economic Journal: Applied Economics,* 5(3)**,** 41-62.

Kerwin, J. T. and Thornton, R., 2015. Making the grade: Understanding what works for teaching literacy in rural Uganda.

Khandker Shahidur, R., Pitt Mark, M. and Fuwa, N., 2003. Subsidy to promote girls' secondary education: The female stipend program in Bangladesh. Munich: Report.

Kharisma, B., Can a school operational assistance fund program (BOS) reduce school drop-outs during the post-rising fuel prices in Indonesia? Evidence from Indonesia. Unpublished working paper.

Khattri, N., Ling, C. and Jha, S., 2010. The effects of school-based management in the Philippines. An initial assessment using administrative data. *Policy Research Working Paper.* Washington D.C.: World Bank.

Khattri, N., Ling, C. and Jha, S., 2012. The effects of school-based management in the Philippines: An initial assessment using administrative data. *Journal of Development Effectiveness,* 4**,** 277-295.

Kim, J., Alderman, H. and Orazem, P., 1998. Can private schools subsidies increase schooling for the poor? The Quetta urban fellowship program. Washington D.C.: The World Bank.

Kleiman-Weiner, M., Luo, R., Zhang, L., Shi, Y., Medina, A. and Rozelle, S., 2013. Eggs versus chewable vitamins: Which intervention can increase nutrition and test scores in rural china? *China Economic Review,* 24**,** 165-176.

Kremer, M., Miguel, E. and Thornton, R., 2004. Incentives to learn. Cambridge, MA: National Bureau of Economic Research.

Kremer, M., Miguel, E. and Thornton, R., 2009. Incentives to learn. *The Review of Economics and Statistics,* 91**,** 437-456.

Kremer, M., Moulin, S. and Namunyu, R., 2003. Decentralization: A cautionary tale. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (J-PAL).

Kremer, M., Ozier, O., Thornton, R. and Miguel, E., 2005. Incentives to learn. Washington D.C.: The World Bank.

Lai, F., Luo, R., Zhang, L., Huang, X. and Rozelle, S., 2011. Does computer-assisted learning improve learning outcomes? Evidence from a randomized experiment in migrant schools in Beijing. Stanford, The United States: Rural Education Action Program (REAP).

Lai, F., Zhang, L., Hu, X., Qu, Q., Shi, Y., Qiao, Y., Boswell, M. and Rozelle, S., 2013. Computer assisted learning as extracurricular tutor? Evidence from a randomised experiment in rural boarding schools in shaanxi. *Journal of Development Effectiveness,* 5 (2)**,** 208-231.

Lai, F., Zhang, L., Qu, Q., Hu, X., Shi, Y., Boswell, M. and Rozelle, S., Teaching the language of wider communication, minority students and overall educational performance: Evidence from a randomized experiment in Qinghai province, china. Unpublished working paper.

Lai, F., Zhang, L., Shi, Y., Boswell, M. and Rozelle, S., 2012. Computer assisted learning as extracurricular tutor? Evidence from a randomized experiment in rural boarding schools in Shaanxi. Stanford the United States: Rural Education Action Program (REAP).

Lakshminarayana, R., Eble, A., Bhakta, P., Frost, C., Boone, P., Elbourne, D. and Mann, V., 2013. The support to rural India’s public education system (stripes) trial: A cluster randomised controlled trial of supplementary teaching, learning material and material support. *PLOS One,* 8 (7).

Lara, B., Mizala, A. and Repetto, A., 2009. Revisiting the school choice debate in Chile. Santiago: Centro de Economía Aplicada, Universidad de Chile.

Lara, B., Mizala, A. and Repetto, A., 2011. The effectiveness of private voucher education: Evidence from structural school switches. *Educational Evaluation and Policy Analysis,* 33**,** 119-137.

Lassibille, G., Tan, J. P., Jesse, C. and Van Nguyen, T., 2010. Managing for results in primary education in Madagascar: Evaluating the impact of selected workflow interventions. *World Bank Economic Review,* 24**,** 303-329.

Leme, M. C., Louzano, P., Ponczek, V. and Souza, A. P., 2009. O impactos dos metodos estruturados de ensino na proficiencia dos alunos da rede publica municipal do estado de Sao Paulo. *Unpublished working paper*.

Leme, M. C., Louzano, P., Ponczek, V. and Souza, A. P., 2012. The impact of structured teaching methods on the quality of education in Brazil. *Economics of Education Review,* 31**,** 850-860.

Levy, D. and Ohls, J., 2003. Evaluation of Jamaica’s path programme: Methodology report.

Levy, D. and Ohls, J., 2007. Evaluation of Jamaica’s path program: Final report. Report.

Levy, D. and Ohls, J., 2010. Evaluation of Jamaica’s path conditional cash transfer programme. *Journal of Development Effectiveness,* 2**,** 421-441.

Levy, D., Sloan, M., Linden, L. L. and Kazianga, H., 2009. Impact evaluation of Burkina Faso’s bright program: Final report. Washington DC: Mathematica Policy Research, Inc.

Li, T., Han, L., Zhang, L. and Rozelle, S., 2014. Encouraging classroom peer interactions: Evidence from Chinese migrant schools. *Journal of Public Economics,* 111**,** 29-45.

Li, T., Li, H., Rozelle, S. and Zhang, L., 2010. Cash incentives, peer tutoring, and parental involvement: A study of three educational inputs in a randomized field experiment in china. Stanford: REAP Working Paper 221.

Linden, L., Banerjee, A. and Duflo, E., 2003. Computer-assisted learning: Evidence from a randomized experiment. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (J-PAL).

Linden, L. L., 2008. Complement or substitute? The effect of technology on student achievement in India. Unpublished working paper.

Liu, C., Yi, H., Luo, R., Bai, Y., Zhang, L., Shi, Y., Wang, H., Chu, J. and Rozelle, S., 2013. The effect of early commitment of financial aid on matriculation to senior high school among poor junior high students in rural China. Stanford, CA: Rural Education Action Project (REAP).

Lo, M. M. and Diagne, A., 2012. The impact of school feeding on students' cognitive acquisitions in rural primary schools in Senegal. Consortium for Economic and Social Research (CRES).

Lockheed, M., Harris, A., Gammill, P., Barrow, K. and Jayasundera, T., 2006. New horizons for primary schools in Jamaica: Inputs, outcomes and impact. Washington D.C.: USAID & Academy for Educational Development (AED).

Lockheed, M., Harris, A. and Jayasundera, T., 2010. School improvement plans and student learning in Jamaica. *International Journal of Educational Development,* 30**,** 54-66.

Lokshin, M. and Yemtsov, M., 2004. Combining longitudinal household and community surveys for evaluation of social transfers: Infrastructure rehabilitation projects in rural Georgia. *Journal of Human Development,* 5 (2)**,** 265-277.

Loyalka, P., Liu, C., Song, Y., Yi, H., Huang, X., Wei, J., Zhang, L., Shi, Y., Chu, J. and Rozelle, S., 2013. Can information and counseling help students from poor rural areas go to high school? Evidence from china. *Journal of comparative economics,* 41**,** 1012-1025.

Loyalka, P., Sylvia, S., Liu, C., Chu, J. and Shi, Y., 2015. Pay by design: Teacher performance pay design and the distribution of student achievement. *Working Paper*.

Lucas, A. M., Mcewan, P. J., Ngware, M. and Oketch, M. Improving primary school quality across countries: Experimental evidence from Kenya and Uganda. Unpublished working paper, 2012

2012 Working paper 2012. Unpublished working paper.

Lucas, A. M., Mcewan, P. J., Ngware, M. and Oketch, M., 2014. Improving early-grade literacy in east Africa: Experimental evidence from Kenya and Uganda. University of Delaware Working Papers: 13-03.

Luseno, W. K., 2013. *Effect of the Mchinji social cash transfer pilot scheme on children's schooling, work and health outcomes: A multilevel study using experimental data.* Dissertation.

Macours, K. and Vakis, R., 2009. Changing households' investments and aspirations through social interactions: Evidence from a randomized transfer program. Working Paper.

Mahawithanage, S. T., Kannangara, K. K., Wickremasinghe, R., Chandrika, U. G., Jansz, E. R., Karunaweera, N. D. and Wickremasinghe, A. R., 2007. Impact of vitamin a supplementation on health status and absenteeism of school children in Sri Lanka. *Asia Pacific Journal of Clinical Nutrition,* 16**,** 94-102.

Maiga, E., 2010. *Essays on the economics of health and education in developing countries.* University of Minnesota.

Maluccio, J., Murphy, A. and Regalia, F., 2009. Does supply matter? Initial supply conditions and the effectiveness of conditional cash transfers for grade progression in Nicaragua. Middlebury, Vermont: Discussion paper.

Martínez, M. and Oscar, A., 2012. Effects of the scholarships of the 'Oportunidades' program on school attendance. The case of the urban zones in northeast Mexico. *Desarrollo y sociedad,* 63**,** 99-132.

Mcewan, P. and Santibáñez, L. 2005. Teacher and principal incentives in Mexico. *In:* Emiliana, V. (ed.) *Incentives to improve teaching: Lessons from Latin America.* Washington D.C.: World Bank.

Mcewan, P. J., 2013. The impact of Chile's school feeding program on education outcomes. *Economics of Education Review,* 32**,** 122-139.

Melo De, G., Machado, A. and Miranda, A., The impact of a one laptop per child program on learning. Evidence from uruguay1. Unpublished working paper.

Miguel, E. and Kremer, M., 2001. Worms: Education and health externalities in Kenya. Cambridge, MA: NBER Working Papers: 8481.

Miguel, E. and Kremer, M., 2004. Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica,* 72**,** 159–217.

Min, Y., Yanqing, D. and Wenbin, H., 2012. Southwest basic education project (SBEP): Analysis of the impact of SBEP on student achievement. Cambridge United Kingdom: Cambridge Education.

Mo, D., L, L. Z., Wang, J., Huang, W., Shi, Y., Boswell, M. and Rozelle, S., 2014. The persistence of gains in learning from computer assisted learning (CAL): Evidence from a randomized experiment in rural schools in Shaanxi province in china. Stanford: Rural Education Action Program (REAP).

Mo, D., Zhang, L., Luo, R., Qu, Q., Huang, W., Wang, J., Qiao, Y., Boswell, M. and Rozelle, S., 2014. Integrating computer-assisted learning into a regular curriculum: Evidence from a randomised experiment in rural schools in Shaanxi. *Journal of Development Effectiveness,* http://dx.doi.org/10.1080/19439342.2014.911770 [online].

Mo, D., Zhang, L., Qu, Q., Boswell, M. and Rozelle, S., 2013. Integrating computer assisted learning into a regular curriculum: Evidence from a randomized experiment in rural schools in Shaanxi. Unpublished working paper.

Mo, D., Zhang, L., Yi, H., Luo, R., Rozelle, S. and Brinton, C., 2011. School dropouts and conditional cash transfers: Evidence from a randomized controlled trial in rural china's junior high schools. Working Paper.

Mouton, J., 1995. Second language teaching for primary school students: An evaluation of a new teaching method. *Evaluation and Program Planning,* 18**,** 391-408.

Moya, S. H., 2012. *Avaliação da eficácia dos métodos estruturados de ensino nas escolas públicas municipais do estado de São Paulo.* Dissertation thesis.

Muradlidharan, K. and Sundararaman, V., 2010. Contract teachers: Experimental evidence from India. UC San Diego Working Paper.

Muralidharan, K., 2011. Long- term effects of teacher performance pay: Experimental evidence from india. Unpublished working paper.

Muralidharan, K. and Sundararaman, V., 2008. Teacher incentives in developing countries: Experimental evidence from India. Nashville: National Center on Performance Incentives, Working Paper 2008-13.

Muralidharan, K. and Sundararaman, V., 2009. Teacher performance pay: Experimental evidence from India. NBER Working Paper: 15323.

Muralidharan, K. and Sundararaman, V., 2010. The impact of diagnostic feedback to teachers on student learning: Experimental evidence from India. *Economic Journal,* 120**,** F187-F203.

Muralidharan, K. and Sundararaman, V., 2011. Teacher performance pay: Experimental evidence from India. *Journal of Political Economy,* 119**,** 39-77.

Muralidharan, K. and Sundararaman, V., 2013. The aggregate effect of school choice: Evidence from a two-stage experiment in India. Cambridge MA: NBER.

Muralidharan, K. and Sundararaman, V., 2013. Contract teachers: Experimental evidence from India. Cambridge, MA: National Bureau of Economic Research.

Murnane, R. J. J., Willet, J. B. and Cardenas, 2006. Did the participation of schools in programa escuelas de calidad (pec) influence student outcomes? Unpublished working paper.

Newman, J., Pradhan, M., Rawlings, L., Ridder, G., Coa, R. and Evia, J. L., 2002. An impact evaluation of education, health, and water supply investments by the Bolivian social investment fund. *World Bank Economic Review,* 16 (2)**,** 241–74.

Nguyen, T., 2008. Information, role models, and perceived returns to education: Experimental evidence from Madagascar. Unpublished working paper.

Nguyen, T. and Lassibille, G., 2008. Improving management in education: Evidence from a randomized experiment in Madagascar. Unpublished working paper.

Nonoyama-Tarumi, Y. and Bredenberg, K., 2009. Impact of school readiness program interventions on children's learning in Cambodia. *International Journal of Educational Development,* 29**,** 39-45.

Ome, A., 2012. The effects of meritocracy for teachers in Colombia.

Ome, A., 2012. Meritocracy for teachers: Evidence from Colombia. Working paper.

Omwami, E. M., Neumann, C. and Bwibo, N. O., 2011. Effects of a school feeding intervention on school attendance rates among elementary schoolchildren in rural Kenya. *Nutrition,* 27**,** 188-93.

Orkin, K., 2013. The effect of lengthening the school day on children’s achievement in Ethiopia. Oxford: Young Lives Working Papers.

Orkin, K., 2013. Lengthening the school day and children’s achievement in Ethiopia. Unpublished working paper.

Pallante, D. H. and Kim, Y. S., 2013. The effect of a multicomponent literacy instruction model on literacy growth for kindergartners and first-grade students in Chile. *International Journal of Psychology,* 48**,** 747-761.

Pandey, P., Goyal, S. and Sundararaman, V., 2009. Community participation in public schools: Impact of information campaigns in three Indian states. *Education Economics,* 17**,** 355-375.

Pandey, P., Goyal, S. and Sundararaman, V., 2011. Does information improve school accountability? Results of a large randomized trial. Washington D.C.: World Bank.

Paqueo, V. and Lopez-Acevedo, G., 2003. Supply-side school improvement and the learning achievement of the poorest children in indigenous and rural schools: The case of pare. Washington DC: The World Bank.

Pedro, C., Oswald, K., Nathalie, L., Costas, M. and Corina, M., 2015. Decentralizing education resources: School grants in Senegal.

Pellerano, L., Moratti, M., Jakobsen, M., Bajgar, M. and Barca, V., 2014. Child grants programme impact evaluation. Follow-up report. Report.

Pianto, D. and Soares, S. Use of survey design for the evaluation of social programs: The PNAD and PETI. Paper presented at the annual meetings of the Latin American and Caribbean Economic Association, October 2003, Puebla, Mexico, 2004/07.

Piper, B. and Korda, M., 2011. Early grade reading assessment (EGRA) plus: Liberia: Program evaluation report. Research Triangle Park, NC: RTI International.

Piper, B. and Mugenda, A., 2013. The primary math and reading (PRIMR) initiative midterm impact evaluation. Washington D.C.: USAID Kenya.

Piper, B. and Mugenda, A., 2014. The primary math and reading (PRIMR) initiative: Endline impact evaluation. Research Triangle Park, NC: RTI International.

Piper, B., Zuilkowski, S. S. and Mugenda, A., 2014. Improving reading outcomes in Kenya: First-year effects of the PRIMR initiative. *International Journal of Educational Development,* 37**,** 11-21.

Ponce, J. and Arjun, S. B., 2008. The impact of a cash transfer program on cognitive achievement: The bono de desarrollo humano of Ecuador. Discussion Paper.

Ponczek, V., 2011. Income and bargaining effects on education and health in Brazil. *Journal of Development Economics,* 94**,** 242-253.

Powell, C. A., Walker, S. P., Chang, S. M. and Grantham-Mcgregor, S. M., 1998. Nutrition and education: A randomized trial of the effects of breakfast in rural primary school children. *American Journal of Clinical Nutrition,* 68**,** 873–879.

Pradhan, M., Suryadarma, D., Beatty, A., Wong, M., Alishjabana, A., Gaduh, A. and Artha Rima, P., 2013. Improving educational quality through enhancing community participation: Results from a randomized field experiment in Indonesia. Washington D.C.: The World Bank.

Pradhan, M., Suryadarma, D., Beatty, A., Wong, M., Alishjabana, A., Gaduh, A. and Artha Rima, P., 2014. Improving educational quality through enhancing community participation: Results from a randomized field experiment in Indonesia. *American Economic Journal: Applied Economics 2014,* 6**,** 105–126.

Quiroga, F. and Bibiana, M., 2006. Trabajo infantil en los ninos y jovenes beneficiarios del programa familias en accion: Una evaluacion de impacto. 0**,** 71-109.

Rau, T. B. and Contreras, D. G., 2011. Tournaments incentives for teachers: The case of chile. Unpublished manuscript.

Raymond, M. and Sadoulet, E., 2003. Educational grants closing the gap in schooling attainment between poor and non-poor.

Reinikka, R. and Svensson, J., 2007. The returns from reducing corruption: Evidence from education in Uganda. Stockholm: CEPR Discussion Papers: 6363.

Renfu, L., Yaojiang, S., Linxiu, Z., Chengfang, L., Scott, R., Brian, S., Ai, Y., Qiran, Z. and Reynaldo, M., 2012. Nutrition and educational performance in rural china's elementary schools: Results of a randomized control trial in Shaanxi province. *Economic Development and Cultural Change,* 60**,** 735-72.

Ribas, R. P., Soares, F. V., Teixeira, C., Silva, E. and Hirata, G., 2011. Externality and behavioural change effects of a non-randomised CCT programme: Heterogeneous impact on the demand for health and education. Working Paper.

Robertson, L., Mushati, P., Eaton, J. W., Dumba, L., Mavise, G., Makoni, J., Schumacher, C., Crea, T., Monasch, R., Sherr, L., Garnett, G. P., Nyamukapa, C. and Gregson, S., 2013. Effects of unconditional and conditional cash transfers on child health and development in Zimbabwe: A cluster-randomised trial. *Lancet,* 381**,** 1283-1292.

Rodriguez, C., Sanchez, F. and Armenta, A., 2010. Do interventions at school level improve educational outcomes? Evidence from a rural program in Colombia. *World development,* 38 (3)**,** 415-428.

Rosati, F. and Rossi, M., 2007. Impact of school quality on child labor and school attendance: The case of CONAFE compensatory education program in Mexico. Rome: Understanding Children's Work (UCW) Project, University of Rome "Tor Vergata".

Saavedra Facusse, T. B., 2013. *Efecto del financiamiento compartido sobre el rendimiento escolar.* Universidad de Chile.

Sabarwal, S., Evans, D. K. and Marshak, A., 2014. The permanent input hypothesis: The case of textbooks and (no) student learning in Sierra Leone. Washington D.C.: World Bank Group Education Global Practice Group & Africa Region.

San, A. and Diosdado, M., 2008. Creating better schools through democratic school leadership. *International Journal of Leadership in Education: Theory and Practice,* 11**,** 43-62.

San Antonio, D. M., Morales, N. S. and Moral, L. S., 2011. Module-based professional development for teachers: A cost-effective Philippine experiment. *Teacher development,* 15**,** 157-169.

Santana, M. I., 2008. An evaluation of the impact of South Africa's child support grant on school attendance. Working Paper.

Santibanez, L., Abreu-Lastra, R. and O'donoghue J., L., 2014. School based management effects: Resources or governance change? Evidence from Mexico. *Economics of Education Review,* 39**,** 97-109.

Santibañez, L., Martínez, J. F., Datar, A., Mcewan, P. J., Setodji, C. M. and Basurto-Dávila, R., 2007. Breaking ground: Analysis of the assessment system and impact of Mexico's teacher incentive program 'carrera magisterial. RAND Corporation.

Schady, N. and Araujo, M. C., 2008. Cash transfers, conditions, and school enrollment in Ecuador. *Economia: Journal of the Latin American and Caribbean Economic Association,* 8**,** 43-70, 76.

Schultz, T. P., 2004. School subsidies for the poor: Evaluating the Mexican progresa poverty program. *Journal of Development Economics,* 74**,** 199-250.

Sean, S., Renfu, L., Linxiu, Z., Yaojiang, S., Alexis, M. and Scott, R., 2013. Do you get what you pay for with school-based health programs? Evidence from a child nutrition experiment in rural china. *Economics of Education Review,* 37**,** 1-12.

Sharma, D., 2011. *The impact of financial incentives on academic achievement and household behavior: Evidence from a randomized trial in Nepal.* Ohio State University.

Sharma, U. Can computers increase human capital in developing countries? An evaluation of Nepal's one laptop per child program. In 2014 Annual Meeting, July 27-29, 2014, Minneapolis, Minnesota (No. 169846). Agricultural and Applied Economics Association, 2014 2014 Minneapolis, Minnesota USA.

Simeon, D. T., Grantham-Mcgregor, S. M., Callender, J. E. and Wong, M. S., 1995. Treatment of trichuris trichiura infections improves growth, spelling scores and school attendance in some children. *Journal of Nutrition,* 125**,** 1875-1883.

Simeon, D. T., Grantham-Mcgregor, S. M. and Wong, M. S., 1995. Trichuris trichiura infection and cognition in children: Results of a randomized clinical trial. *Parasitology,* 110**,** 457-464.

Simwaka, B. N., Simwaka, K. and Bello, G., 2009. Retrospective analysis of a school-based malaria treatment programme demonstrates a positive impact on health and education outcomes in Mangochi district, Malawi. *Journal of Development Effectiveness,* 1**,** 492-506.

Skoufias, E. and Parker, S. W., 2001. Conditional cash transfers and their impact on child work and schooling: Evidence from the progresa program in Mexico. Food Consumption and Nutrition Devision, International Food Policy Research Institute.

Skoufias, E. and Shapiro, J., 2006. Evaluating the impact of Mexico's quality schools program: The pitfalls of using non-experimental data. Washington, D.C.: World Bank.

Sparrow, R., 2007. Protecting education for the poor in times of crisis: An evaluation of a scholarship programme in Indonesia*. *Oxford Bulletin of Economics and Statistics,* 69**,** 99-122.

Spratt, J., King, S. and Bulat, J., 2013. Independent evaluation of the effectiveness of institut pour l'education populaire's "read-learn-lead" (RLL) program in Mali. Endline report. Research Triangle Park, NC: RTI International (Prepared for William and Flora Hewlett Foundation).

Tan, J. P., Lane, J. and Lassibille, G., 1999. Student outcomes in Philippine elementary schools: An evaluation of four experiments. *World Bank Economic Review,* 13**,** 493–508.

Tokman, A., 2002. Evaluation of the p900 program: A targeted education program for underperforming schools. Central bank of Chile. Santiago Chile.

Valenzuela, J. P., 2005. *Partial evaluation of a big reform in the Chilean education system: From a half day to a full day schooling. University of Michigan doctoral dissertation.*

Vegas, E. and De Laat, J., 2003. Do differences in teacher contracts affect student performance? Evidence from Togo. Washington D.C.: World Bank/Brown University.

Watkins, W. E., Cruz, J. R. and Pollitt, E., 1996. The effects of deworming on indicators of school performance in Guatemala. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 90**,** 156-161.

Wolf, S., Torrente, C., Frisoli, P., Weisenhornc, N., Shivshanker, A., Annan, J. and Aber, J. L., 2015. Preliminary impacts of the "learning to read in a healing classroom" intervention on teacher well-being in the Democratic Republic of the Congo. *Working paper*.

Wong, H. L., Shi, Y. J., Luo, R. F., Zhang, L. X. and Rozelle, S., 2014. Improving the health and education of elementary schoolchildren in rural china: Iron supplementation versus nutritional training for parents. *Journal of Development Studies,* 50**,** 502-519.

World, B. and Ausaid, 2013. Basic education public expenditure review phase ii. School based management in the Philippines: An empirical investigation. Washington D.C.: World Bank/AusAID.

Yamauchi, F., 2014. An alternative estimate of school-based management impacts on students' achievements: Evidence from the Philippines. Washington D.C.: World Bank.

Yang, Y., Zhang, L., Zeng, J., Pang, X., Lai, F. and Rozelle, S., 2013. Computers and the academic performance of elementary school-aged girls in china's poor communities. *Computers & Education,* 60**,** 335-346.

Yi, H., Song, Y., Liu, C., Huang, X., Zhang, L., Bai, Y., Shi, Y., Yang, C., Wang, H., Loyalka, P., Chu, J. and Rozelle, S., 2012. Giving kids a head start: The impact of early commitment of financial aid on poor seventh grade students in rural china. Stanford, CA: Rural Education Action Project (REAP).

Yi, H., Yingquan, S., Chengfang, L., Xiaoting, H., Linxiu, Z., Yunli, B., Baoping, R., Yaojiang, S., Prashant, L., James, C. and Scott, R., 2015. Giving kids a head start: The impact and mechanisms of early commitment of financial aid on poor students in rural china. *Journal of Development Economics,* 113**,** 1–15.

Zavakou, A., (n.d.). An evaluation of the CCT program familias en acción on grade failure and school drop-out rates. *Working Paper for European University Institute, Florence.* Working Paper.

Zeitlin, A., Bategeka, L., Guloba, M., Kasirye, I. and Mugisha, F., 2012. Management and motivation in Ugandan primary schools: Impact evaluation final report. University of Oxford: Centre for the Study of African Economies.

Zhang, H., 2009. Magnet schools and student achievement: Evidence from a randomized natural experiment in china. Working paper.

Cash transfer impact evaluations: includable but not used

Angelucci, M., De Giorgi, G., Rangel, M. A. and Rasul, I., 2009. Family networks and school enrolment: Evidence from a randomized social experiment. Working Paper.

Angelucci, M., De Giorgi, G., Rangel, M. A. and Rasul, I., 2010. Family networks and school enrolment: Evidence from a randomized social experiment. *Journal of Public Economics,* 94**,** 197-221.

Attanasio, O., Battistin, E., Fitzsimons, E. and Vera-Hernandez, M., 2005. How effective are conditional cash transfers? Evidence from Colombia. London: Briefing Note.

Attanasio, O., Meghir, C. and Santiago, A., 2005. Education choices in Mexico: Using a structural model and a randomised experiment to evaluate progresa.

Attanasio, O., Meghir, C. and Santiago, A., 2010. Education choices in Mexico: Using a structural model and a randomized experiment to evaluate progresa. Working paper.

Attanasio, O. P., Meghir, C. and Santiago, A., 2011. Education choices in Mexico: Using a structural model and a randomized experiment to evaluate progresa. *Review of economic studies,* 75(1)**,** 37-66.

Baird, S., Chirwa, E., Mcintosh, C. and Özler, B., 2010. The short-term impacts of a schooling conditional cash transfer program on the sexual behavior of young women. Working paper.

Barham, T., Macours, K. and Maluccio, J. A., 2012. More schooling and more learning? Effects of a 3-year conditional cash transfer program in Nicaragua after 10 years. Report.

Barham, T., Macours, K. and Maluccio, J. A., 2013. More schooling and more learning? Effects of a three-year conditional transfer program in Nicaragua after 10 years. Working paper.

Behrman, J., Sengupta, P. and Todd, P., 2002. Progressing through progresa: An impact assessment of a school subsidy experiment. Report.

Behrman, J. R., Gallardo-Garcia, J., Parker, S. W., Todd, P. E. and Velez-Grajales, V., 2011. Are conditional cash transfers effective in urban areas? Evidence from Mexico. Working paper.

Behrman, J. R., Gallardo-Garcia, J., Parker, S. W., Todd, P. E. and Velez-Grajales, V., 2012. Are conditional cash transfers effective in urban areas? Evidence from Mexico. *Education Economics,* 20**,** 233-259.

Behrman, J. R., Gallardo-García, J., Parker, S. W., Todd, P. E. and Vélez-Grajales, V., 2010. How conditional cash transfers impact schooling and working behaviors of children and youth in urban Mexico. Working paper.

Behrman, J. R., Parker, S. and Todd, P. E., 2005. Long-term impacts of the oportunidades conditional cash transfer programme on rural youth in Mexico. Report.

Behrman, J. R., Parker, S. W. and Todd, P. E., 2004. Medium-term effects of the oportunidades program package, including nutrition, on education of rural children age 0-8 in 1997. Report.

Behrman, J. R., Parker, S. W. and Todd, P. E., 2007. Do school subsidy programs generate lasting benefits? A five-year follow-up of oportunidades participants. Working Paper.

Behrman, J. R., Parker, S. W. and Todd, P. E., 2009. Medium-term impacts of the oportunidades conditional cash-transfer program on rural youth in Mexico. Working paper.

Behrman, J. R., Parker, S. W. and Todd, P. E., 2009. Symposium: Impacts of the oportunidades program: Schooling impacts of conditional cash transfers on young children: Evidence from Mexico. *Economic Development and Cultural Change,* 57**,** 439-77.

Behrman, J. R., Parker, S. W. and Todd, P. E., 2011. Do conditional cash transfers for schooling generate lasting benefits? A five-year followup of Progresa/Oportunidades. *Journal of human resources,* 46**,** 93-122.

Behrman, J. R., Sengupta, P. and Todd, P., 2005. Progressing through progresa: An impact assessment of a school subsidy experiment in rural Mexico. *Economic Development and Social Change,* 54**,** 238–275.

Bobba, M. and Gignoux, J., 2011. Policy-induced social interactions and schooling decisions. Working Paper.

Bobonis, G. and Finan, F., 2002. Transfers to the poor increase the schooling of the non-poor: The case of Mexico's progresa program. Working paper.

Buddelmeyer, H. and Skoufias, E., 2004. An evaluation of the performance of regression discontinuity design on progresa. Working paper.

Bustelo, M., 2011. *Three essays on investments in children's human capital.* University of Illinois.

Cardoso, E. and Souza, A. P., 2003. The impact of cash transfers on child labor and school attendance in Brazil. Working paper.

Cattaneo, A. and Lalive, R., 2006. Social interactions and schooling decisions. Working paper.

Coady, D., 2000. Final report: The application of social cost-benefit analysis to the evaluation of progresa. Working paper.

Coady David, P. and Parker Susan, W., 2002. A cost-effectiveness analysis of demand- and supply-side education interventions: The case of progresa in Mexico. Washington D.C.: IFPRI.

Dammert Ana, C., 2009. Heterogeneous impacts of conditional cash transfers: Evidence from Nicaragua. *Economic development and cultural change,* 58**,** 53-84.

De Janvry, A., Finan, F., Sadoulet, E. and Vakis, R., 2004. Can conditional cash transfers serve as safety nets to keep children at school and out of the labor market?

De Janvry, A., Finan, F., Sadoulet, E. and Vakis, R., 2006. Can conditional cash transfer programs serve as safety nets in keeping children at school and from working when exposed to shocks? *Journal of Development Economics,* 79**,** 1709-1753.

Dominguez, V. and Marcos, E., 2011. Does the impact of Oportunidades program increase in highly competitive regions? 30**,** 79-111.

Dubois, P., De Janvry, A. and Sadoulet, E., 2004. Effects on school enrollment and performance of a conditional transfers program in Mexico. Berkeley.

Dubois, P. and Rubio-Codina, M., 2009. Child care provision: Semiparametric evidence from a randomized experiment in Mexico. CEPR Discussion Paper No. DP7203.

Dubois, P. and Rubio-Codina, M., 2011. Child care provision: Semiparametric evidence from a randomized experiment in Mexico.

Dubois, P. and Rubio-Codina, M., 2012. Child care provision: Semiparametric evidence from a randomized experiment in Mexico. *Annals of economics and statistics,* 105-106**,** 155-184.

Edmonds Eric, V. and Schady, N., 2009. Poverty alleviation and child labor. Cambridge, MA: National Bureau of Economic Research.

Ferreira Francisco, H. G., Filmer, D. and Schady, N., 2009. Own and sibling effects of conditional cash transfer programs: Theory and evidence from Cambodia.

Fitzsimons, E. and Mesnard, A., 2013. Can conditional cash transfers compensate for a father's absence? Washington D.C.: World Bank

Galiani, S. and Mcewan Patrick, J., 2013. The heterogeneous impact of conditional cash transfers. *Journal of Public Economics,* 103**,** 85-96.

Gignoux, J., 2009. Spillovers of conditional cash transfers across localities: Evidence from progresa. Paris.

Gitter Seth, R. and Barham, B., 2007. Women's power, conditional cash transfers and schooling in Nicaragua. University of Wisconsin, Agricultural and Applied Economics Working Paper.

Gitter Seth, R. and Barham, B., 2009. Conditional cash transfers, shocks, and school enrolment in Nicaragua. *Journal of development studies,* 45**,** 1747-1767.

Glewwe, P. and Kassouf Ana, L., 2008. The impact of the Bolsa Escola/Familia conditional cash transfer program on enrollment, grade promotion and dropout rates in Brazil.

Glewwe, P., Olinto, P. and De Souza, P. Z., 2003. Evaluating the impact of conditional cash transfers on schooling in honduras: An experimental approach.

Hasan, A., 2010. Gender-targeted conditional cash transfers: Enrollment, spillover effects and instructional quality. Working paper.

Heinrich Carolyn, J., 2007. Demand and supply-side determinants of conditional cash transfer program effectiveness. *World Development,* 35**,** 121-43.

Independent Evaluation Group, 2011. Do conditional cash transfers lead to medium-term impacts? Evidence from a female school stipend program in Pakistan. Washington DC.

Lalive, R. and Cattaneo, A., 2006. Social interactions and schooling decisions. IZA working paper

Lalive, R. and Cattaneo, M. A., 2009. Social interactions and schooling decisions. *Review of economics and statistics,* XCI**,** 457-477.

Maluccio, J. A., 2003. Education and child labor: Experimental evidence from a Nicaraguan conditional cash transfer program. Working paper.

Maluccio, J. A., 2005. Coping with the" coffee crisis" in Central America: The role of the Nicaraguan red de protección social. Washington D.C.: IFPRI

Manley, J., Fernald, L. and Gertler, P., 2015. Wealthy, healthy and wise: Does money compensate for being born into difficult conditions? *Applied Economics Letters,* 22**,** 121-126.

Miller, C. and Tsoka, M., 2012. Cash transfers and children's education and labour among Malawi's poor. *Development Policy Review,* 30**,** 499-522.

Oosterbeek, H., Ponce, J. and Schady, N., 2008. The impact of cash transfers on school enrollment: Evidence from Ecuador. Washington D.C.: World Bank

Parker, S., 2003. Evaluación del impacto de oportunidades sobre la inscripción escolar: Primaria, secundaria y media superior. Evaluación de resultados de impacto del programa de desarrollo humano oportunidades. Morelos.

Parker, S. and Skoufias, E., 2000. The impact of progresa on work, leisure and time allocation. IFPRI: Washington, D.C.

Parker, S. W., Todd, P. and Wolpin, K. I., 2006. Within-family program effect estimators: The impact of oportunidades on schooling in Mexico.

Ponce, J. N., 2008. *Education policy and performance. Evaluating the impact of targeted education programs in.* Shaker Publishing.

Ponczek, V., 2011. Income and bargaining effects on education and health in Brazil. Unpublished working paper.

Rubio-Codina, M., 2009. Intrahousehold time allocation in rural Mexico: Evidence from a randomized experiment. Unpublished working paper.

Sadoulet, E., Finan, F., De Janvry, A. and Vakis, R., 2004. Can conditional cash transfer programs improve social risk management? Lesson for education and child labor outcomes. Washington D.C.: World Bank.

Schady, N. and Araujo, M. C., 2006. Cash transfers, conditions, school enrollment, and child work: Evidence from a randomized experiment in Ecuador. Washington, DC.

Schultz, T. P., 2000. Final report: The impact of progresa on school enrollments. IFPRI: Washington, D.C.

Schultz, T. P., 2000. Impact of progresa on school attendance rates in the sampled population. IFPRI: Washington, D.C.

Schultz, T. P., 2000. School subsidies for the poor: Evaluating a Mexican strategy for reducing poverty. 54.

Schultz, T. P., 2001. School subsidies for the poor: Evaluating the Mexican progresa poverty program. Economic Growth Center Discussion Paper No. 834

Skoufias, E., 2001. Progresa and its impacts on the human capital and welfare of households in rural Mexico: A synthesis of the results of an evaluation by IFPRI. Washington, D.C.

Skoufias, E., 2005. Progresa and its impacts on the welfare of rural households in Mexico. 84.

Todd, P. E., Gallardo-Garcia, J., Behrman, J. R. and Parker, S. W. 2005. Impacto de oportunidades sobre la educación de niños y jóvenes de áreas urbanas después de un año de participación en el programa. *In:* Hernández-Prado, B.andHernández-Ávlia, M. (eds.) *Evaluación externa de impacto del programa oportunidades.* Mexico.

Valdés, N., 2008. Did progresa send drop-outs back to school? Unpublished working paper.

included additional documents to answer objective 2

Altman, E., 2013. Childhood nutrition in Chile: An assessment of a national school food program*. Thesis submitted to the Faculty of Emory College of Arts and Sciences of Emory University in partial fulfilment of the requirements of the degree of Bachelor of Arts with Honors.* Atlanta: Emory College.

Asociacion Computadores Para, E., 2010. Informe de auditoría gubernamental con enfoque integral, modalidad regular. Bogotá: Contraloria.

Azim Premji Foundation Andhra Pradesh school choice research study objectives and methodology. Bangalore: Azim Premji Foundation.

Azim Premji, F., 2010. Findings from the Andhra Pradesh randomized evaluation series: No silver bullets. Bangalore: Azim Premji Foundation.

Baines, S., 2005. Towards more transparent financial management: Scholarships and grants programmes in Indonesia. Paris: International Institute for Educational Planning, UNESCO.

Banerjee, A., Banerji, R., Duflo, E., Glennerster, R. and Khemani, S., 2006. Can information campaigns spark local participation and improve outcomes? A study of primary education in Uttar Pradesh, India.

Bellei, C., Osses, A. and Valenzuela, J. P., 2010. Asistencia técnica educativa: De la intuición a la evidencia. Santiago: CONICYT; Universidad de Chile; FONDEF.

Bellei, C., Raczynski, D. and Osses, A., 2010. Asistencia técnica educativa en Chile: ¿Aporte al mejoramiento escolar? Santiago: Universidad de Chile.

Beyer, H. and Eyzaguiree, S., 2014. Fin del financiamiento compartido - análisis del proyecto de ley y propuestas. *Puntos de Referencia,* 374.

Borja, S. 2004. *Orientaciones para las instituciones educativas huascarán año escolar, directiva nº 06-2004/p. Huascarán* [Online]. Available: http://www.minedu.gob.pe/normatividad/directivas/dir06-2004-PHuascaran.php.

Bracho, T., Mendieta, G., Martínez, A., Camacho, A., Jacobo, M. and Núñez, M., 2006. Evaluación externa del programa escuelas de calidad (pec) 2005-2006. Mexico City: Centro de Investigación y Docencia Económicas, A.C., Administración Pública.

Branson, N. and Leibbrandt, M., 2013. Education quality and labour market outcomes in South Africa. Paris: OECD Economics Department.

Burde, D. 2011. Innovative methods in education in emergencies research: A randomized trial assessing community-based schools in Afghanistan. *In:* Mundy, K. and Dryden-Petersen, S. E. (eds.) *Educating children in conflict zones.*

Caoli-Rodriguez, R. B., 2007. The Philippines country case study - country profile prepared for the education for all global monitoring report 2008 education for all by 2015: Will we make it? Country profile commissioned for the EFA global monitoring report 2008, education for all by 2015. Paris: UNESCO.

Carlos, D. and Quispe, L., 2014. El programa ulpn del Perú: Caracterización de las prácticas de enseñanza con incorporación de computadoras portátiles. *CINTED-UFRGS,* 12**,** 1-11.

Castañeda, T., 2005. Targeting social spending to the poor with proxy–means testing: Colombia's Sisben system. Washington DC USA: The World Bank.

Cf, B. T. E. T., 2010. Foundation assisted schools, Pakistan. Cambridge: CfBT Education Trust.

Chauhan, S. D., 2011. *A study of mid-day meal programme in the government primary schools of the Gwalior city of Madhya Pradesh (a thesis submitted to the Maharaja Sayajirao University of Baroda for the degree of doctor of philosophy in education).* Center of Advanced Study in Education, Faculty of Education and Psychology.

Creative Associates International, I., 2004. A healthy child in a healthy environment: A look at the changes program in Zambia. Washington D.C.: Creative Associates International.

CrisóStomo, A. L. M., 2004. Plano de desenvolvimento da escola – PDE: O que mudou nas escolas públicas estaduais de teresina? : Unpublished manuscript.

Cristia, J., 2010. Evaluación cualitativa del programa “una laptop por niño”. Unpublished manuscript.

Cyan, M. F. and Young, A., Contract employment policy review. Prepared for Punjab government efficiency improvement program. Government of Punjab with the support of the Asian Development Bank.

De Pascual Pola, D. M. N., 2009. Análisis de las políticas para maestros de educación básica en México. OCDE.

Departamento Nacional De Planeación República De, C., 1999. Programa de donación de computadores a colegios públicos "computadores para educar". Santa Fe de Bogotá: Departamento Nacional de Planeación, República de Colombia.

Dhushyanth, R., 2011. The PPRS program. Unpublished manuscript (provided by author).

Directorate of Primary, E., 2012. Bangladesh reaching out-of-school children (ROSC) project-detailed procurement plan & processing schedule for goods & services 2004-2012. Dhaka: Directorate of Primary Education.

Dos Santos, F. A., 2012. *O plano de desenvolvimento da escola e a produção do consentimento ativo.*

Dreze, J. and Aparajita, G., 2003. Future of mid-day meals. *Economic and Political Weekly,* 38**,** 4673-4683.

Du Bois, F., Cusato, A. and Charez, J. F., 2004. Un balance de las políticas sociales. Programas sociales, salud y educación en el perú. Democratizando el gasto social. Lima: Instituto Peruano de Economía Social de Mercado.

Dyer, C., 1996. The improvement of primary school quality in India: Successes and failures of 'operation blackboard. Edinburgh: Centre for South Asian Studies.

Dyer, C., 1999. Researching the implementation of educational policy: A backward mapping approach. *Comparative Education,* 35**,** 45-61.

Eduardo Villanueva, M. and Paz, O., 2012. Institutional barriers to development innovation: Assessing the implementation of xo-1 computers in two peri-urban schools in Peru. *Information Technologies & International Development***,** 177-189.

Ekaju, J., 2011. *An investigation into the relationship between the 1997 universal primary education (UPE) policy and regional poverty and educational inequalities in Uganda (1997-2007).* University of Glasgow.

Entwicklungsbank, K. F. W., 2005. Bolivia: Social investment fund I + II. Frankfurt: KFW.

Epstein, R., Henriquez, L., Catalan, J., Weintraub, G. Y., Martinez, C. and Espejo, F., 2004. A combinatorial auction improves school meals in Chile: A case of or in developing countries. *International Transactions in Operational Research,* 11**,** 593–612.

Essama-Nssah, B., 2011. Achieving universal primary education through school fee abolition: Some policy lessons from Uganda. Yes Africa Can, 465.

Fernando Reimers Sergio, C., 2007. Who benefits from school-based management in Mexico? *Prospects,* 37**,** 37-56.

Freund, P., Graybill, E. and Keith, N., 2005. Health and education working together: A case study of a successful school health and nutrition model. Washington D.C: Creative Associates International Inc.

Giese, S., Zide, H., Koch, R. and Hall, K., 2009. A study on the implementation and impact of the no-fee and exemption policies. Cape Town South Africa: Alliance for Children's Entitlement to Social Security.

GutiéRrez, J. A., 2005. Reforma educativa en méxico: El programa escuelas de calidad. *REICE - Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación,* 1**,** 1-15.

Hollingsworth, S. and Gains, P., 2009. Integrated education program: The systematic method for reading success (SMRS) in South Africa (lessons learned from SMRS mastery tests and teacher performance check lists). Washington D.C.: USAID.

Honeyman, C. A., 2010. Social responsibility and community development: Lessons from the Sistema de Aprendizaje tutorial in Honduras. *International Journal of Educational Development,* 30**,** 599–613.

Hove, J. K., 2007. *Barriers to girls' secondary school participation in rural Bangladesh.*

Inter-American Development, B., Aligning learning incentives - plan of operations. Washington D.C.: Inter-American Development Bank.

Inter-American Development, B., 2007. Ali - technical cooperation profile. Washington D.C.: Inter-American Development Bank.

International, C., 2006. Measuring effectiveness of mid-day meal scheme in Rajasthan: Participatory expenditure tracking survey. Jaipur: CUTS Centre for Consumer Action, Research & Training.

International, R. T. I., 2011. Understanding the causal mechanisms: EGRA plus's effectiveness. Washington D.C.: USAID.

International, R. T. I., 2013. USAID-Kenya primary math and reading (PRIMR) initiative: Annual report.

Irarrázaval, I. and Paredes, R., 2012. Evaluación de los primeros años de implementación del programa de subvención escolar preferencial, de la subsecretaría de educación; marjorie murray; gabriel gutiérrez; francisca bogolasky; cristóbal contreras. *Centro de Políticas Públicas.* Santiago: Pontificia Universidad Católica de Chile.

J, P. a. L., 2011. J-pal test scores cea Kenya ETP budget & cost-effectiveness analysis. Provided on request by J-PAL.

Jacoby, E. R., Cueto, S. and Pollitt, E., 1998. When science and politics listen to each other: Good prospects from a new school breakfast program in Peru. *American Journal of Clinical Nutrition,* 67**,** 795S-797S.

Juan, A. L., 2007. *The no-fee schools policy: A case study of policy implementation in four KwaZulu-Natal schools.* University of KwaZulu-Natal, South Africa.

Kain, J., Uauy, R. and Taibo, M., 2002. Chile’s school feeding programme: Targeting experience. *Nutrition Research,* 22**,** 599–608.

Kampuchean Action for Primary, E., 2004. School readiness program: Assessment report. Phnom Penh: Kampuchean Action for Primary Education (KAPE).

Karlstrom, K., School readiness program Cambodia. UNICEF.

Khera, R., 2008. Mid-day meals in primary schools: Achievements and challenges. *Economic and Political Weekly,* 41**,** 4742-4750.

Kimura, I., 2008. Policy-level improvement and institutionalization of field-level trials: Achievement of third elementary education project in the Philippines. Unpublished manuscript.

King, E., Rawlings, L., Gutierrez, M., Pardo, C. and Torres, C., 1997. Colombia’s targeted education voucher program: Features, coverage and participation. Washington D.C.: World Bank.

King, E. M., Orazem, P. F. and Wohlgemuth, D., 1999. Central mandates and local incentives: The colombia education voucher program. *World Bank Economic Review,* 13**,** 467-491.

Larreta, H. R., 2001. Transferencia de recursos para programas alimentarios en las provincias: Un analisis de lo sucedido en los años ‘90. Unpublished manuscript.

Leaps. *Learning and educational achievement in Punjab schools (leaps): Project overview* [Online]. LEAPS project document. Available: http://leapsproject.org/assets/publications/LEAPS_project_overview.pdf.

Liang, X., 1996. Bangladesh: Female secondary school assistance. Washington D.C.: World Bank.

Louzano, P., Kameoka, M., Rocha, V. and Borges, A., 2008. Sistemas estruturados de ensino e redes municipais do estado de São Paulo. PowerPoint provided by the authors.

Mahmud, S., 2003. Female secondary school stipend programme in Bangladesh: A critical assessment. Dhaka: Bangladesh Institute of Development Studies.

Marishane, R. N., 2013. Empowerment and accountability in implementing a ‘no-fee school’ policy: A challenge for school governing bodies. *Africa Education Review,* 10**,** 224-237.

Martinic, S., Huepe, D. and Madrid, A., 2008. Jornada escolar completa en chile. Representaciones de los profesores sobre sus efectos en los aprendizajes. *Revista Iberoamericana de Evaluación Educativa 2008,* 1**,** 124-139.

Mayer, P., 2004. An evaluation of the Programa de Ampliación de la Educación Secundaria (PACES): The use of educational vouchers in Colombia. Unpublished working paper.

Mejía, M. I. and Bernal, P., 2003. Computadores para educar: Enriqueciendo la formación de las nuevas generaciones de colombianos. Instituto para la Conectividad en las Américas (ICA).

Mestry, R., 2013. A critical analysis of legislation on the financial management of public schools: A South African perspective. *De Jure,* 1**,** 162-177.

Ministerio de Economía y Finanzas Peru, 2010. *Evaluation de diseno y ejecucion presupuestal - edep programa "una laptop por nino"* [Online]. Lima: Ministerio de Economía y Finanzas Peru. Available at: http://www.gobernabilidad.org.pe/buen_gobierno/galleries/181729586_095%20EDEP%20Una%20laptop%20por%20nino.pdf.

Ministry of Education and Sports, the Republic of Uganda, 1999. The Ugandan experience of universal primary education (UPE). Kampala: Ministry of Education and Sports.

Ministry of Human Resource Development; Department of School Education, Government of India, 2006. National programme of nutritional support to primary education, 2006 (mid-day meal scheme) guidelines. New Delhi: Ministry of Human Resource Development, Department of School Education, Literacy, Government of India.

Ministry Of, P. and Government of People’s Republic Of, B., 2014. Impact evaluation study of reaching out of school children (ROSC) project of the ministry of primary and mass education. Ministry of Planning, Government of People's Republic of Bangladesh.

Mizala, A. and Urquiola, M., 2013. School markets: The impact of information approximating schools’ effectiveness. *Journal of Development Economics,* 103**,** 313-335.

Muralidharan, Andhra Pradesh school choice project proposal. Unpublished manuscript.

Murnane, R. J., Page, L. and Vegas, E., 2010. Distribution of student achievement in Chile baseline analysis for the evaluation of the subvención escolar preferencial, SEP (preferential school subsidy). *Education Sector Review.* Unpublished manuscript.

Murphy-Graham, E. and Lample, J., 2014. Learning to trust: Examining the connections between trust and capabilities friendly pedagogy through case studies from Honduras and Uganda. *International Journal of Educational Development,* 36**,** 51-62.

Nkosi, C. S., 2011. *The implications of being declared a no fee school.* The Faculty of Education, University of Pretoria.

Orazem, P. F., 2000. The urban and rural fellowship school experiments in Pakistan: Design, evaluation, and sustainability. Washington D.C.: World Bank.

Pratham InfoTech Foundation, 2009. Computer assisted learning program: Annual report. Mumbai: Pratham InfoTech Foundation.

Programme Evaluation, O., Planning, C. and Government Of, I., 2010. Performance evaluation of the cooked mid-day meal (cmdm). New Delhi India: PEO Report 202.

Puhl, C., 1991. Does the teaching of thinking promote language acquisition? *PER LINGUAM,* 7**,** 27-36.

Rao, M. S., 2011. Achieving universal quality primary education in India - lessons from the Andhra Pradesh randomized evaluation studies (AP Rest). Azim Premji Foundation.

Reinikka, R. and Svensson, J., 2004. Local capture: Evidence from a central government transfer program in Uganda. *The Quarterly Journal of Economics,* 119**,** 679-705.

Richards, S., 2010. Improving maths education in rural Honduras- a research study on maths teaching and learning in the Sistema de Aprendizaje tutorial (SAT). New York: Ford Foundation.

Romaguera, P. and Gallegos, S., 2010. Financiando la educación de grupos vulnerables: La subvención escolar preferencial□. Unpublished working paper.

Sapelli, C. and Vial, B., 2002. The performance of private and public schools in the Chilean voucher system. *Cuadernos de Economía-Latin American Journal of Economics,* 39**,** 423-454.

Secretaria De, E., Programa escuelas de calidad.

Secretaria De, E., 2015. Information on the concession schools program in Colombia. Documentation provided by the Secretery of Education of Bogotà, 2015 (direct communication).

Secretaría De Educación De, H. and Asociación De Desarrollo Socio-Económico, I., 2004. Sistema de aprendizaje tutorial. Secretaría de Educación de Honduras.

Secretaría De Educación, P. B., 2005. Reglas de operación del programa escuelas de calidad. Secretaría de Educación Pública.

Setoaba, M. P., 2009. *The implementation of the 'no-fee' school policy in selected primary schools in Limpopo.*

Sindh Education, F. *Promoting private schooling in rural Sindh - a public-private partnership project based on per-child subsidy model* [Online]. Karachi: Sindh Education Foundation. Available from: http://www.sef.org.pk/pdf/current_initiatives/PPRS_Brochure.pdf.

Society for Community Strengthening; Promotion of Education; Balochistan SCSPEB, n.d. Fellowship schools. Unpublished manuscript.

Sosa, A., Hernando, H., GóNgora, M. and Paola, G., Acompañamiento educativo en el proceso de apropiación de la tecnología de la información por comunidades colombianas. - proyecto computadores para educar.

Sotomayor, C., 2006. Public programs to improve the quality of primary schools in vulnerable urban contexts: Intervention strategies evolution and learnings (1990-2005). *Revista Pensamiento Educatitivo,* 39.

Sotomayor, C. and Dupriez, V., 2007. Desarrollar competencias docentes en la escuela: Aprendizaje de una experiencia chilena de asesor´ia a escuelas de alta vulnerabilidad social y educativa. *Les Cahiers de Recherche en E´ducation et Formation.* Les Cahiers de Recherche en Éducation et Formation Number 61.

Spratt, J. E., 2013. Independent evaluation of the effectiveness of institut pour l'education populaire's "read-learn-lead" (RLL) program in mali - cost analysis of key elements of the RLL program (educator training, pedagogical support, and pedagogical materials inputs). Research Triangle Park, NC: RTI International.

Spratt, J. E. and Ralaingita, W., 2013. Evaluation of Mali's mother-tongue early grade "read learn read" program year 2 technical follow-up report with focus on teacher and classroom practice. Durham, NC: RTI International.

Suharyo, W. I., 2006. A rapid appraisal of the PKPS-BBM education sector: School operational assistance (BOS). Jakarta: SMERU Research Institute.

Sumarto, S., Suryadarma, D., Widyanti, W. and Suryahadi, A. Local monitoring system during the implementation of Indonesia's social safety net programs with special reference to the BKKBN system. Paper presented at Poverty and Economic (PEP) Research Network Meeting, Senegal, June 16-20, 2004. Poverty and Economic Policy (PEP) research network working paper.

Sumarto, S., Suryahadi, A. and Widyanti, W., 2002. Designs and implementation of the Indonesian social safety net programs. *The Developing Economies,* XL-1**,** 3-31.

The Azim Premji Foundation, 2011. *Research studies (research test bed in Andhra Pradesh (APRest)) [online]* [Online]. Azim Premji Foundation (Website) [Accessed on the 25th of September 2014]. Available: http://www.azimpremjifoundation.org/Research_Studies.

The World, B., 2006. Colombia: Contracting education services [report no. 31841-co]. Washington D.C.: World Bank.

Unknown. *Examples of best practices in science and mathematics teaching and learning in Peru* [Online]. Unpublished manuscript. Available at: http://hrd.apec.org/images/f/f9/45.18.pdf. Available: http://hrd.apec.org/images/f/f9/45.18.pdf.

USAID, 2010. Audit of USAID/Afghanistan's partnership for advancing community-based education in Afghanistan (pace-a) program. Manila: Office of Inspector General, USAID.

Vasilades, K., 1993. Memorandum recommendation of the president of the international development association to the executive directors on a proposed credit of SDR 28.3 million to the republic of Bolivia for the second social investment fund project (SIF). Washington D.C.: World Bank.

Villa, L. and Duarte, J., 2002. Los colegios en concesión de Bogotá, Colombia: Una experiencia innovadora de gestión escolar reformas o mejoramiento continuo. Washington, D.C.: Banco Interamericano de desarrollo.

Villagran, L. 2013. *Test the teacher? Educators balk at Mexico's reforms* [Online]. Available: http://www.csmonitor.com/World/Americas/2013/0301/Test-the-teacher-Educators-balk-at-Mexico-s-reforms.

Winch, R., 2009. International approaches to school feeding: Country experiences from Mali, Chile, and India. Seattle: Global Child Nutrition Foundation.

World Bank, 1996. Memorandum and recommendation of the president of the international bank for reconstruction and development to the executive directors on a proposed loan in the amount of us $113.4 million to the republic of the Philippines for a third elementary education project. Washington D.C.: World Bank.

World Bank, 1996. Philippines - third elementary education project. Washington DC USA: World Bank.

World Bank, Operations Evaluations, Department, 1997. Ten years of social fund activities in Bolivia. *OED Precis.* Washington D.C.: World Bank.

World Bank, 1998. Report no. Pid691. Washington D.C.: World Bank.

World Bank, 2001. Colombia - secondary education project. Implementation, completion and results report. Washington D.C.: World Bank.

World Bank, 2001. Implementation completion report on a credit in the amount of sdr12.5 million to Georgia for a municipal infrastructure rehabilitation project. Washington D.C.: World Bank.

World, Bank, 2001. Implementation completion report on a loan/credit/grant in the amount of US$40 million to Bolivia for a social investment fund ii. Washington D.C.: World Bank.

World, Bank, 2013. Implementation completion and results report on a grant in the amount of SDR 35.2 million for a reaching out-of-school children project and additional financing credit in the amount of SDR 23.1 million to the people's republic of Bangladesh. Washington D.C.: World Bank.

World, Bank, Human and Social Development, G., 1996. Implementation completion report, Bolivia, social investment fund (credit 2127-bo). Washington D.C.: World Bank.

World Food Programme, 2006. World food programme Lao PDR: Annual report 2006. Rome: World Food Programme.

World Food Programme, 2007. World food programme Lao PDR: Annual report 2007. Rome: World Food Programme.

Additional references

Adelman, W., Gilligan, D. O. and Lehrer, K., 2008. How effective are food for education programs? A critical assessment of the evidence from developing countries. Washington D.C.: International Food Policy Research Institute.

Agarwal, S., Heltberg, R. and Diachok, M., 2009. Scaling up social accountability in World Bank operations. Washington, D.C.: World Bank.

Alderman, H., Kim, J. and Orazem, P. F., 2003. Design, evaluation, and sustainability of private schools for the poor: The Pakistan urban and rural fellowship school experiments. *Economics of Education Review,* 22**,** 265–274.

Al-Samarrai, S. and Zaman, H., 2007. Abolishing school fees in Malawi: The impact on education access and equity. *Education Economics,* 15**,** 359-375.

Andrabi, T., Das, J. and Khwaja, A., 2009. Report cards: The impact of providing school and child test-scores on educational markets. Unpublished manuscript.

Armstrong, R. and Waters, E., 2007. Systematic reviews of health promotion and public health interventions guidelines [online].

ASER, 2011. The annual status of education report (rural) 2010. New Delhi: Pratham.

Baird, S., Ferreira, F., Özler, B. and Woolcock, M., 2012. Relative effectiveness and cost-effectiveness of conditional and unconditional cash transfers for schooling outcomes in developing countries: A systematic review (protocol). Campbell Collaboration.

Baird, S., Ferreira, F., Özler, B. and Woolcock, M., 2013. Relative effectiveness of conditional and unconditional cash transfers for schooling outcomes in developing countries: A systematic review. Campbell Systematic Reviews.

Barrera-Osorio, F., Bertrand, M., Linden, L. L. and Perez, F., 2007. Using conditional transfers in education to investigate intra family decisions: Evidence from a randomised experiment. Unpublished working paper.

Barrera-Osorio, F., Bertrand, M., Linden, L. L. and Perez-Calle, F., 2008. Conditional cash transfers in education design features, peer and sibling effects evidence from a randomized experiment in Colombia. Cambridge, MA: National Bureau of Economic Research.

Barrera-Osorio, F., Fasih, T. and Patrinos, A., 2009. Decentralized decision-making in schools: The theory and evidence on school-based management. Washington D.C.: World Bank.

Barrera-Osorio, F. and Filmer, D., 2013. Incentivizing schooling for learning: Evidence on the impact of alternative targeting approaches. Washington D.C.: The World Bank.

Barro, R. J., 1991. Economic growth in a cross section of countries. *The Quarterly Journal of Economics,* 106**,** 407-443.

Becker, B., Hedges, L. and Pigott, T., Statistical analysis policy brief. Oslo: Campbell Collaboration.

Bennell, P., 2004. Teacher motivation and incentives in sub-Saharan Africa and Asia. Brighton: Knowledge and Skills for Development.

Bentaouet, K., 2006. Implementation of free basic education policy. Washington D.C.: World Bank.

Bergeron, G. and Del Rosso, J. M., 2001. Food for education indicator guide. Washington D.C.: USAID.

Birdsall, N. and Orivel, F., 1996. Demand for primary schooling in rural Mali: Should user fees be increased? *Education Economics,* 4**,** 279-296.

Birdthistle, I., Dickson, K. and Freeman, M. L. J., 2011. What is the impact of separate toilets for girls at schools on girls' educational outcomes? A systematic review of the evidence. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Bloem, S., 2015. Pisa for low-and middle-income countries. *Compare: A Journal of Comparative and International Education,* 45**,** 481-486.

Borenstein, M., Hedges, L. V., Higgins, J. P. T. and Rothstein, H., 2009. Introduction to meta-analysis (statistics in practice). Chichester: John Wiley & Sons.

Borkum, E., 2012. Can eliminating school fees in poor districts boost enrollment? Evidence from South Africa. *Economic Development and Cultural Change,* 60**,** 359-398.

Bouillon, C. and Tejerina, L., 2007. Do we know what works? A systematic review of impact evaluations of social programs in Latin America and the Caribbean. Washington D.C.: Inter-American Development Bank. Sustainable Development Dept. Poverty and Inequality Unit.

Bruns, B., Filmer, D. and Patrinos, H. A., 2011. Making schools work: New evidence on accountability reforms. Washington D.C.: World Bank.

Bundy, D., 2005. School health and nutrition: Policy and programs. *Food and Nutrition Bulletin,* 26**,** S186-S192.

Bundy, D., Burbano, C., Grosh, M., Gelli, A., Jukes, M. and Drake, L., 2009. Why implement school feeding? Rethinking school feeding: Social safety nets, child development, and the education sector. Washington, DC: World Bank.

Bundy, D., Shaeffer, S., Jukes, M., Beegle, K., Gillespie, A., Drake, L., Lee, S. F., Hoffman, A., Jones, J., Mitchell, A., Barcelona, D., Camara, B., Golmar, C., Savioli, L., Sembene, M., Takeuchi, T. and Wright, C. 2006. School-based health and nutrition programs. *In:* Jamison, D. T., Breman, J. G., Measham, A. R., Alleyne, G., Claeson, M., Evans, D. B., Jha, P., Mills, A.and Musgrove, P. (eds.) *Disease control priorities in developing countries.* Washington D.C: World Bank.

Cameron, L., 2009. Can a public scholarship program successfully reduce school drop-outs in a time of economic crisis? Evidence from Indonesia. *Economics of Education review,* 28**,** 308-317.

Carnoy, M., 2004. Education for all and the quality of education: A reanalysis. Paris: UNESCO. Background paper for EFA Global Monitoring Report 2005.

Carr, S., Leggart-Cook, C., Clarke, M., Maclachlan, Papola, T. S., Pais, J., Thomas, S., Mcauliffe, E. and Normand, C., 2001. What is the evidence of the impact of increasing salaries on improving the performance of public servants, including teachers, nurses and mid-level occupations, in low- and middle income countries: Is it time to give pay a chance? London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Chandler, A. M. K., Walker, S. P., Connolly, K. and Grantham-Mcgregor, S. M., 1995. School breakfast improves verbal fluency in undernourished Jamaican children. *Journal of Nutrition,* 125**,** 894-900.

Chang, S. M., Walker, S. P., Himes, J. and Grantham-Mcgregor, S. M., 1996. The effects of breakfast on classroom behaviours in rural Jamaican school children. *Food and Nutrition Bulletin,* 17**,** 248–57.

Chapman, D. W. and Mushlin, S., 2008. Do girls' scholarship programs work? Evidence from two countries. *International Journal of Educational Development,* 28**,** 460-472.

Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K. and Rogers, F. H., 2006. Missing in action: Teacher and health worker absence in developing countries. *Journal of Economic Perspectives,* 20**,** 91-116.

Colclough, C., Rose, P. and Tembon, M., 2000. Gender inequalities in primary schooling: The roles of poverty and adverse cultural practice. *International Journal of Educational Development,* 20**,** 5-27.

Cook, T. D., Shadish, W. R. and Wong, V. C., 2008. The conditions under which experimental & observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management,* 27**,** 724-750.

Critical Appraisal Skills Programme, 2006. 10 questions to help you make sense of qualitative research. Public Health Resource Unit. Available at: http://www.biomedcentral.com/content/supplementary/2046-4053-3-139-S8.pdf.

Crossley, M. and Murby, M., 1994. Textbook provision and the quality of the school curriculum in developing countries: Issues and policy options. *Comparative Education,* 30**,** 99-114.

Das, J., Do, Q. and Özler, B., 2005. Reassessing conditional cash transfer programs. *The World Bank Research Observer,* 20**,** 57-80.

Davies, P. 2006. What is needed from research synthesis from a policy-making perspective? *In:* Popay, J. (ed.) *Moving beyond effectiveness in evidence synthesis – methodological issues in the synthesis of diverse sources of evidence.* London: National Institute for Health and Clinical Excellence.

De Fraja, G. and Landeras, P., 2006. Could do better: The effectiveness of incentives and competition in schools. *Journal of Public Economics,* 90**,** 189-213.

De La Rue, L., Polanin, J. R., Espelage, D. L. and Pigott, T. D., 2013. Protocol: School-based interventions to reduce dating and sexual violence: A systematic review. Oslo: The Campbell Collaboration.

De Vibe, M., Bjørndal, A., Tipton, E., Hammerstrøm, K. T. and Kowalski, K., 2012. Mindfulness based stress reduction (MBSR) for improving health, quality of life and social functioning in adults. Campbell Systematic Reviews.

Department for International Development, 2013. Education position paper: Improving learning, expanding opportunities. London: Department for International Development.

Duflo, E., 2000. Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. Cambridge, MA: National Bureau of Economic Research.

Duflo, E. and Hanna, R., 2005. Monitoring works: Getting teachers to come to school. Cambridge MA: National Bureau of Economic Research.

Duval, S. and Tweedie, R., 2000. A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association,* 95**,** 89-98.

Effective Practice and Organisation of Care (EPOC), 2013. Epoc resources for review authors. Oslo: Norwegian Knowledge Centre for the Health Services. Available at: http://epoc.cochrane.org/epoc-specific-resources-review-authors.

Erisen, E., 2008. *School vouchers: Choice, competition and accountability: The Chilean case.* Stony Brook University.

Farrell, J. P. and Heyneman, S. P., 1989. Textbooks in the developing world: Economic and educational choices. Washington, D.C.: Economic Development Institute of the World Bank.

Filmer, D. and Schady, N., 2008. Getting girls into school: Evidence from a scholarship program in Cambodia. *Economic Development and Cultural Change,* 56**,** 581-618.

Friedman, W., Kremer, M., Miguel, E. and Thornton, R., 2011. Education as liberation? Cambridge, MA: National Bureau of Economic Research.

Fu, R., Gartlehner, G., Grant, M., Shamliyan, T., Sedrakyan, A., Wilt, T. J., Griffith, L., Oremus, M., Raina, P., Ismaila, A., Santaguida, P., Lau, J. and Trikalinos, T. A., 2011. Conducting qualitative synthesis when comparing medical interventions: Ahrq and the effective health care program. *Journal of Clinical Epidemiology,* 64**,** 1187-1197.

Glewwe, P., 2013. What education policies work best to increase student learning? Lessons from three recent reviews on the evidence. Health and Education Advice and Resource Team (HEART). Prepared for DFID Education Advisers 16th January 2013.

Glewwe, P. and Kremer, M., 2006. Schools, teachers, and education outcomes in developing countries. *Handbook of the Economics of Education,* 2**,** 945-1017.

Glewwe, P. W., Hanushek, E. A., Humpage, S. D. and Ravina, R., 2011. School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010. Cambridge, MA: National Bureau of Economic Research.

Greenhalgh, T., Kristjansson, E. and Robinson, V., 2007. Realist review to understand the efficacy of school feeding programmes. *BMJ,* 335**,** 858-861.

Guerrero, G., Leon, J., Zapata, M., Sugimaru, C. and Cueto, S., 2012. What works to improve teacher attendance in developing countries? A systematic review. London: EPPI Centre, Social Science Research Unit, Institute of Education, University of London.

Hammerstrøm, K., Wade, A. and Klint Jørgensen, A. M., 2010. Searching for studies: A guide to information retrieval for campbell systematic reviews. Oslo: Campbell Collaboration.

Hannum, E. and Buchmann, C., 2003. The consequences of global educational expansion. Social science perspectives. Cambridge, MA: American Academy of Arts and Sciences.

Hedges, L. V. 2011. Effect sizes in nested designs. *In:* Cooper, H., Hedges, L. V. and Valentine, J. C. (eds.) *The handbook of research synthesis.* New York: Russell Sage Foundation.

Hedges, L. V., Tipton, E. and Johnson, M. C., 2010. Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods,* 1**,** 39-65.

Herbert, R. D. and Bø, K., 2005. Analysis of quality of interventions in systematic reviews. *British Medical Journal,* 331**,** 507–509.

Higgins, J. and Green, S. 2011. *Cochrane handbook for systematic reviews of interventions. (version 5.0.2, updated September 2009)*, The Cochrane Collaboration.

Hillman, A. L. and Jenkner, E., 2004. Educating children in poor countries. *Economic Issues,* 33.

Holmstrom, B. and Milgrom, P., 1991. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization,* 7**,** 24-52.

Hombrados, G. J. and Waddington, H., 2012. Internal validity in social experiments and quasi-experiments: An assessment tool for reviewers. Unpublished working document. International Initiative for Impact Evaluation (3ie)

Hunt, F., 2008. Dropping out from school: A cross country review of literature. Consortium for Research on Educational Access, Transitions and Equity (CREATE), Falmer, UK.

International Development Coordinating Group, 2012. Protocol and review guidelines. Oslo: Campbell Collaboration.

Jomaa, L. H., Mcdonnell, E. and Probart, C., 2011. School feeding programs in developing countries: Impacts on children's health and educational outcomes. *Nutrition Reviews,* 69**,** 83–98.

Jukes, M., Drake, L. and Bundym, D., 2008. Levelling the playing field: School health, nutrition and education for all. Washington, D.C.: The World Bank.

Kabeer, N., Piza, C. and Taylor, L., 2012. What are the economic impacts of conditional cash transfer programmes? A systematic review of the evidence. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Kandel, E. and Lazear, E., 1992. Peer pressure and partnerships. *Journal of Political Economy,* 100**,** 801-817.

Kandori, M., 1992. Social norms and community enforcement. *The Review of Economic Studies,* 59**,** 63-80.

Keef, S. P. and Roberts, L. A., 2004. The meta-analysis of partial effect sizes. *British Journal of Mathematical and Statistical Psychology,* 57**,** 97-129.

Kingdon, G., Aslam, M., Rawal, S. and Das, S., 2012. Are contract and para-teachers a cost effective intervention to address teacher shortages and improve learning outcomes? Protocol. London: EPPI-Centre, Social Science Research Unit, Institute of Education.

Kingdon, K., Aslam, M., Rawal, S. and Das, S., 2013. Are contract teachers and para-teachers a cost-effective intervention to address teacher shortage and improve learning outcomes? London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Kremer, M., Brannen, C. and Glennester, R., 2013. The challenge of education and learning in the developing world. *Science,* 340**,** 297-300.

Kremer, M., Edward, M. and Thornton, R., 2009. Incentives to learn. *The Review of Economics and Statistics,* 91**,** 437-456.

Krishnaratne, S. and White, H., 2013. Quality education for all children. What works in education in developing countries. Delhi: International Initiative for Impact Evaluation (3ie).

Kristjansson, B., Petticrew, M., Macdonald, B., Krasevec, J., Janzen, L., Greenhalgh, T., Wells, G. A., Macgowan, J., Farmer, A. P., Shea, B., Mayhew, A., Tugwell, P. and Welch, V., 2009. School feeding for improving the physical and psychosocial health of disadvantaged students (review). The Cochrane Library.

Lavy, V., 2007. Using performance-based pay to improve the quality of teachers. *The Future of Children,* 17**,** 87-109.

Lawson, T. M., 2012. *Impact of school feeding programs on educational, nutritional, and agricultural development goals: A systematic review of literature.*

Le Grand, J. 2006. *Motivation, agency, and public policy: Of knights and knaves, pawns and queens,* Oxford, Oxford University Press.

Learning Metrics Task Force, 2013. Toward universal learning a global framework for measuring learning report. Paris: United Nations Educational, Scientific and Cultural Organisation (UNESCO).

Levacic, R., 2009. Teacher incentives and performance: An application of principal-agent theory. *Oxford Development Studies,* 37**,** 33-46.

Ma, X., Sylvia, S., Boswell, M. and Rozelle, S., 2013. Ordeal mechanisms and information in the promotion of health goods in developing countries: Evidence from rural china. Stanford: Rural Education Action Program (REAP).

Macdonald, G., Higgins, J. P. T., Ramchandani, P., Valentine, J. C., Bronger, L. P., Klein, P., O'daniel, R., Pickering, M., Rademaker, B., Richardson, G. and Taylor, M., 2012. Cognitive-behavioural interventions for children who have been sexually abused: A systematic review. Campbell Systematic Reviews.

Mañelele, I. and Muya, M., 2008. Risk identification on community-based construction projects in Zambia. *Journal of Engineering, Design and Technology,* 6**,** 145-161.

Mankiw, G., Romer, D. and Weil, D., 1992. A contribution to the empirics of economic growth. *The Quarterly Journal of Economics***,** 407-437.

Manley, J., Gitter, S. and Slavchevska, V., 2013. How effective are cash transfers at improving nutritional status? *World Development,* 48**,** 133-155.

Mansuri, G. and Rao, V., 2004. Community-based and-driven development: A critical review. *The World Bank Research Observer,* 19**,** 19-39.

Mares, M. and Pan, Z., 2013. Effects of Sesame Street: A meta-analysis of children's learning in 15 countries. *Journal of Applied Developmental Psychology,* 34**,** 140–151.

Mathews, R., 1996. Importance of breakfast to cognitive performance and health. *Perspectives in Applied Nutrition,* 3**,** 204-12.

Mcewan, P., 2014. Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments. *Review of Educational Research,* 85**,** 353-394.

Mcewan, P. J., 2012. Cost-effectiveness analysis of education and health interventions in developing countries. *Journal of Development Effectiveness,* 4**,** 189-213.

Miguel, M. M. and Barsaga, E. B. 1997. Multi-grade schooling in the philippines, a strategy for improving access to and quality of primary education. *In:* Chapman, D. W., Mahlck, L. O.andSmulders, A. E. M. (eds.) *From planning to action: Government initiatives for improving school-level practice.* Oxford / Paris: Pergamon / UNESCO.

Miles, M. B. and Huberman, A. M. 1994. *An expanded sources book: Qualitative data analysis,* London, Sage Publications.

Molina, E., Pacheco, A., Gasparini, L., Cruces, G. and Rius, A., 2013. Community monitoring interventions to curb corruption and increase access and quality in service delivery in low -and middle-income countries: A systematic review protocol. Campbell Collaboration Library.

Montgomery, P., Underhill, K., Gardner, F., Operario, D. and Mayo-Wilson, E., 2013. The oxford implementation index: A new tool for incorporating implementation data into systematic review and meta-analyses. *Journal of Clinical Epidemiology,* 66**,** 874-82.

Morgan, C., Petrosino, A. and Fronius, T., 2012. A systematic review of the evidence of the impact of eliminating school user fees in low-income developing countries.

Morgan, C., Petrosino, A. and Fronius, T., 2013. A systematic review of the evidence of the impact of school voucher programmes in developing countries.

Mourshed, M., Chijioke, C. and Barber, M., 2011. How the world's most improved school systems keep getting better. McKinsey & Company.

National Education Commission On Time and Learning, 2000. *Prisoners of time: Too much to teach, not enough time to teach it,* Peterborough, Crystal Springs Books.

National Institute for Health and Care Excellence. 2013. *Guide to the methods of technology appraisal* [Online]. NICE. Available online: http://publications.nice.org.uk/pmg9. Available: http://publications.nice.org.uk/pmg9.

Noyes, J., Booth, A., Hannes, K., Harden, A., Harris, J., Lewin, S. and Lockwood, C. E., 2011. Supplementary guidance for inclusion of qualitative research in Cochrane systematic reviews of interventions. Cochrane Collaboration Qualitative Methods Group. Available from: http://cqrmg.cochrane.org/supplemental-handbook-guidance.

Oketch, M., Mutisya, M., Ngware, M. and Ezeh, A. C., 2010. Why are there proportionately more poor pupils enrolled in non-state schools in urban Kenya in spite of fpe policy? *International Journal of Educational Development,* 30**,** 23-32.

Oketch, M. and Ngware, M., 2010. Free primary education still excludes the poorest of the poor in urban Kenya. *Development in Practice,* 20**,** 603-610.

Orr, D., Westbrook, J., Pryor, J., Durrani, N., Sebba, J. and Adu-Yeboah, C., 2013. What are the impacts and cost-effectiveness of strategies to improve performance of untrained and under-trained teachers in the classroom in developing countries? London: EPPI Centre, Social Science Research Centre, Institute of Education, University of London.

Petal, M., Green, R., Kelman I Shaw, R. and Pixit, A. 2008. Community-based construction for disaster risk reduction (chapter 10). *In:* Bosher, L. E. (ed.) *Hazards and the built environment: Attaining built-in resilience.* Abingdon: Taylor and Francis.

Petrosino, A., Morgan, C., Fronius, T. A., Tanner-Smith, E. E. and Boruch, R. F., 2012. Interventions in developing nations for improving primary and secondary school enrolment of children: A systematic review. Campbell Systematic Reviews.

Pollitt, E., Jacoby, E. and Cueto, S. 1998. Evaluation of a school breakfast program in Peru. *In:* Pan American Health, O. (ed.) *Nutrition, health, and child development.* USA: PAHO.

Pritchett, L., 2013. The rebirth of education: Schooling ain't learning. Washington D.C.: Center for Global Development.

Pritchett, L. and Sandfur, J., 2013. Context matters for size: Why external validity. Claims and development practice don't mix. Washington, DC: Center for Global Development.

Psacharopoulos, G., 1985. Returns to education: A further international update and implications. *The Journal of Human Resources,* 20**,** 583-604.

Psacharopoulos, G. and Patrinos, H. A., 2004. Returns to investment in education: A further update. *Education Economics,* 12**,** 111-134.

R Development Core Team, 2008. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

Reimers, F., Da Silva, C. D. and Trevino, E., 2006. Where is the "education" in conditional cash transfers in education? Montreal: UNESCO Institute for Statistics.

Reinikka, R. and Svensson, J., 2004. Local capture: Evidence from a government transfer program in Uganda. *The Quarterly Journal of Education,* 119**,** 679-705.

Robinson, J. P., 2011. A global compact on learning: Taking action on education in developing countries. Washington D.C.: Center for Universal Education at Brookings.

Roen, K., 2006. Extending systematic reviews to include evidence on implementation. *Social science and medicine,* 63**,** 1060–1071.

Rubenstein, L. V., Williams, J., Danz, M. and Shekelle, P., 2009. Determining key features of effective depression interventions. Los Angeles, CA.: Greater Los Angeles Veterans Affairs Healthcare System/Southern California/RAND Evidence-based Practice Center.

Saavedra, J. E. and García, S., 2012. Impacts of conditional cash transfer programs on educational outcomes in developing countries: A meta-analysis. Santa Monica: RAND Labor and Population working paper.

Sandieson, R., 2006. Pathfinding in the research forest: The pearl harvesting method for effective information retrieval. *Education and Training in Developmental Disabilities,* 41**,** 401-409.

Scriven, M., 2007. Key evaluation checklist. Unpublished manuscript. Available from: http://www.wmich.edu/evalctr/archive_checklists/kec_feb07.pdf.

Sen, A. 1999. *Development as freedom,* Oxford, Oxford University Press.

Shadish, W. and Myers, D., 2004. Research design policy brief. Oslo: Campbell Collaboration.

Shadish, W. R., 2011. Randomized controlled studies & alternative designs in outcome studies: Challenges and opportunities. *Research on Social Work Practice,* 21**,** 636-643.

Shemilt, I., Mugford, M., Byford, S., Drummond, M., Eisenstein, E., Knap, M. and Walker, D., 2008. The Campbell collaboration economics methods policy brief. Oslo: Campbell Collaboration.

Shemilt, I., Vale, L., Ruiz, F., Drummond, M. F., Brunetti, M., Lord, J., Graybill, E., Herramm, K. H. and Schünemann, H. J. *Grade guidelines: 17. Evidence to recommendations – considering resource use and modelling evidence for cost-effectiveness*.

Shemilt, I., Valentine, J. C., Pössel, P., Mugford, M. and Wooldridge, D. T., 2012. Costing program implementation using systematic reviews: Interventions for the prevention of adolescent depression. *Research synthesis methods,* 3**,** 191-201.

Sherman, L. W., Farrington, D. P., Welsh, B. C. and Layton Mackenzie, D. 2002. *Evidence based crime prevention (revised edition),* London, Routledge.

Slavin, R. E. 2009. *Can financial incentives enhance educational outcomes? Evidence from international experiments,* York, Institute for Effective Education.

Snilstveit, B., 2012. Systematic reviews: From 'bare bones' reviews to policy relevance. *Journal of Development Effectiveness,* 4**,** 388–40.

Song, F., Altman, D. G., Glenny, A. and Deeks, J. J., 2003. Validity of indirect comparison for estimating efficacy of competing interventions: Empirical evidence from published meta-analysis. *British Medical Journal,* 326**,** 472-475.

Spaull, N., 2011. Primary school performance in Botswana, Namibia, and South Africa. SACMEQ Working Paper. Retrieved 2014, from Southern & Eastern Africa Consortium for Monitoring and Educational Quality.

Steiner, P. M., Cook, T, D., Shadish, W. R. and Clark, M. H., 2010. The importance of covariate selection in controlling for selection bias in observational studies. *Psychological Methods,* 15**,** 250–267.

Sylva, K., Melhuish, E., Sammons, P. and Siraj-Blatchford, I., 2004. Brenda Taggart the effective provision of pre-school education (eppe) project: Final report: A longitudinal study funded by the DfES 1997-2004. London: Institute of Education.

Tanner-Smith, E. and Tipton, E., 2014. Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in stata and SPSS. *Research Synthesis Methods,* 5**,** 13-30.

Taylor-Robinson, D. C., Maayan, N., Soares-Weiser, K., Donegan, S. and Garner, P., 2012. Deworming drugs for soil-transmitted intestinal worms in children: Effects on nutritional indicators, haemoglobin and school performance. Cochrane Database of Systematic Reviews.

Tikly, L., 2011. Towards a framework for researching the quality of education in low-income countries. *Comparative Education,* 47**,** 1-23.

Tomasevski, K., 2003. School fees as a hinderance to universalizing primary education. Paris: UNESCO.

Tooley, J., Dixon, P. and Stanfield, J., 2008. Impact of free primary education in Kenya: A case study of private schools in kibera. *Educational Management Administration & Leadership,* 36**,** 449-469.

Treasury, H. M., 2011. Magenta book. London: HM Treasury.

Umansky, I. 2005. A literature review of teacher quality and incentives. *In:* Ed, E. V. (ed.) *Incentives to improve teaching.* Washington D.C.: The World Bank.

UN, 2013. *The universal declaration of human rights* [Online]. United Nations. Available at: http://www.un.org/en/documents/udhr/index.shtml#a26.

UN, 2015. Millennium Development Goals Report [online]. Available at: http://www.un.org/millenniumgoals/2015_MDG_Report/pdf/MDG%202015%20rev%20(July%201).pdf [Accessed on the 18th of November 2015]

UNESCO Institute for Statistics, 2005. Children out of school: Measuring exclusion from primary education. Montreal: UNESCO.

UNESCO, 2012. Reaching out-of-school children is crucial for development. Paris: UNESCO. UIS Fact Sheet, No.18.

UNESCO, 2014a. Region Profile (World) UNESCO Institute for Statistics Data Centre [online]. Available at: http://www.uis.unesco.org/DataCentre/Pages/region-profile.aspx?regioncode=40510 [Accessed on the 18th of November 2015].

UNESCO, 2014b. Region Profile (Sub-Saharan Africa) UNESCO Institute for Statistics Data Centre [online]. Available at: http://www.uis.unesco.org/DataCentre/Pages/region-profile.aspx?regioncode=40540 [Accessed on the 18th of November 2015].

UNESCO, 2015. EFA Global Monitoring Report [online]. Available at: http://unesdoc.unesco.org/images/0023/002322/232205e.pdf [Accessed on the 18th of November 2015].

UNICEF, 2013b. *At a glance: Burkina Faso [online]* [Online]. Available: http://www.unicef.org/infobycountry/burkinafaso_statistics.html. [Accessed on the 15th of December, 2015].

UNICEF. 2013a. *At a glance: Mexico [online]* [Online]. Available: http://www.unicef.org/infobycountry/mexico_statistics.html. [Accessed on the 15th of December, 2015].

USAID, 2011. School dropout prevention pilot program: Review of the literature. Washington D.C.: USAID.

Uwezo, 2013. Are our children learning? Literacy and numeracy across east Africa. Nairobi: UWEZO.

Waddington, H., Snilstveit, B., Hombrados, J., Vojtkova, M., Phillips, D. and White, H., 2014. Farmer field schools for improving farming practices and farmer outcomes in low- and middle-income countries: A systematic review. *Campbell Systematic Reviews,* 10.

Westhorp, G., Walker, B. and Rogers, P., 2013. Under what circumstances does enhancing community accountability and empowerment improve education outcomes, particularly for the poor? A realist synthesis protocol. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Westhorp, G., Walker, D. W., Rogers, P., Overbeeke, N., Ball, D. and Brice, G., 2014. Enhancing community accountability, empowerment and education outcomes in low and middle-income countries: A realist review. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

White, H., 2009. Theory-based impact evaluation: Principles and practice. New Delhi: International Initiative for Impact Evaluation.

White, H., Krishnaratne, S. and Carpenter, E., 2013. Quality education for all children? What works in education in developing countries. New Delhi: International Initiative for Impact Evaluation (3ie).

Wild, L. and Harris, D., 2011. The political economy of community scorecards in Malawi. London: Overseas Development Institute.

Wilson, D. B., Weisburd, D. and Mclure, D., 2011. Use of DNA testing in police investigative work for increasing offender identification, arrest, conviction and case clearance. *Campbell Systematic Reviews,* 7.

World, B., 2004. World development report 2004: Making services work for poor people. Washington, D.C.: World Bank.

Yoong, J., Rabinovich, L. and Diepeveen, S., 2012. The impact of economic resource transfers to women versus men: A systematic review. London: London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Appendix A: Search strategy

Search Terms

LMICs

1. (Afghanistan or Albania or Algeria or Angola or Antigua or Barbuda or Argentina or Armenia or Armenian or Aruba or Azerbaijan or Bahrain or Bangladesh or Barbados or Benin or Byelarus or Byelorussian or Belarus or Belorussian or Belorussia or Belize or Bhutan or Bolivia or Bosnia or Herzegovina or Hercegovina or Botswana or Brasil or Brazil or Bulgaria or "Burkina Faso" or "Burkina Fasso" or "Upper Volta" or Burundi or Urundi or Cambodia or "Khmer Republic" or Kampuchea or Cameroon or Cameroons or Cameron or Camerons or "Cape Verde" or "Central African Republic" or Chad or Chile or China or Colombia or Comoros or "Comoro Islands" or Comores or Mayotte or Congo or Zaire or "Costa Rica*" or "Cote d'Ivoire" or "Ivory Coast" or Croatia or Cuba or Czechoslovakia or "Czech Republic" or Slovakia or "Slovak Republic" or Djibouti or "French Somaliland" or Dominica or "Dominican Republic" or "East Timor" or "East Timur" or "Timor Leste" or Ecuador or Egypt or "United Arab Republic" or "El Salvador" or Eritrea or Estonia or Ethiopia or Fiji or Gabon or "Gabonese Republic" or Gambia or Gaza or "Georgia Republic" or "Georgian Republic" or Ghana or "Gold Coast" or Greece or Grenada or Guatemala or Guinea or Guam or Guiana or Guyana or Haiti or Honduras or India or Maldives or Indonesia or Iran or Iraq or Jamaica or Jordan or Kazakhstan or Kazakh or Kenya or Kiribati or Korea or Kosovo or Kyrgyzstan or Kirghizia or "Kyrgyz Republic" or Kirghiz or Kirgizstan or "Lao PDR" or Laos or Latvia or Lebanon or Lesotho or Basutoland or Liberia or Libya or Lithuania or Macedonia or Madagascar or "Malagasy Republic" or Malaysia or Malaya or Malay or Sabah or Sarawak or Malawi or Nyasaland or Mali or Malta or "Marshall Islands" or Mauritania or Mauritius or "Agalega Islands" or Mexico or Micronesia or "Middle East" or Moldova or Moldovia or Moldovian or Mongolia or Montenegro or Morocco or Ifni or Mozambique or Myanmar or Myanma or Burma or Namibia or Nepal or "Netherlands Antilles" or "New Caledonia" or Nicaragua or Niger or Nigeria or "Northern Mariana Islands" or Oman or Muscat or Pakistan or Palau or Palestine or Panama or Paraguay or Peru or Philippines or Philipines or Phillipines or Phillippines or "Puerto Ric*" or Romania or Rumania or Roumania or Russia or Russian or Rwanda or Ruanda or "Saint Kitts" or "St Kitts" or "Nevis" or "Saint Lucia" or "St Lucia" or "Saint Vincent" or "St Vincent" or Grenadines or Samoa or "Samoan Islands" or "Navigator Island" or "Navigator Islands" or "Sao Tome" or "Saudi Arabia" or Senegal or Serbia or Montenegro or Seychelles or "Sierra Leone" or Slovenia or "Sri Lanka" or Ceylon or "Solomon Islands" or Somalia or "South Africa" or Sudan or Suriname or Surinam or Swaziland or Syria or Tajikistan or Tadzhikistan or Tadjikistan or Tadzhik or Tanzania or Thailand or Togo or Togolese Republic or Tonga or Trinidad or Tobago or Tunisia or Turkey or Turkmenistan or Turkmen or Uganda or Ukraine or Uruguay or USSR or "Soviet Union" or "Union of Soviet Socialist Republics" or Uzbekistan or Uzbek or Vanuatu or "New Hebrides" or Venezuela or Vietnam or "Viet Nam" or "West Bank" or Yemen or Yugoslavia or Zambia or Zimbabwe or Rhodesia) NOT ("African-American*" OR "African-American*" OR "Mexican American*" OR "American Indian*" OR "Asian American*" OR "native american*")

2. (developing or "less* developed" or "under developed" or underdeveloped or under-developed or "middle income" or "low* income") NEAR/3 (countr* or nation*)

3. (low NEAR/3 (middle NEAR/3 (countr*))

4. (Africa or Asia or Caribbean or “West Indies” or “South America” or “Latin America” or “Central America”)

5. (lmic or lmics or "third world" or "lami countr*" OR "transitional countr*")

6. 1 OR 2 OR 3 OR 4 OR 5

General Search

Young Students (Population)

1. (student* OR pupil* OR child* OR youth* OR youngster* OR “young person*” OR “young people” OR teen* OR adolescen* OR schoolchild*)

Study Methods

2. ("random* control* trial*" OR "random* trial*" OR RCT OR "cluster random* trial" OR "propensity score matching" OR PSM Or "regression discontinuity design" OR RDD OR "difference in difference*" OR DID OR "systematic* review*" OR meta-analy* OR "meta analy*" OR SR OR "control* random* trial*" OR "case control" OR matching OR "interrupted time series" OR "random* allocation*" OR (random* NEAR/3 (allocat*)) OR "instrumental variable*" OR IV OR "research synthesis" OR "scoping review" OR "rapid evidence assessment" OR "systematic literature review" OR evaluation OR assessment OR ((quantitative OR "comparison group" OR counterfactual OR "counter factual" OR counter-factual OR experiment*) NEAR/3 (design OR study OR analysis)) OR QED)

Outcomes

3. (outcome* OR effect* OR impact* OR attain* OR enrol* OR attend* OR progress* OR achiev* OR result OR results OR complet* OR improve* OR assess* OR perform* OR test* OR mark OR marks OR marking OR learn* OR exam OR exams OR examination* OR graduat* OR matriculat* OR retention OR retain* OR grade* OR grading OR score* OR scoring OR absen* OR truan* OR "drop out*" OR "drop-out*" OR "dropped out" OR qualif* OR cost* OR "cost-effect*" OR "cost-benefit" OR "cost-utility")

Education

4. (educat* OR teach* OR academ* OR schol* OR school* OR class room OR classes OR classroom* OR class-room* OR pedagog* OR learn* OR lesson* OR curricul*)

5. 1 AND 2 AND 3 AND 4

Reducing Costs

1. ("cash transfer*" OR "cash-transfer*" OR (cash NEAR/3 (transfer*)) OR (cash NEAR/3 (payment*)) OR pension OR pensions OR (cash NEAR/3 (incentive*)) OR CCT* OR UCT* OR ((cash OR asset* OR monetary OR economic OR pecuniary OR capital) NEAR/3 (pay* OR transfer* OR incentiv* OR hand-out* OR handout* OR grant* OR aid OR assistance OR benefit* OR help)) OR ("child support" NEAR/3 grant*) or (cash NEAR/3 subsid*) OR "social safety" or "welfare grant*" or "social protection" or "transfer payment*" or "transfer program*" or "poverty alleviation transfer*" OR Oportunidades OR PROGRESA OR "Bolsa familia" OR "Bolsa escola" OR "familias en accion" OR "escuela nueva")

2. (scholarship* OR subvention* OR subsid* OR stipend* OR grant* OR donation OR bursary OR bursaries OR “tuition relief” OR “user payment*” OR “merit aid” OR “merit based aid” OR “merit-based aid” OR “merit award”)

3. ((Uniform OR uniforms) NEAR/3 school) OR (User NEAR/3 (payment* OR fee* OR finance*)) OR (education NEAR/3 (charg* OR payment*)))

4. ((Voucher* OR credit*) NEAR/3 (national OR program* OR plan* OR education* OR school* OR choice)) OR scholarship* OR “equal education” OR “private school aid” OR subsid*)

5. (((Fee* OR tuition) NEAR/3 (reduc* OR abolish* OR abolition* OR stop* OR eliminat* OR cancel* OR cut OR waiv*)) OR “tuition tax credit” OR scholarship OR “fee free” OR “fee-free” OR “non-fee paying”)

6. 1 OR 2 OR 3 OR 4 OR 5

Providing Information

((mentor* OR peer OR volunteer* OR "role model*" OR "role-model" OR "scholarship plus" OR "study counsel*" OR "directive counsel*" OR feedback) NEAR/3 (school* OR educat*)) OR (((provis* OR dissem* OR invest*) NEAR/3 (inform* OR stat*)) OR ((provide OR providing) AND information)) OR ((perceive* OR perception* OR expect* OR estimat*) NEAR/3 (return* OR benefit*) NEAR/3 (educat* OR school*))

Report Card

(“report card*” OR scorecard OR score-card OR “score card” OR “assessment systems” OR “student assessment” OR “school-based information” OR “school based information” OR “school quality information” OR “information for accountability” OR “information campaign*” OR (school AND (monitoring OR inspection*))) OR ((“active citizenship” OR ranking OR “school accountability” OR “social accountability” OR “beneficiary accountability” OR “rights-based accountability” OR “community accountability” OR overs* OR monitor* OR decentralis* OR decentraliz* OR transparen* OR “parent-teacher partnership*” OR “parent teacher partnership*” OR PTP OR audit) NEAR/3 (educat* OR school*)) OR ((Communit* OR civil OR citizen* OR local*) NEAR/3 (empower* OR accountab* OR transparen*) NEAR/3 (educat* OR school*))

Teacher-related Supply Side

((teacher* or schoolteacher* OR school-teacher* OR "school teacher*" OR tutor OR tutors OR educator) NEAR/3 (hire OR hiring OR hired OR recruit* OR supervis* OR monitor* OR attend* OR absen* OR truan* OR shirk* OR presen* OR drop-out* OR "drop out*" OR "dropped out" OR perform* OR employ* OR retention OR retain* OR accountab* OR report* OR learn* OR course* OR “professional development” OR training OR qualif* OR experience OR educat* OR bonus* OR reward OR rewards OR merit OR pay OR payment OR incentiv* OR remunerat* OR salary OR salaries OR wage OR wages OR emolument* OR earning* OR contract* OR work-load OR workload OR “work* environment*” OR “work* conditions” OR mentor*)) OR (((educat* OR teach* OR academ* OR schol* OR school* OR pedagog*) NEAR/3 (assistant* OR staff OR personnel OR temp*)) NEAR/3 (hire OR hiring OR hired OR recruit* OR supervis* OR monitor* OR attend* OR absen* OR truan* OR shirk* OR presen* OR drop-out* OR "drop out*" OR "dropped out" OR perform* OR employ* OR retention OR retain* OR accountab* OR report* OR learn* OR course* OR “professional development” OR training OR qualif* OR experience OR educat* OR bonus* OR reward OR rewards OR merit OR pay OR payment OR incentiv* OR remunerate* OR salary OR salaries OR wage OR wages OR emolument* OR earning* OR contract* OR work-load OR workload OR “work* environment*” OR “work* conditions” OR mentor*))

School-based Management

((educat* OR teach* OR academ* OR schol* OR school* OR pedagog*) NEAR/3 (“site-based management” OR “ site based management” OR accountabil* OR managed OR management OR managing OR administrating OR administration OR administrated OR organisation OR organization OR decentral* OR governance OR budget* OR expenditure OR allocate* OR autonomy OR “decision-making” OR “decision making”)) OR (((community OR parent*) NEAR/3 (association OR board* OR council* OR committee*)) NEAR/3 (educat* OR teach* OR academ* OR schol* OR school* OR pedagog*)) OR (((share* OR sharing) NEAR/3 decision*) NEAR/3 (educat* OR teach* OR academ* OR schol* OR school* OR pedagog*)) OR ((parent* NEAR/3 particip*) NEAR/3 (educat* OR teach* OR academ* OR schol* OR school* OR pedagog*)) OR (SBM OR "school-based management" OR "school-based-management" OR "school based management" OR “school-based budgeting” OR “school based budgeting” OR “collaborative school management” OR “shared school governance”)

Buildings and Infrastructure, Equipment and Materials

((educat* OR teach* OR academ* OR schol* OR school* OR pedagog*) NEAR/3 (electric* OR aid* OR equipment OR materials OR supplies OR stationery OR book* OR desk* OR chair* OR flipchart* OR flip-chart* OR “flip chart*” OR chalkboard OR whiteboard OR blackboard OR chalk-board OR white-board OR black-board OR “chalk board” OR “white board” OR “black board” OR computer* OR PC OR laptop OR internet OR tech*)) OR ((transport* OR bus*) NEAR/3 (school* OR educat* OR student* OR pupil*)) OR ((educat* OR teach* OR academ* OR schol* OR school*) NEAR/3 (input* OR upgrad* OR infrastructure OR building OR structure* OR facility OR facilities OR house OR houses OR housing OR residential OR residence* OR accommodation OR classroom* OR class-room* OR "class room*" OR toilet* OR latrine* OR WC OR lavator* OR washroom* OR "wash room*" OR pump* OR garden Or playground Or "play area" OR play-ground OR play-area OR "play ground" OR librar* OR lab OR labs OR laborator*))

Teaching Methods

((educat* OR curricul* OR pedagog* OR teach* OR instruct*) NEAR/3 (method* OR approach* OR improv* OR develop* OR reform* OR change*)) OR ((class* OR lesson*) NEAR/3 (plan* OR preparation OR preparing OR guide)) OR ((educat* OR teach* OR class* OR pedagog* OR stud* OR learn* OR instruct*) NEAR/3 (stream* OR multigrade OR multi-grade OR "multi grade" OR "multiple grade" OR group OR cooperative OR co-operative)) OR ((educat* OR teach* OR class* OR pedagog* OR stud* OR learn* OR instruct*) NEAR/3 (homework OR home-work OR "home work" OR tutoring OR remedia* OR developmental OR "basic skill*" OR compensatory OR supplement* OR additional OR after-school OR "after school")) OR (("computer assisted learning" OR computer-assisted-learning OR "computer-assisted learning" OR "computer based learning" OR computer-based-learning OR "computer-based learning" OR "computer game" OR "electronic game") NEAR/3 (educat* OR teach* OR pedagog* OR learn*)) OR (((educat* OR teach* OR class* OR pedagog* OR stud* OR class* OR learn* OR instruct*) NEAR/3 (computer* OR internet OR tech*)) OR CAL) OR ((“Class size” NEAR/3 (reduc* OR small*)) OR (class NEAR/3 size) OR ((student* OR pupil*) NEAR/3 number*) OR ((ratio NEAR/3 teacher*) NEAR/3 (student* OR pupil*)) OR “school size”) OR ((lesson* OR learn* OR educat* OR teach* OR class* OR pedagog* OR instruct* OR “school term”* OR school-term OR “school day*” OR “school week”) NEAR/3 (hours OR time* OR timing* OR length OR duration OR flexible)) OR ((educat* OR curricul* OR pedagog* OR teach* OR instruct*) NEAR/3 (language OR dialect))

Example full search strategy:

Web of Science (Social Sciences Citation Index/Arts & Humanities Citation Index) – Searched 19th Nov 2013 & 12th Dec 2013 (Health aspects)

#36 #35 NOT (#34 OR #32)

#35 #33 AND #22 AND #4 AND #3 AND #1

#34 TS=(universit* or "medical school*" or college or " higher education" or (medical or nursing or pharmacy or veterinary) NEXT/1 (student*))

#33 TS=(Ivermectin or Albendazole or Mebendazole or Piperazine* or Levamisole or pyrantel or tiabendazole or anthelmint* or Anticestodal or Antiplatyhelmintic or Anti-platyhelmintic or Albendazole or Dichlorophen or Niclosamide or Quinacrine or Bithionol or Diamfenetide or Nitroxinil or Oxyclozanide or Rafoxanide or Schistosomicide* or "Antimony Potassium Tartrate" or "Antimony Sodium Gluconate" or Hycanthone or Lucanthone or Niridazole or Oxamniquine) OR TS=(deworm* or de-worm* or whipworm* or "whip worm*" or hookworm* or "hook worm*" or roundworm* or "round worm*" or pinworm* or "pin worm*" or flukes or helmint* or geohelminth* or ancylostoma or Necator* or Ascaris or Ascaridida or Ancylostoma or "Necator americanus" or Enterobius or Oxyuroidea or Oxyurida or Trichuris or Trichuroidea or Capillaria or Trichinella or Strongyloid* or Oesophagostomum or Oesophagostomiasis or Strongylus or Acanthocephala or Moniliformis or Adenophorea or Enoplida or Secernentea or Ascaridida or Rhabditida or Nematoda or Cestoda or Trematod* or Turbellaria or Platyhelminth* or Rotifera or trichuriasis or ascariasis or trichinellosis or Trichostrongyloidiasis or ancylostomiasis or enterobiasis or nematode* or cestode* or trematode* or ascarid* or Toxocara* or toxocariasis or schistosomiasis or Schistosoma*) OR TS=(Food OR Diet OR “dietary Supplement*” OR “diet therapy” OR “diet fortif*” OR “Functional Food” OR Nutri* OR Supplement* OR “Food For Education” OR (in-school OR “in school” OR Extra OR take-home OR “take home” OR takehome NEAR/1 (food OR feed* OR ration* OR meal*)) OR Feed* OR Ration* OR Lunch* OR dinner* OR break-fast* OR breakfast* OR break fast* OR supper* OR snack* OR meal* OR Milk OR milk-powder OR Milk Powder OR Cereal* OR Flour OR Maize OR Porridge OR Biscuit* OR Vitameal OR (Fortif* OR Enrich* NEAR/1 (food OR diet OR spread OR flour OR cereal*))) OR TS=((supplement* OR complement* NEAR/1 (food OR feed OR diet OR nutrition OR nutrient* OR micronutrient* OR micro-nutrient*)) OR Vitamin* OR Mineral* OR iron OR “iron supplement*” OR iron fortific* OR (RUTF OR Therapeutic NEAR/1 (feed* OR food* OR Plumpy* OR Nutrispread OR LNS OR “Lipid Nutrient Supplement*”)) OR (supplement* NEAR/1 (“Lipid based” OR Lipid-based))) OR TS=(eyeglass* OR eye-glass OR glasses OR spectacles OR specs OR “vision correction” OR “vision screening” OR “eye test”OR “glasses-wearing” NEAR/3 (educat* OR school)) OR TS=(“intermittent screening and treatment” OR IST OR “intermittent preventive treatment” OR IPT OR school-based NEAR/3 (malaria)) OR TS=(anaemi* OR anemi* OR “iron deficiency” NEAR/3 (educat* OR school*))

OR TS=(health* or nutrition* or well-being or illness* or sickness* or sick or malnutrition* or malnourished or undernutrition)

#32 #31 OR #30 OR #29 OR #28 OR #27 OR #26 OR #25 OR #24

#31 #22 AND #4 AND #3 AND #2 AND #1

Refined by: Web of Science Categories=(EDUCATION EDUCATIONAL RESEARCH OR EDUCATION SPECIAL OR PSYCHOLOGY DEVELOPMENTAL OR ETHICS OR PSYCHOLOGY MULTIDISCIPLINARY OR MANAGEMENT OR ETHNIC STUDIES OR PSYCHOLOGY APPLIED OR ECONOMICS OR BEHAVIORAL SCIENCES OR EDUCATION SCIENTIFIC DISCIPLINES OR PSYCHOLOGY EDUCATIONAL OR SOCIAL SCIENCES INTERDISCIPLINARY OR SOCIAL ISSUES OR BUSINESS FINANCE OR FAMILY STUDIES OR PSYCHOLOGY SOCIAL OR SOCIOLOGY OR POLITICAL SCIENCE OR TRANSPORTATION OR URBAN STUDIES OR PLANNING DEVELOPMENT OR PUBLIC ADMINISTRATION)

#30 #22 AND #16 AND #4

#29 #22 AND #15 AND #4

#28 #22 AND #14 AND #4

#27 #22 AND #13 AND #4 AND #1

#26 #22 AND #12 AND #4 AND #1

#25 #22 AND #11 AND #4 AND #1

#24 #22 AND #10 AND #4 AND #1

#23 #22 AND #4 AND #3 AND #2 AND #1

#22 #21 OR #20 OR #19 OR #18 OR #17

#21 TS=((lmic or lmics or "third world" or "lami countr*")) OR TS=("transitional countr*")

#20 TS=(Africa or Asia or Caribbean or “West Indies” or “South America” or “Latin America” or “Central America”)

#19 TS=(low NEAR/3 (middle NEAR/3 (countr*)))

#18 TS=((developing or "less* developed" or "under developed" or underdeveloped or under-developed or "middle income" or "low* income") NEAR/3 (countr* or nation*))

#17 TS=((Afghanistan or Albania or Algeria or Angola or Argentina or Armenia or Armenian or Aruba or Azerbaijan or Bahrain or Bangladesh or Benin or Byelarus or Byelorussian or Belarus or Belorussian or Belorussia or Belize or Bhutan or Bolivia or Bosnia or Herzegovina or Hercegovina or Botswana or Brasil or Brazil or Bulgaria or "Burkina Faso" or "Burkina Fasso" or "Upper Volta" or Burundi or Urundi or Cambodia or "Khmer Republic" or Kampuchea or Cameroon or Cameroons or Cameron or Camerons or "Cape Verde" or "Central African Republic" or Chad or China or Colombia or Comoros or "Comoro Islands" or Comores or Mayotte or Congo or Zaire or "Costa Rica*" or "Cote d'Ivoire" or "Ivory Coast" or Cuba or Djibouti or "French Somaliland" or Dominica or "Dominican Republic" or "East Timor" or "East Timur" or "Timor Leste" or Ecuador or Egypt or "United Arab Republic" or "El Salvador" or Eritrea or Ethiopia or Fiji or Gabon or "Gabonese Republic" or Gambia or Gaza or "Georgia Republic" or "Georgian Republic" or Ghana or Grenada or Guatemala or Guinea or Guiana or Guyana or Haiti or Honduras or India or Maldives or Indonesia or Iran or Iraq or

Jamaica or Jordan or Kazakhstan or Kazakh or Kenya or Kiribati or Korea or Kosovo or Kyrgyzstan or Kirghizia or "Kyrgyz Republic" or Kirghiz or Kirgizstan or "Lao PDR" or Laos or Lebanon or Lesotho or Basutoland or Liberia or Libya or Macedonia or Madagascar or "Malagasy Republic" or Malaysia or Malaya or Malay or Sabah or Sarawak or Malawi or Mali or "Marshall Islands" or Mauritania or Mauritius or "Agalega Islands" or Mexico or Micronesia or "Middle East" or Moldova or Moldovia or Moldovian or Mongolia or Montenegro or Morocco or Ifni or Mozambique or Myanmar or Myanma or Burma or Namibia or Nepal or "Netherlands Antilles" or "New Caledonia" or Nicaragua or Niger or Nigeria or Muscat or Pakistan or Palau or Palestine or Panama or Paraguay or Peru or Philippines or Philipines or Phillipines or Phillippines or "Puerto Ric*" or Romania or Rumania or Roumania or Rwanda or Ruanda or "Saint Lucia" or "St Lucia" or "Saint Vincent" or "St Vincent" or Grenadines or Samoa or "Samoan Islands" or "Navigator Island" or "Navigator Islands" or "Sao Tome" or Senegal or Serbia or Montenegro or Seychelles or "Sierra Leone" or "Sri Lanka" or Ceylon or "Solomon Islands" or Somalia or "South Africa" or Sudan or Suriname or Surinam or Swaziland or Syria or Tajikistan or Tadzhikistan or Tadjikistan or Tadzhik or Tanzania or Thailand or Togo or Togolese Republic or Tonga or Tunisia or Turkey or Turkmenistan or Turkmen or Uganda or Ukraine or Uzbekistan or Uzbek or Vanuatu or "New Hebrides" or Venezuela or Vietnam or "Viet Nam" or "West Bank" or Yemen or Yugoslavia or Zambia or Zimbabwe or Rhodesia) NOT ("African-American*" OR "African-American*" OR "Mexican American*" OR "American Indian*" OR "Asian American*" OR "native american*"))

#16 TS=((educat* OR curricul* OR pedagog* OR teach* OR instruct*) NEAR/3 (method* OR approach* OR improv* OR develop* OR reform* OR change*)) OR TS=((class* OR lesson*) NEAR/3 (plan* OR preparation OR preparing OR guide)) OR TS=((educat* OR teach* OR class* OR pedagog* OR stud* OR learn* OR instruct*) NEAR/3 (stream* OR multigrade OR multi-grade OR "multi grade" OR "multiple grade" OR group OR cooperative OR co-operative)) OR TS=((educat* OR teach* OR class* OR pedagog* OR stud* OR learn* OR instruct*) NEAR/3 (homework OR home-work OR "home work" OR tutoring OR remedia* OR developmental OR "basic skill*" OR compensatory OR supplement* OR additional OR after-school OR "after school")) OR TS=(("computer assisted learning" OR computer-assisted-learning OR "computer-assisted learning" OR "computer based learning" OR computer-based-learning OR "computer-based learning" OR "computer game" OR "electronic game") NEAR/3 (educat* OR teach* OR pedagog* OR learn*)) OR TS=(((educat* OR teach* OR class* OR pedagog* OR stud* OR class* OR learn* OR instruct*) NEAR/3 (computer* OR internet OR tech*)) OR CAL) OR TS=((“Class size” NEAR/3 (reduc* OR small*)) OR (class NEAR/3 size) OR ((student* OR pupil*) NEAR/3 number*) OR ((ratio NEAR/3 teacher*) NEAR/3 (student* OR pupil*)) OR “school size”) OR TS=((lesson* OR learn* OR educat* OR teach* OR class* OR pedagog* OR instruct* OR “school term”* OR school-term OR “school day*” OR “school week”) NEAR/3 (hours OR time* OR timing* OR length OR duration OR flexible)) OR TS=((educat* OR curricul* OR pedagog* OR teach* OR instruct*) NEAR/3 (language OR dialect))

#15 TS=((educat* OR teach* OR academ* OR schol* OR school* OR pedagog*) NEAR/3 (electric* OR aid* OR equipment OR materials OR supplies OR stationery OR book* OR desk* OR chair* OR flipchart* OR flip-chart* OR “flip chart*” OR chalkboard OR whiteboard OR blackboard OR chalk-board OR white-board OR black-board OR “chalk board” OR “white board” OR “black board” OR computer* OR PC OR laptop OR internet OR tech*)) OR TS=((transport* OR bus*) NEAR/3 (school* OR educat* OR student* OR pupil*)) OR TS=((educat* OR teach* OR academ* OR schol* OR school*) NEAR/3 (input* OR upgrad* OR infrastructure OR building OR structure* OR facility OR facilities OR house OR houses OR housing OR residential OR residence* OR accommodation OR classroom* OR class-room* OR "class room*" OR toilet* OR latrine* OR WC OR lavator* OR washroom* OR "wash room*" OR pump* OR garden Or playground Or "play area" OR play-ground OR play-area OR "play ground" OR librar* OR lab OR labs OR laborator*))

#14 TS=((educat* OR teach* OR academ* OR schol* OR school* OR pedagog*) NEAR/3 (“site-based management” OR “ site based management” OR accountabil* OR managed OR management OR managing OR administrating OR administration OR administrated OR organisation OR organization OR decentral* OR governance OR budget* OR expenditure OR allocate* OR autonomy OR “decision-making” OR “decision making”)) OR TS=(((community OR parent*) NEAR/3 (association OR board* OR council* OR committee*)) NEAR/3 (educat* OR teach* OR academ* OR schol* OR school* OR pedagog*)) OR TS=(((share* OR sharing) NEAR/3 decision*) NEAR/3 (educat* OR teach* OR academ* OR schol* OR school* OR pedagog*)) OR TS=((parent* NEAR/3 particip*) NEAR/3 (educat* OR teach* OR academ* OR schol* OR school* OR pedagog*)) OR TS=(SBM OR "school-based management" OR "school-based-management" OR "school based management" OR “school-based budgeting” OR “school based budgeting” OR “collaborative school management” OR “shared school governance”)

#13 TS=((teacher* or schoolteacher* OR school-teacher* OR "school teacher*" OR tutor OR tutors OR educator) NEAR/3 (hire OR hiring OR hired OR recruit* OR supervis* OR monitor* OR attend* OR absen* OR truan* OR shirk* OR presen* OR drop-out* OR "drop out*" OR "dropped out" OR perform* OR employ* OR retention OR retain* OR accountab* OR report* OR learn* OR course* OR “professional development” OR training OR qualif* OR experience OR educat* OR bonus* OR reward OR rewards OR merit OR pay OR payment OR incentiv* OR remunerat* OR salary OR salaries OR wage OR wages OR emolument* OR earning* OR contract* OR work-load OR workload OR “work* environment*” OR “work* conditions” OR mentor*)) OR TS=(((educat* OR teach* OR academ* OR schol* OR school* OR pedagog*) NEAR/3 (assistant* OR staff OR personnel OR temp*)) NEAR/3 (hire OR hiring OR hired OR recruit* OR supervis* OR monitor* OR attend* OR absen* OR truan* OR shirk* OR presen* OR drop-out* OR "drop out*" OR "dropped out" OR perform* OR employ* OR retention OR retain* OR accountab* OR report* OR learn* OR course* OR “professional development” OR training OR qualif* OR experience OR educat* OR bonus* OR reward OR rewards OR merit OR pay OR payment OR incentiv* OR remunerate* OR salary OR salaries OR wage OR wages OR emolument* OR earning* OR contract* OR work-load OR workload OR “work* environment*” OR “work* conditions” OR mentor*))

#12 TS=(“report card*” OR scorecard OR score-card OR “score card” OR “assessment systems” OR “student assessment” OR “school-based information” OR “school based information” OR “school quality information” OR “information for accountability” OR “information campaign*” OR (school AND (monitoring OR inspection*))) OR TS=((“active citizenship” OR ranking OR “school accountability” OR “social accountability” OR “beneficiary accountability” OR “rights-based accountability” OR “community accountability” OR overs* OR monitor* OR decentralis* OR decentraliz* OR transparen* OR “parent-teacher partnership*” OR “parent teacher partnership*” OR PTP OR audit) NEAR/3 (educat* OR school*)) OR TS=((Communit* OR civil OR citizen* OR local*) NEAR/3 (empower* OR accountab* OR transparen*) NEAR/3 (educat* OR school*))

#11 TS=((mentor OR peer OR volunteer* OR "role model*" OR "role-model" OR "scholarship plus" OR "study counsel*" OR "directive counsel*" OR feedback) NEAR/3 (school* OR educat*)) OR TS=(((provis* OR dissem* OR invest*) NEAR/3 (inform* OR stat*)) OR ((provide OR providing) AND information)) OR TS=((perceive* OR perception* OR expect* OR estimat*) NEAR/3 (return* OR benefit*) NEAR/3 (educat* OR school*))

#10 #9 OR #8 OR #7 OR #6 OR #5

#9 TS=(((Fee* OR tuition) NEAR/3 (reduc* OR abolish* OR abolition* OR stop* OR eliminat* OR cancel* OR cut OR waiv*)) OR “tuition tax credit” OR scholarship OR “fee free” OR “fee-free” OR “non-fee paying”)

#8 TS=((Voucher* OR credit* NEAR/3 (national OR program* OR plan* OR education* OR school* OR choice)) OR scholarship* OR “equal education” OR “private school aid” OR subsid*)

#7 TS=((Uniform OR uniforms NEAR/3 (school)) OR (User NEAR/3 (payment* OR fee* OR finance*)) OR (education NEAR/3 (charg* OR payment*)))

#6 TS=(scholarship* OR subvention* OR subsid* OR stipend* OR grant* OR donation OR bursary OR bursaries OR “tuition relief” OR “user payment*” OR “merit aid” OR “merit based aid” OR “merit-based aid” OR “merit award”)

#5 TS=("cash transfer*" OR "cash-transfer*" OR (cash NEAR/3 (transfer*)) OR (cash NEAR/3 (payment*)) OR pension OR pensions OR (cash NEAR/3 (incentive*)) OR CCT* OR UCT* OR ((cash OR asset* OR monetary OR economic OR pecuniary OR capital) NEAR/3 (pay* OR transfer* OR incentiv* OR hand-out* OR handout* OR grant* OR aid OR assistance OR benefit* OR help)) OR ("child support" NEAR/3 grant*) or (cash NEAR/3 subsid*) OR "social safety" or "welfare grant*" or "social protection" or "transfer payment*" or "transfer program*" or "poverty alleviation transfer*" OR Oportunidades OR PROGRESA OR "Bolsa familia" OR "Bolsa escola" OR "familias en accion" OR "escuela nueva")

#4 TS=(student* OR pupil* OR child* OR youth* OR youngster* OR “young person*” OR “young people” OR teen* OR adolescen* OR schoolchild*)

#3 TS=("random* control* trial*" OR "random* trial*" OR RCT OR "cluster random* trial" OR "propensity score matching" OR PSM Or "regression discontinuity design" OR RDD OR "difference in difference*" OR DID OR "systematic* review*" OR meta-analy* OR "meta analy*" OR SR OR "control* random* trial*" OR "case control" OR matching OR "interrupted time series" OR "random* allocation*" OR (random* NEAR/3 (allocat*)) OR "instrumental variable*" OR IV OR "research synthesis" OR "scoping review" OR "rapid evidence assessment" OR "systematic literature review" OR evaluation OR assessment OR ((quantitative OR "comparison group" OR counterfactual OR "counter factual" OR counter-factual OR experiment*) NEAR/3 (design OR study OR analysis)) OR QED)

#2 TS=(outcome* OR effect* OR impact* OR attain* OR enrol* OR attend* OR progress* OR achiev* OR result OR results OR complet* OR improve* OR assess* OR perform* OR test* OR mark OR marks OR marking OR learn* OR exam OR exams OR examination* OR graduat* OR matriculat* OR retention OR retain* OR grade* OR grading OR score* OR scoring OR absen* OR truan* OR "drop out*" OR "drop-out*" OR "dropped out" OR qualif* OR cost* OR "cost-effect*" OR "cost-benefit" OR "cost-utility")

#1 TS=(educat* OR teach* OR academ* OR schol* OR school* OR class room OR classes OR classroom* OR class-room* OR pedagog* OR learn* OR lesson* OR curricul*

Appendix B: Search results

Academic Databases

Initial search: November 2013 – January 2014

Database	Hits	Date of search
Web of Science	12346	19/11/2013
CAB Abstracts	11705	26/11/2013
Econlit	1954	27/11/2013
PsycInfo	6684	28/11/2013
Africa-Wide (Ebsco)	1233	29/11/2013
SocIndex	4959	29/11/2013
IBSS (Proquest)	3802	5/12/2013
Econlit (Health terms search)	662	11/12/2013
Africa-Wide (Health terms search)	5415	12/12/2013
Web of Science (Health terms search – SSCI and AHCI)	1331	12/12/2013
PsycInfo (Health terms search)	1523	12/12/2013
CAB Abstracts (Health terms search)	5473	15/12/2013
Medline	3263	18/12/2013
Global Health	2696	18/12/2013
SocIndex (Health terms search)	2298	16/12/2013
ERIC	10140	08/1/2014
WHO GHL	1382	08/1/2014
ASSIA (simplified search)	1372	09/1/2014
TOTAL	**78238**	

Search update: June 2015

Database	Hits	Date of search
Web of Science	2652	13/06/2015
CAB Abstracts	1034	12/06/2015
Econlit	507	12/06/2015
PsycInfo	791	13/06/2015
Africa-Wide (Ebsco)	127	15/06/2015
SocIndex	137	15/06/2015
IBSS (Proquest)	124	15/06/2015
Econlit (Health terms search)	134	12/06/2015
Africa-Wide (Health terms search)	104	15/06/2015
Web of Science (Health terms search – SSCI and AHCI)	310	13/06/2015
PsycInfo (Health terms search)	340	13/06/2015
CAB Abstracts (Health terms search)	1021	09/06/2015
Medline	561	12/06/2015
Global Health	169	13/06/2015
SocIndex (Health terms search)	13	15/06/2015
ERIC	2207	12/06/2015
WHO GHL	7	15/06/2015
ASSIA (simplified search)	48	15/06/2015
TOTAL	**10286**	

Appendix C: Data extraction tools

Context, implementation and cost coding tool

Category No.	ID	Description	Question	Coding
1	ID	Unique study identification #		
2	AUTHOR	First Author	Surname, Initial	Surname, Initial
3	COMMENTS	General comments	**(1) General comments** Any general comments on study not coded elsewhere **(2) Issues of comparability** Please report any potential issues of comparability between different documents (e.g. different documents assess a programme/intervention at different scales [geographic/time scale]). If the issue of comparability related only to a certain secion of a document (e.g. cost data), please put in brackets in relevant cell.	Open answer
4	PUB DATE	Publication date	Year, letter	XXXX (a)
5	PUB TYPE	Publication type	What is the impact evaluation publication type?	1= Peer-reviewed journal 2= Book chapter/book 3= Conference paper 4= Report (other grey literature) 5= working paper 6= implementation document 7= other grey 8= PhD thesis / dissertation

6	FUNDER	Funding agency	Who is funding the evaluation/study?	1= Public institution (e.g. govt, NGO, university, research institute) 2= Private institution (e.g. private company) 3= Multilateral Organisation (World Bank, UN) 4 = Foundations 8= Not clear 9= Not applicable (Non-funded)
7	FUNDER NAME	Name of funding agency	Please add name of the agency funding the evaluation	Open answer
8	INDEPENDENT EVALUATION	Independent evaluation	Is it an independent evaluation (not funded by the implementing agency)?	1=Yes 2=No 8=Not clear
9	INDEPENDENT DATA COLLECTION	Independent data collection	Has the data been collected by an independent party?	1= Yes 2=No 8=Not clear
10	CONFLICT	Conflict of interest	Is there a potential conflict of interest associated with study which could influence results collected/reported? (eg. Is there a declaration of conflict of interest? Is any of the authors related in any way to the funding or implenting institution?)	1=Yes 2=No 8=Not clear
11	CONFLICT COMMENTS	Comments on Conflict of interest	Please add reason for your answer to whether there is a conflict of interest.	Open answer
12	LANGUAGE	Language of publication	Language of publication of the impact evaluation, e.g. Spanish, English etc.	Open answer
13	INTERV LEVEL	Intervention level	Indicate intervention level as per protocol	Child level; Household level; School level; System level; Teacher Level

14	INTERV TYPE	Intervention type	Indicate type of intervention	School based health programme; School Feeding; Providing Information (child-level); Merit based scholarships (child-level), Interventions reducing costs: Scholarships (household-level), Interventions reducing costs: Fee reduction/ elimination, Interventions reducing costs;Conditional Cash transfers; Unconditional Cash Transfers; Providing Information(household-level), Investing in teachers: training; Investing in teachers: employing teachers; Pedagogy; New schools and infrastructure; Providing Materials Teacher incentives & accountability, Decentralisation and local community participation: School based management; Decentralisation and local community participation: Community based monitoring and accountability; Public private

				partnerships and private provision of schooling (including school choice, vouchers) Other (describe interventions if other), Multiple interventions
15	INTERV NAME	Programme or project name	State the programme or project name. If no name, then list the location (e.g. Town, village etc.).	Open answer
16	INTERV IMPLEM AGENCY	Intervention Implementing Agency	Who is implementing the intervention? State the name (and department) of the implementing agency.	Open answer
17	INTERV FUNDER	Intervention Funding Agency	Name of intervention funding agency	Open answer
18	INTERV DESCRIPTION - WHAT	Intervention description	Provide descriptive details about what is delivered to participants as part of the intervention	Open answer
19	INTERV- INPUTS	Intervention inputs	Describe details of types and quantity of inputs provided, including total number of english text books, computers etc. *Also non-physical inputs, e.g. hours of staff time.*	Open answer
20	INTERV - SCALE	Intervention scale	Describe total number of beneficiaries/ schools/ households covered by the intervention	Open answer
21	INTERV DESCRIPTION - WHO	Intervention description	Describe in detail the characteristics of those delivering the intervention (profession, training level, number of staff etc)	Open answer

22	INTERV DESCRIPTION - FREQUENCY	Intervention description	Provide descriptive details about the duration/ frequency/ intensity of the intervention. *This refers to the frequency/duration of the intevention, e.g. number of weeks/months of teacher training, or how regularly cash transfers were delivered, rather than the time period during which the intervention was being implemented (this is asked in questions 26/27)*	Open answer
23	INTERV OBJECTIVES	Objectives of intervention	State any objectives stated in study or project document	
24	PROGRAM THEORY	Program theory	Report any description/statement of program theory as stated by author(s).	Open answer
25	INTERV TARGET	Intervention target group	What were the characteristics of beneficiaries used to target the intervention?	1=Place of Residence 2=Race/Ethnicity, 3=Occupation, 4=Gender 5=Religion 6=Education 7=Socioeconomic Status 8=Social Capital 9=School type (eg: public, religious etc) 10=School quality 11=School level/ grade 12= Age 13= Not clear 14= Other 15 = No targeting
26	TARGET METHODS	Targeting methods	How were beneficiaries targeted for the programme (Eg: how was the targeting implemented)?	Open answer
27	INTERV START	Intervention start	Start date (if not stated, state study date) of intervention	XX/XXXX
28	INTERV END	Intervention end	State end date	

29	FIDELITY/ ADHERENCE	Information about implementation fidelity/ service delivery quality	Is there any information on implementation fidelity/ service delivery quality? *Commentary by authors should be used when information on program adherence etc. is not backed up by some sort of research / when the authors do not report that/how they collected data to assess these areas. Also applies to question 31, TAKE-UP.*	1=Yes, commentary from author; 2=No; 4= Yes, formally assessed
30	FIDELITY/ ADHERENCE - METHODS	Methods of assessing intervention fidelity	Which methods are used to assess implementation fidelity/ service delivery quality	1= Observation by intervention staff 2= Reporting by participants 3= Other 4= Commentary from author 9= N/A
31	FIDELITY/ ADHERENCE - RESULTS	Results of the assessment of intervention fidelity	What is the result/ information provided of the assessment of implementation fidelity/ service delivery quality	Open answer
32	TAKE-UP	Information about program take-up/adherence (among beneficiaries)	Is there any information about program take-up/adherence (among beneficiaries)?	1=Yes, commentary from author; 2=No; 4= Yes, formally assessed
33	TAKE UP - METHODS	Methods of assessing take-up/adherence	Which methods are used to assess program take-up/adherence?	1= Observation by intervention staff 2= Reporting by participants 3= Other 4= Commentary from author 9= N/A
34	TAKE UP - RESULTS	Results of the assessment of take-up/adherence	What is the result/ information provided of the assessment of program take-up/adherence?	Open answer

35	PROCESS - OTHER DECR	Other description of process factors	Any other description of process factors not covered above	Open answer
36	OTHER INTERV T	Other interventions treatment group	Describe other education related interventions undertaken in treatment group. State if there is no other intervention (NO OTHER INTERVENTION), or no other intervention reported by author(s) (NO OTHER INTERVENTION REPORTED)	Open answer
37	INTERV C	Intervention comparison group	Describe education interventions available to comparison group. State if there is no other intervention (NO OTHER INTERVENTION), or no other intervention reported by author(s) (NO OTHER INTERVENTION REPORTED)	Open answer
38	COST	Cost	Are any unit cost data / cost-effectiveness estimates provided?	1=Yes 2=No
39	COST DATA	Cost details	If yes, report any details of unit cost and/or total cost. Please also report year and currency.	Open answer
40	COUNTRY	Country	List countries the study was conducted in	Country 1, Country 2, etc.
41	DETAILED LOCATION	Detailed location	If provided, give detailed information on where the study took place within a country, for example regions/districts covered	Open answer
42	REGION	Wolrd Bank Region	List region(s) the study was conducted in according to World Bank. For more info on region classification see http://data.worldbank.org/country	1= East Asia & Pacific 2=Europe& Central Asia 3=Latim America & Caribbean 4=Middle East & North Africa 5=South Asia

				6=Sub-Saharan Africa
43	RURAL	Rural	Is study conducted in rural areas?	1=Yes 2=No 3=Not clear
44	URBAN	Urban	Is study conducted in urban areas?	1=Yes 2=No 3=Not clear
45	PERI	Peri-urban	Is study conducted in peri-urban areas? (eg, city outskirts, adjoining urban areas, just outside city boundaries)	1=Yes 2=No 3=Not clear
46	SCHOOL TYPE	School type	Which school level(s) is covered by the intervention?	1=Primary school; 2=Secondary school; 3= Other
47	PROVIDER	Provider of School	Who is the school provider? *If not reported, note as "not clear", even if it is a plausible assumption that the intervention applies to multiple providers since it is a national programme. Please use 7: no schools at baseline, when the intervention involves building new schools*	1=Public; 2= Private for profit; 3=Private NGO; 4= Religious; 5=Multiple; 6= Other; 7= no schools at baseline 8=not clear
48	SCHOOL CHARACT	School characteristic s	Report any characteristics of schools reported, including class size/ pupil teacher ratio, infrastructure, materials available etc	Open answer
49	EXTERNAL EVENT	Significant external event external to school/ education system	Did any significant external events (external defined as outside of the school/ education system) occur at the time of intervention?	Open answer
50	EDUCATION POLICY	Description of education policy	Please provide any info provided about relevant education policy, including information about administrative arrangements	Open answer

			such as decentralisation, legistation etc	
51	OTHER CONTEXT	Description of other contextual factors	Please provide any additional contextual information that may be of importance to the intervention effectiveness and/ or implementation, such as security issues, cultural factors etc.	Open answer
52	STUDENT GENDER	Gender of students in sample	What percentage of females in sample?	Open answer
53	STUDENT AGE	Average age of students in sample	What was the (average) age of students in the sample? (Averages are preferable. If not available, report any other available data e.g. age range).	Open answer
54	STUDENT GRADE	Grade of students in sample	What was the grade(s) (school year) of students in the sample?	Open answer
55	STUDENT CHARACTERISTICS	Student characteristics	Report any (average) student characteristics reported by the authors	Open answer
56	HOUSEHOLD CHARACT	Household characteristics	Report any (average) household characteristics reported by the authors (for example, household size, education of household leader, number of school age children, etc).	Open answer
57	TEACHER CHARACT	Teacher characteristics	Report any (average) teacher characteristics reported by the authors (Eg: age, education level, gender etc)	Open answer
58	STUDY A & O	Study aims and objectives	Please state the aims and objectives as stated by the authors	Open answer
61	OTHER METHODS	Other methods	If the study address other questions than effectiveness note questions and methods used here.	Open answer

62	PRIMARY OUTCOMES	Primary outcomes measured	Which primary outcomes are measured in the study? *Primary and secondary outcome coding categories refer to "our" primary and secondary outcomes as defined in the EER protocol, not the authors' primary outcomes*	1=enrolment; 2=attendance; 3= drop-out; 4= completion; 5= learning
63	PRIMARY OUTCOMES - DEF	Definition of primary outcomes	Please provide the authors definition of each primary outcome included in the study.	Open answer
64	SECONDARY OUTCOMES	Secondary outcomes measured	Which secondary outcomes are measured in the study? Please report any other education related secondary outcomes, including: (1) teacher attendance: defined as a measure of the proportion of total school days for which teachers are present; (2) teacher performance: defined as any measure of teachers' knowledge, practice, motivation or satisfaction (Orr et al., 2013).	1= teacher attendance; 2= teacher performance; 3=Other (state outcomes in brackets); 9=Not applicable
65	SECONDARY OUTCOMES	Definition of secondary outcomes	Please provide the authors definition of each secondary outcome included in the study.	Open answer
66	OTHER OUTCOMES	Outcomes measured in study	Which other secondary and intermediate outcomes are measured in the study? This can include for instance health status of children, child labour, parental participation etc.	Open answer
67	OTHER OUTCOMES	Definition of other outcomes	Please provide the authors definition of any other outcomes included in the study.	Open answer
68	OTHER EDU OUTCOMES	Other education outcomes	Please report any excluded learning outcomes	Open answer

69	DATA METHODS	Methods of data collection	Describe methods of data collection *Please report what methods of data collection were used for each type of analyses (if the impact evaluation also contains relevant qualitative analysis, for example implementation assessmenet or targeting analysis)*	Open answer
70	DATA FREQ	Data collection frequency	What is the frequency of outcome data collection?	Open answer question
71	STUDY START	Study start	Start date of collection of data on outcome	XX/XXXX
72	STUDY END	Study end	End date of collection of data on outcome	XX/XXXX
73	LENGTH	Length of study	Length of study in months (Where study length not reported, code as length of intervention, noting that in brackets)	# months
74	UNIT ASSIGNMENT	Unit of assignment	At which level was assignment to treatment and control group conducted?	1=Individual 2=Household 3=School/ cluster 9= N/A
75	EFFICACY/ EFFECTIVENESS	Efficacy or effectiveness trial	Was the intervention implemented under "real world" conditions?	1=Yes 2=No 9= N/A
76	EFFICACY/ EFFECTIVENESS METHODS	Methods of implementin g trial	Desription of methods for implementing the trial to corroborate answer to question above	Open answer
77	PERSONELL	Personell implementin g the programme	Who was in charge of implementing the program?	1=PI/ researchers (study authors); 2= implementing agency staff, 3= external agency (eg: survey firm); 4=Others; 8= Not clear
78	STUDY SAMPLING FRAME	Sampling frame for the study	State the sampling frame (list of all those within a population who can be sampled, ie. Students, households, schools,	Open answer

			communities) for selection of study participants (i.e. Census, membership list of parents' association, list of students, etc).	
79	DESCRIBE SAMPLE SELECTION	Description of how the sample was selected	How was the study population sampled and why? Include description of sampling and any justification provided by the authors	Open answer
80	REPRESENTATIVE SAMPLE	Representative sample used in study	What population was the sample representative of?	Open answer

Effect size data extraction and risk of bias coding tool

Category No.	ID	Description	Question	Coding
1	ID	Unique study identification #		
58	STUDY DESIGN	Design type	What type of study design is used?	1= Randomised controlled trial (RCT) (experiment with random assignment to households/individuals) 2= Cluster-RCT 3= Quasi-RCT (experiment with quasi-random assignment to households/individuals) 4= Cluster-quasi-RCT 5= RDD (quasi-experiment with discontinuity assignment) 6 = CBA (quasi-experiment with baseline and endline data collection) 7=Natural experiment 8= Interupted time series 9=Other

59	STUDY ANALYSIS	Methods used for analysis	Which methods are used to control for selection bias and confounding?	1=PSM 2=Covariate matching 3=DID 4=IV-regression 5=Heckman selection model 6= Fixed effects regression 7= Other regression 8=Other
80	T&C SELECTION	Comments on treatment and control selection	Provide details on the treatment and control group selection (eg, school lottery, households selected from local assotiation memberlist)	Open answer
81	T&C COMPAR	Comparability of treatment & control	Is discussion of treatment and control comparability given?	1=Yes 2=No 9= N/A
82	T&C VARIABLES	Variables used in assessing similarity between treatment and control groups	Does the study state variables on which comparability of treatment and control is assessed?	1=Yes 2=No 9= N/A
83	LIST VARIABLES	List of variables used in control group selection	Variables considered in assessment of similarity (e.g. location, socioeconomic status, baseline schooling conditions; education levels)	Variable 1, variable 2, etc.
84	MATCHING	Assessment of covariates balance	Are covariates in treatment and control groups assessed as balanced, and if unbalanced controlled in adjusted analysis?	1=Yes 2=No 9= N/A

85	MATCHING METHOD	Method used to match or control for covariates	List techniques used to match (incl matching variables)	Open answer (eg propensity score matching; matching variables include socio-economic status, location, gender, age, household size)
86	CONTROL ADEQ	Control adequate	Control is of adequate comparability, moderate adequacy, or not adequate	1= Yes, control is adequte, either through randomisation of selection to intervention and control, or matching, or adjustment in multivariate regression analysis, or comparability of characteristics which are reported on and are sufficiently similar 2= Adequacy of control is moderate; general statements made on similarity of some variables between treatment and control groups, no adjustment for confounders in multivariate analysis 3= Control is inadequate; nothing reported on similarities between treatment and control groups, or control not random representative sample of non-users
87	SPILLOVERS	Likelihood of spillovers and crossovers	Is control group geographically separated from treatment, or if not separated is it unlikely that comparisons received the intervention?	1=Yes 2=No 8= Not clear 9= N/A
88	COMP GROUP	Type of comparison group		1=No intervention (business as usual) 2=Other Education intervention 3=Placebo control

				4=Pipeline (wait-list) control
89	CONTAMINATION	Other intervention differentially received by comparison group	Describe any non-education comparison group intervention received which treatment group does not?	Open answer
90	CONTAM METHODS	Contamination methods	If yes, how do authors control for contamination? Describe methods to assess contamination	Open answer
91	BLIND PARTICIPANTS	Blinded participants	Blinding of participants?	1=Yes 2=No 9= N/A
92	BLIND OBSERVERS	Blinded observers	Blinding of outcome assessors?	1=Yes 2=No 9= N/A
93	BLIND ANALYSTS	Blinded analysts	Blinding of data analysts	1=Yes 2=No 9= N/A
94	BLIND METHOD	Method used to blind	Describe method(s) used to blind	Open answer (including describe method of placebo control)
95	UOA	Unit of analysis	Are there any unit of analysis errors which you are not able to recalculate?	1=Yes 2=No 8=Not clear 9= N/A
96	EFFECT SIZE LOCATION		Which page(s) contain the effect size data?	Open answer
97	S SIZE METRIC	Sample size metric	Sample size unit of analysis	1= Children 2= Households 3=Teachers 4= Groups (e.g. Class, school) 5= Other 6= Not clear
98	S SIZE TREAT	Sample size (treatment)	Initial sample size treatment group	#
99	S SIZE CONTR	Sample size (control)	Initial sample size control group	#
100	ATTRIT TREAT	Treatment attrition	Number of drop-outs	#

101	ATTRIT CONTR	Control attrition	Number of drop-outs	#
102	# TREAT	Observations (treatment)	Number of treatment observations after attrition (individuals)	#
103	# CONTROL	Observations (control)	Number of control observations after attrition (individuals)	#
104	TREAT EFFECT	Treatment effect estimated	What treatment effect is estimated?	1=ITT 2=ATET 3=ATE 4=LATE
105	OUTCOME DEF	Outcome	Does the study give a precise definition of outcome X?	1=Yes 2=No 3=Partially
106	DEFINITION	Definition of outcome	What definition of outcome x given	Open answer
107	BASE T MEAN	Baseline outcome treatement	State result of baseline outcome for treatment group	#
108	BASE T SD	SD Baseline outcome treatement	State SD of baseline outcome measure for treatment group	#
109	N BASE T	Sample size baseline treatment	State sample size at baseline	#
110	BASE C MEAN	Baseline outcome control	State result of baseline outcome for control group	#
111	BASE C SD	SD Baseline outcome control	State SD of baseline outcome measure for contol group	#
112	N BASE C	Sample size baseline control	State sample size at baseline	#
113	OUTCOME POST INTERV T	Outcome in treatment post intervention	State result of post intervention outcome for treatment group	#
114	SD POST INTERV T	SD Outcome in treatment post intervention	State SD of post intervention outcome measure for treatment group	#
115	N OUTCOME POST INTERV T	Number with outcome in	State sample size post intervention	#

		treatment post intervention		
116	OUTCOME POST INTERV C	Outcome in control post intervention	State result of post intervention outcome for control group	#
117	SD POST INTERV C	SD Outcome in control post intervention	State SD of post intervention outcome measure for control group	#
118	N OUTCOME POST INTERV C	Number with outcome in contol post intervention	State sample size post intervention	#
119	OUTCOME 1st FOLLOW UP T	Outcome in treatment 1st follow up	State result of 1st follow up outcome measure for treatment group	#
120	SD 1st FOLLOW UP T	SD Outcome in treatment 1st follow up	State SD 1st follow up outcome measure for treatment group	#
121	N OUTCOME 1st FOLLOW UP T	Number with outcome in treatment 1st follow up	State sample size first follow up	#
122	OUTCOME 1st FOLLOW UP C	Outcome in control 1st follow up	State result of 1st follow up outcome measure for treatment group	#
123	SD 1st FOLLOW UP C	SD Outcome in control 1st follow up	State SD 1st follow up outcome measure for treatment group	#
124	N OUTCOME 1st FOLLOW UP C	Number with outcome in control 1st follow up	State sample size first follow up	#
125	OTHER		Repeat the above for any additional follow up measures	
126	OUTCOME DEF	Outcome	Does the study give a precise definition of outcome X?	1=Yes 2=No 3=Partially
127	DEFINITION	Definition of outcome	What definition of outcome x given	Open answer

128	BASE OUTCOME T	Baseline number with outcome in treatement	State result of baseline outcome for treatment group	#
129	N BASE T	Sample size baseline treatment	State sample size at baseline	#
130	PROP BASE OUTCOME T	Proportion with outcome at baseline in treatment	State proportion with outcome at baseline in treatment	#
131	BASE OUTCOME C	Baseline number with outcome in control	State result of baseline outcome for treatment group	#
132	N BASE C	Sample size baseline control	State sample size at baseline	#
133	PROP BASE OUTCOME C	Proportion with outcome at baseline in control	State proportion with outcome at baseline in contol	#
134	No W/ OUTCOME POST INTERV T	Number with outcome in treatment post intervention	State number with outcome post intervention for treatment group	#
135	N POST INTERV T	Sample size post intervention treatment	State sample size for treatment group post intervention	#
136	PROP OUTCOME POST INTERV T	Proportion with outcome in treatment group post intervention	State proportion with outcome post intervention in control group	#
137	No W/ OUTCOME POST INTERV C	Number with outcome in control post intervention	State number with outcome post interventionfor control group	#
138	N POST INTERV C	Sample size post intervention control	State sample size for control group post intervention	#
139	PROP OUTCOME POST INTERV C	Proportion with outcome in control	State proportion with outcome post intervention in control group	#

		group post intervention		
140	No W/ OUTCOME 1st FOLLOW UP T	Number with outcome in treatment 1st follow up	State number with outcome at 1st follow up for treatment group	#
141	N 1st FOLLOW UP T	Sample size 1st follow up treatment	State sample size at 1st follow up for treatment group	#
142	PROP OUTCOME 1st FOLLOW UP T	Proportion with outcome in treatment group 1st follow up	State proportion with outcome at 1st follow up in treatment group	#
143	No W/ OUTCOME 1st FOLLOW UP C	Number with outcome in contro 1st follow up	State number with outcomeat 1st follow up for control group	#
144	N 1st FOLLOW UP C	Sample size 1st follow up control	State sample size at for control group at 1st follow up	#
145	PROP OUTCOME 1st FOLLOW UP C	Proportion with outcome in contol group 1st follow up	State proportion with outcome at 1st follow up in control group	#
146	OTHER		Repeat the above for any additional follow up measures	
147	SUB GROUP	Sub group analysis	Does the study conduct sub group analysis	1=Yes 2=No
148	SUB GROUP OUTCOMES	Types of sub-groups included	State any sub-groups for which the study includes outcome measures	
149	OLS	OLS	OLS used?	1=Yes 2=No
150	LOGISTIC 1	Logistic	Logistic used?	1=Yes 2=No
151	LOGISTIC 2	Type of logistic	What type of logistic regression?	1=binomial 2=multinomial
152	GLS	GLS/WLS	GLS or WLS used?	1=Yes 2=No
153	POISSON	Poisson	Poisson regression used?	1=Yes 2=No

154	OTHER REG	other regression types	Other regression type used? Specify	open answer
155	MULTILEVEL	multilevel models	Is this a multilevel model?	1=Yes 2=No
156	PREDIDS	number of predictors	How many predictors/covariates (not including the intercept) are in the model?	#
157	CONTOUTCOME	continous outcome	Is the outcome continous?	1=Yes 2=No
158	DIOUTCOME	dichotomus outcome	Is the outcome dichotomus?	1=Yes 2=No
159	MULTICATEGORIES	multiple outcome categories	Does the outcome have more than 2 categories?	1=Yes 2=No 3=Continous
160	VAREXPLAINED	variance explained	What is the variance explained in the model?	#
161	COEFFTYPE	type of coefficient	What is the coefficient type?	1=raw 2=standardized 3=other
162	COEFF	coefficient	What is the coefficient estimate?	#
163	STANDARDERROR	standard error	What is the standard error of the coefficient estimate?	#
164	TTEST	t test	What is the t statistic associated with the focal predictor?	#
165	WALD	Wald test	What is the Wald statistic associated with the focal predictor?	#
166	TRTPOR	treatment proportion	What is the treatment proportion?	#
167	CONPOR	control proportion	what is the control proportion?	#

Appendix D: Critical appraisal of studies included to answer questions 2a and 2b[148]

Critical appraisal of quantitative and qualitative studies

Is the research aim clearly stated? (Yes/No)

REPORTING:

2. Description of the context? (Yes/No)
3. Description of sampling procedures? (Yes/No)

- *How have the participants been selected, were they the most appropriate?*

4. Are sample characteristics sufficiently reported? (sample size, location, and at least one additional characteristic) (Yes/No)
5. Is it clear how the data were collected (eg: for interviews, is there an indication of how interviews were conducted? (Yes/No)
6. Methods of recording of data reported? (Yes/No)
7. Methods of analysis explicitly stated? (Yes/No)

METHODOLOGY:

8. Is there a clear link to relevant literature/theoretical framework? (Yes/No)
9. Is the design appropriate to answer the research question? (Yes/No)

- *Has the researcher justified the research design?*

10. Was the sampling strategy appropriate to the aims of the research? (Yes/No)

- *Have the researchers explained how the participants were selected?*
- *Have the researchers explained why the participants they selected were the most appropriate to provide access to the type of knowledge sought by the study?*
- *Have the researchers discussed issues around recruitment? (e.g. why some people chose not to take part)*

11. Were the data collected in a way that addressed the research issue? (Yes/No)

- *Were the methods used appropriate and justified?*
- *Did the researcher discuss saturation of data?*

12. Was the data analysis sufficiently rigorous? (Yes/No)

- *Is there a detailed description of the analysis process?*
- *Does the data support the findings?*

[148] The appraisal tool is and adapted version of CASP (2006), adapted by Waddington et al (2012).

- *Is the relationship between the researcher and the participants adequately considered?*

- *To what extent is contradictory data are taken into account?*

- *If the findings are based on quantitative analysis of survey data, are multivariate techniques used to control for potential confounding variables?*

13. Has triangulation been applied? (Yes/No)

- *Data triangulation (location, time and participants)*

- *Investigator triangulation*

- *theory triangulation (several theories)*

- *methodological triangulation*

14. Is the analysis and conclusions clearly presented? (Yes/No)

- *Have the researchers discussed the credibility of their findings? (e.g. triangulation, respondent validation, more than one analyst)*

- *Is there adequate discussion of the evidence both for and against the researcher's arguments?*

- *Are the findings explicit?*

- *Are the findings discussed in relation to the original research question?*

15. Was there potential for conflict of interest and if so, was this considered and addressed? (Yes/No)

16. Does the paper discuss ethical considerations related to the research? (Yes/No)

Critical appraisal of process evaluations

Process evaluations assess whether a policy is being implemented as intended and what, in practice, is felt to be working more or less well, and why. Process evaluations often include the collection of qualitative and quantitative data from different stakeholders to cover subjective issues (perceptions of policy success) or objective aspects (how a policy has operated). They might also be used to collect organisational information.

1. Is the research aim clearly stated? (Yes/No)

REPORTING:

2. Description of the context? (Yes/No)
3. Description of sampling procedures? (Yes/No)

- *How have the participants been selected, was the approach appropriate?*

4. Are sample characteristics sufficiently reported? (sample size, location, and at least one additional characteristic) (Yes/No)

5. Is it clear how the data were collected (eg: for interviews, is there an indication of how interviews were conducted? (Yes/No)

6. Methods of recording of data reported? (Yes/No)

7. Methods of analysis explicitly stated? (Yes/No)

METHODOLOGY:

8. Is the design appropriate to answer the research question? (Yes/No)

9. Was the sampling strategy appropriate to the aims of the research? (Yes/No)

- Have the researchers explained how the participants were selected?

- Have the researchers explained why the participants they selected were the most appropriate to provide access to the type of knowledge sought by the study?

- Have the researchers discussed issues around recruitment? (e.g. why some people chose not to take part)

10. Were the data collected in a way that addressed the research issue? (Yes/No)

- *Were the methods used appropriate and justified?*

11. Was the data analysis sufficiently rigorous? (Yes/No)

- *Is there a description of the analysis process?*

- Does the data support the findings?

- Is the relationship between the researcher and the participants adequately considered?

- To what extent are contradictory data taken into account?

- If the findings are based on quantitative analysis of survey data, are multivariate techniques used to control for potential confounding variables?

12. Has triangulation been applied? (Yes/No)

- Data triangulation (location, time and participants)

- Investigator triangulation

- Methodological triangulation

13. Are the analysis and conclusions clearly presented? (Yes/No)

- Is there adequate discussion of the evidence both for and against the researcher's arguments?

- *Are the findings explicit?*

- Are the findings discussed in relation to the original research question?

14. Was there potential for conflict of interest and if so, was this considered and addressed? (Yes/No)

15. If appropriate, does the paper discuss ethical considerations related to the research? (Yes/No)=

Appendix E: Efficacy studies tool

Efficacy Trials

We will include studies that are primarily designed to determine the effects of an intervention implemented as part of a program under circumstances that approach usual, 'real- world' practice, so- called effectiveness studies. These types of studies stand in contrast to efficacy trials which test an intervention under ideal and controlled conditions in order to maximise the likelihood of observing an effect, if one exists.

Although there exists broad agreement on the type of study design characteristics of effectiveness (pragmatic) trials and efficacy (explanatory) trials, there is currently no validated definition of 'effectiveness studies' (Treweek et al.. 2009; Gerthlener et al.. 2006; Singal et al. 2014). Furthermore, as argued by Thorpe et al. (2009), the distinction between the two types of trials should be regarded as a continuum rather than a dichotomy as very few trials are purely pragmatic or explanatory.

In order to distinguish effectiveness from efficacy studies we developed five criteria which are based on two existing tools. The first, developed by Gartlehner et al. (2006), proposes an instrument based on seven criteria of study design to distinguish effectiveness from efficacy trials while conducting systematic reviews. The second tool is the pragmatic-explanatory continuum indicator summary (PRECIS) tool developed to help trial designers assess the degree to which design decisions align with the trial's stated purpose (Thorpe et al.: 2009).

Studies will be considered efficacy trials and will therefore be excluded if they fulfil at least one of the criteria outlined below:

Research Objective:

(1) Is the study primarily designed to determine to what extent a specific technique, technology, treatment, procedure or service works under ideal condition rather than attempt to answer a question relevant to the roll- out of a large program?

Population:

(2) Are the participants highly selected and therefore unrepresentative of the general population (Are strict inclusion and exclusion criteria used to enrol a homogenous population which may limit the generalizability of the results? e.g. students that truly have a disease of interest or are more likely to adhere to the treatment)?

Providers:

(3) Is the intervention primarily delivered by the research study team rather than trained laypersons (parents/ teachers/ community members/ NGOs) who don't have extensive expertise?

Delivery of intervention

(4) Is the intervention delivered with high degree of assurance of delivery of the treatment? (Is the delivery tightly monitored/ supervised by the researcher following specific protocols; Is adherence to the treatment monitored closely with frequent follow- ups?)

(5) Are concurrent interventions restricted to the study population in order for a witnessed effect to be attributed to the intervention of interest?

Appendix F: Risk of bias – Full results of assessment

Study	Intervention area	Study design	Selection bias and confounding	Spill-overs, cross-overs and contamination	Outcome reporting	Analysis reporting	Performance bias	Other risks of bias
Banerjee et al., 2007	CAL	Cluster RCT	Low	Unclear	Low	Low	High	Unclear
Barrera-Osorio & Linden, 2009	CAL	Cluster RCT	Low	Low	Low	Low	High	Unclear
Cristia et al., 2012	CAL	Cluster RCT	Unclear	Unclear	Low	Low	Unclear	High
Quispe, 2013	CAL	CBA	NA	NA	NA	NA	NA	NA
Cristia et al. 2014	CAL	CBA	Unclear	Low	Low	Low	Low	Unclear
Linden, 2008	CAL	Cluster RCT	Low	Unclear	Low	High	High	Unclear
Mo et al., 2014	CAL	Cluster RCT	Low	Low	Low	Low	Low	Unclear
Lai et al. ND; Yang et al. 2013	CAL	Cluster RCT	Low	Low	Low	Low	Unclear	Unclear
Lai et al. 2013; Mo et al. 2014	CAL	Cluster RCT	Low	Low	Low	Low	Unclear	Unclear
Yang et al. 2013	CAL	Cluster RCT	Low	Low	Low	Low	Unclear	Unclear
Carillo et al. 2010	CAL	Cluster RCT	Low	Unclear	Unclear	Unclear	High	Unclear
De Melo et al. ND	CAL	CBA	Unclear	Unclear	Low	Low	Low	Low
Linden, 2008	CAL	Cluster RCT	Low	Unclear	Low	Low	High	Unclear
Berlinski & Busso, 2013	CAL	Cluster RCT	Low	Unclear	Low	Low	High	Unclear

Study	Intervention area	Study design	Selection bias and confounding	Spill-overs, cross-overs and contamination	Outcome reporting	Analysis reporting	Performance bias	Other risks of bias
Sharma, 2014	CAL	CBA	High	Unclear	Low	Low	Low	Low
Humpage, 2013	CAL	Cluster RCT	Unclear	Unclear	Low	Low	Unclear	Unclear
Imbrogno, 2014	CAL	Cluster RCT	High	Unclear	Low	Unclear	Low	Unclear
Imbrogno, 2014	CAL	Cluster RCT	High	Unclear	Low	Unclear	Low	Unclear
Banerjee et al. 2010	Community based monitoring	Cluster RCT	Low	Unclear	Low	Low	Low	High
Glewwe & Maïga 2011	Community based monitoring	Cluster RCT	Low	Unclear	Low	Low	High	High
Lassibille et al. 2010	Community based monitoring	Cluster RCT	Low	Unclear	Low	Low	High	High
Andrabi et al., 2013	Community based monitoring	Cluster RCT	Low	Unclear	Low	Low	High	High
Reinikka & Svensson, 2007	Community based monitoring	Natural Experiment	Unclear	Unclear	Unclear	Low	Unclear	Unclear
Bjorkman, 2006	Community based monitoring	Natural Experiment	Unclear	Unclear	Low	Low	Unclear	Unclear

Study	Intervention area	Study design	Selection bias and confounding	Spill-overs, cross-overs and contamination	Outcome reporting	Analysis reporting	Performance bias	Other risks of bias
Pandey et al., 2011	Community based monitoring	Cluster RCT	High	Low	Low	Unclear	Unclear	High
Zeitlin et al., 2012	Community based monitoring	Cluster RCT	Low	Unclear	Low	Low	Unclear	Low
Mizala & Urquiola, 2013	Community based monitoring	RDD	Low	Unclear	Low	Low	Low	High
Camargo et al., 2012	Community based monitoring	RDD	Unclear	Low	Low	Low	Low	High
Duflo et al. 2007	Community based monitoring	Cluster RCT	High	High	Low	Unclear	Low	High
Muralidharan & Sundararaman, 2010	Diagnostic feedback	Cluster RCT	Low	High	Low	Low	Low	Low
Duflo et al. 2015	Diagnostic feedback	Cluster RCT	Low	Low	Low	Low	Low	High
Chen et al. 2010	Grade retention	CBA	Unclear	Unclear	Low	Low	High	Unclear

Study	Intervention area	Study design	Selection bias and confounding	Spill-overs, cross-overs and contamination	Outcome reporting	Analysis reporting	Performance bias	Other risks of bias
Muralidharan & Sundararaman, 2013	Hiring additional teachers	Cluster RCT	Low	Unclear	Low	Low	High	High
Duflo et al., 2012	Hiring Additional Teachers	Cluster RCT	High	High	Low	Unclear	Unclear	Low
Estrada, 2013	Hiring additional teachers	CBA	Unclear	Unclear	Low	Low	High	High
Vegas & Laat, 2003	Hiring additional teachers	CBA	Unclear	Unclear	Low	Low	High	High
Bold et al., 2013	Hiring additional teachers	Cluster RCT	Low	Unclear	Low	Low	High	High
Bau, 2014	Hiring additional teachers	RDD	Unclear	Low	Low	Unclear	Low	Low
Chin, 2005	Hiring additional teachers	CBA	High	Unclear	Unclear	Unclear	Unclear	Unclear
Ome, 2012	Hiring additional teachers	Natural Experiment	Unclear	Low	Low	Low	Low	Low
Kremer et al., 2009	Merit-based scholarships	Cluster RCT	Low	Unclear	Low	Low	High	High
Chen et al., 2013	Merit-based scholarships	Cluster RCT	Low	Unclear	Low	Unclear	High	Unclear

Study	Intervention area	Study design	Selection bias and confounding	Spill-overs, cross-overs and contamination	Outcome reporting	Analysis reporting	Performance bias	Other risks of bias
Yi et al., 2015	Merit-based scholarships	RCT	Low	Low	Low	Low	High	Unclear
Yi et al., 2015	Merit-based scholarships	RCT	Low	Low	Low	Low	High	Unclear
Sharma, 2011	Merit-based scholarships	Cluster RCT	Unclear	Unclear	Low	Low	Unclear	Low
Blimpo, 2014	Merit-based scholarships	Cluster RCT	Low	Unclear	Unclear	Low	High	Low
Behrman et al., 2012	Merit-based scholarships	Cluster RCT	Low	Unclear	Low	Low	High	High
Barrera-Osorio, 10299438	Merit-based scholarships	Cluster RCT	High	Unclear	High	Low	Unclear	Low
Li et al., 2014	Merit-based scholarships	Cluster RCT	Low	Low	Low	Low	Low	Low
Li et al., 2014	Merit-based scholarships	Cluster RCT	Low	Low	Low	Low	Low	Low
Berry, 2013	Merit-based scholarships	RCT	Low	High	Low	Unclear	High	Low
De Hoop & Rosati, 2012	Multiple Interventions	RDD	Unclear	Low	Low	Low	Low	High
Kazianga et al., 2013	Multiple Interventions	RDD	Low	Low	Low	Low	Unclear	High

Study	Intervention area	Study design	Selection bias and confounding	Spill-overs, cross-overs and contamination	Outcome reporting	Analysis reporting	Performance bias	Other risks of bias
Cerdan, Infantes & Vermeersch, 2007	Multiple interventions	CBA	Unclear	Low	Low	Unclear	Unclear	High
Tokman, 2002	Multiple interventions	CBA	Unclear	Low	Low	Unclear	Unclear	Low
Chay et al., 2005	Multiple interventions	RDD	Unclear	Low	Low	Unclear	High	High
Kremer et al., 2003	Multiple interventions	Cluster RCT	Unclear	High	Low	Low	High	High
Lockheed et al., 2010	Multiple interventions	CBA	Unclear	High	Low	Low	Unclear	Low
Min et al., 2012	Multiple interventions	Natural experiment	High	Unclear	Low	Low	Unclear	High
Paqueo, 2003	Multiple interventions	CBA	Unclear	Low	High	Low	Unclear	Low
Rodriguez et al. 2010	Multiple interventions	CBA	Unclear	Low	Low	Low	Unclear	Low
Rosati & Rossi, 2007	Multiple interventions	Natural experiment	Unclear	Low	Low	Low	Unclear	Low
Bellei, 2013	Multiple interventions	CBA	Unclear	Low	Low	Low	Unclear	Low

Study	Intervention area	Study design	Selection bias and confounding	Spill-overs, cross-overs and contamination	Outcome reporting	Analysis reporting	Performance bias	Other risks of bias
Burde & Linden, 2013	New schools and infrastructure	Cluster RCT	Low	Unclear	Low	Unclear	High	Unclear
Newman et al., 2002	New schools and infrastructure	Cluster RCT, CBA	Unclear	Unclear	Low	Low	Low	High
Lokshin & Yemtsov, 2003	New schools and infrastructure	CBA	Unclear	Unclear	Low	Unclear	High	Low
Borkum et al., 2013	New schools and infrastructure	Cluster RCT	Low	Low	Low	High	High	Unclear
Dumitrescu et al., 2011	New schools and infrastructure	Cluster RCT	Unclear	Unclear	Low	High	Unclear	High
Adukia, 2014	New schools and infrastructure	CBA	Unclear	Unclear	Low	Low	High	Unclear
Freeman et al., 2012	New schools and infrastructure	Cluster RCT	Unclear	Low	Low	Low	Unclear	High
He et al. 2008	Pedagogy	Cluster RCT	Low	Unclear	Low	Unclear	High	Unclear
He et al. 2008	Pedagogy	Cluster RCT	Low	Unclear	Low	Unclear	High	Unclear

Study	Intervention area	Study design	Selection bias and confounding	Spill-overs, cross-overs and contamination	Outcome reporting	Analysis reporting	Performance bias	Other risks of bias
Tan et al. 1999	Pedagogy	Cluster RCT	Low	Unclear	Low	Low	Low	High
Abeberese et al. 2011	Pedagogy	Cluster RCT	Low	Unclear	Low	Low	High	High
San Antonio et al. 2011	Pedagogy	Cluster RCT	Low	Unclear	Unclear	Unclear	High	High
He et al. 2009	Pedagogy	Cluster RCT	Low	Unclear	Low	Low	High	Unclear
Dixon et al. 2011	Pedagogy	Cluster RCT	High	Unclear	Low	Low	High	Unclear
Irwing et al., 2008	Pedagogy	CBA	High	Unclear	Unclear	High	High	Unclear
Nonoyama-Tarumi & Bredenberg 2009	Pedagogy	CBA	High	Low	Low	Unclear	High	Unclear
Lucas et al. 2014	Pedagogy	Cluster RCT	Low	Unclear	Low	Low	High	Unclear
Lucas et al. 2014	Pedagogy	Cluster RCT	Low	Unclear	Low	Low	High	Unclear
Piper & Korda 2010	Pedagogy	Cluster RCT	Unclear	Low	Low	Low	Unclear	High
Piper & Mugenda 2014	Pedagogy	Cluster RCT	High	Low	Low	Unclear	Unclear	High
Spratt et al. 2013	Pedagogy	RCT	Low	Unclear	Low	Low	High	Unclear
Leme et al. 2012	Pedagogy	CBA	Unclear	Low	Low	Low	High	Unclear
Moya, 2012	Pedagogy	NA	NA	NA	NA	NA	NA	NA

Study	Intervention area	Study design	Selection bias and confounding	Spill-overs, cross-overs and contamination	Outcome reporting	Analysis reporting	Performance bias	Other risks of bias
Berlinski & Busso 2013	Pedagogy	Cluster RCT	Unclear	Unclear	Low	Low	Unclear	Low
Brooker et al. 2013 Jukes & Dubeck 2015	Pedagogy	Cluster RCT	Low	Unclear	Unclear	High	Unclear	Low
Mouton 1995	Pedagogy	RCT	Low	Unclear	Low	Low	High	Unclear
Kerwin et al., 2015	Pedagogy	Cluster RCT	Low	Low	Low	Low	High	Low
Pallante, 2013	Pedagogy	CBA	Unclear	High	Low	Unclear	Unclear	Unclear
RTI International, 2014	Pedagogy	Cluster RCT	Unclear	Low	Low	High	High	Low
Bellei, 2009	Pedagogy - extra time	Natural experiment	Unclear	Low	Low	Low	Low	High
Valenzuela, 2005	Pedagogy - extra time	CBA	Unclear	Unclear	Low	Low	Low	Unclear
Orkin, 2013	Pedagogy - extra time	CBA	Unclear	Unclear	Low	Low	Low	Unclear
Duflo et al., 2011	Pedagogy - tracking	Cluster RCT	Low	Unclear	Low	Unclear	Unclear	Unclear
Duflo et al., 2015	Pedagogy - tracking	Cluster RCT	Low	Low	Low	Low	Low	High
Loyalka et al., 2013	Providing information	Cluster RCT	High	Unclear	Low	Low	Unclear	High

Study	Intervention area	Study design	Selection bias and confounding	Spill-overs, cross-overs and contamination	Outcome reporting	Analysis reporting	Performance bias	Other risks of bias
Dinkleman & Martinez, 2011	Providing information	Cluster RCT	High	Unclear	Low	Low	Unclear	High
Nguyen, 2008	Providing information	Cluster RCT	High	Low	Low	Low	Low	Low
Jensen, 2010	Providing information	Cluster RCT	Unclear	Unclear	High	Low	Unclear	High
Glewwe et al. 2007	Providing materials	Cluster RCT	Low	Unclear	Unclear	High	High	Unclear
Glewwe et al. 2004	Providing materials	Cluster RCT	Low	Unclear	Low	Low	Unclear	Unclear
Das et al. 2013	Providing materials	Cluster RCT	Low	Low	Low	Low	Unclear	Low
Sabarwal et al. 2014	Providing materials	RCT	Low	Unclear	Low	Low	High	Unclear
Angrist et al. 2002; Angrist et al. 2006	Public private partnerships and private provision of schooling	Natural Experiment	Unclear	Unclear	Unclear	Unclear	High	Unclear
Alderman et al., 2003	Public private partnerships and private	Cluster RCT	Unclear	Unclear	Low	Unclear	High	Low

Study	Intervention area	Study design	Selection bias and confounding	Spill-overs, cross-overs and contamination	Outcome reporting	Analysis reporting	Performance bias	Other risks of bias
	provision of schooling							
Barrera-Osorio, 2007	Public private partnerships and private provision of schooling	RDD	Unclear	Unclear	Low	Unclear	Unclear	Unclear
Barrera-Osorio et al., 2011	Public private partnerships and private provision of schooling	Cluster RCT	Low	Unclear	High	Low	Unclear	Low
Barrera-Osorio, 2006	Public private partnerships and private provision of schooling	Natural experiment	Unclear	Unclear	Low	Unclear	Low	High
Muralidharan & Sundararaman, 2013	Public private partnerships and private provision of schooling	Cluster RCT	Unclear	Low	Low	Low	High	High

Study	Intervention area	Study design	Selection bias and confounding	Spill-overs, cross-overs and contamination	Outcome reporting	Analysis reporting	Performance bias	Other risks of bias
Lara et al., 2009	Public private partnerships and private provision of schooling	CBA	High	Low	Low	Low	High	Unclear
Dang et al., 2011	Public private partnerships and private provision of schooling	Cluster RCT	High	Low	Low	Low	Unclear	Low
Zhang, 2009	Public private partnerships and private provision of schooling	Natural experiment	High	Low	Low	Low	Low	High
Correa et al., 2014	Public private partnerships and private provision of schooling	CBA	Unclear	Low	Low	Low	Low	High
Saavedra Facusse, 2013	Public private partnerships and private	CBA	NA	NA	NA	NA	NA	NA

Study	Intervention area	Study design	Selection bias and confounding	Spill-overs, cross-overs and contamination	Outcome reporting	Analysis reporting	Performance bias	Other risks of bias
	provision of schooling							
Barrera-Osorio et al., 2015	Public private partnerships and private provision of schooling	Cluster RCT	Unclear	High	Low	Low	Unclear	Low
Pianto & Soares, 2004	Cash transfers	CBA	Unclear	Low	Low	Unclear	High	Low
Filmer & Schady, 2009	Cash transfers	RDD	Low	Unclear	Low	Unclear	High	Unclear
Barrera-Osorio et al., 2008	Cash transfers	Cluster RCT	Unclear	Low	Low	Low	Unclear	Low
Filmer & Schady, 2009, 2014	Cash transfers	RDD	Low	Unclear	Low	Low	Unclear	Unclear
Attanasio et al., 2004	Cash transfers	Natural Experiment	Unclear	Unclear	Low	Low	Unclear	High
Baez & Camacho, 2011	Cash transfers	RDD	Unclear	Unclear	Low	Unclear	Unclear	High
Garcia & Hill, 2010	Cash transfers	CBA	Unclear	Unclear	Low	Unclear	Unclear	Low

Study	Intervention area	Study design	Selection bias and confounding	Spill-overs, cross-overs and contamination	Outcome reporting	Analysis reporting	Performance bias	Other risks of bias
Zavakou (nd)	Cash transfers	CBA	Low	Low	Low	Low	High	Low
Ponce & Schady, 2008	Cash transfers	RDD	Low	Unclear	Low	Low	Unclear	Unclear
Schady & Arujo, 2006	Cash transfers	RCT	Low	Unclear	Low	Low	Unclear	High
Glewwe & Olinto, 2004	Cash transfers	Cluster RCT	High	Unclear	Low	Low	Unclear	High
Sparrow, 2007	Cash transfers	Natural Experiment	Unclear	Unclear	Low	Low	High	Unclear
Levy & Ohls, 2007	Cash transfers	RDD	Unclear	Unclear	Low	Low	Unclear	Low
Baird et al., 2011	Cash transfers	Cluster RCT	Low	Low	Low	Low	Unclear	Low
Macours & Vakis, 2009	Cash transfers	Cluster RCT	Low	Unclear	Unclear	Low	Unclear	Low
Maluccio & Flores, 2004	Cash transfers	Cluster RCT	High	High	Low	Low	Unclear	Low
Alam & Baez, 2010	Cash transfers	RDD	Low	Low	Low	Low	Unclear	High
Chaudhury & Parajuli, 2006	Cash transfers	RDD	Unclear	Low	Low	Low	Low	Low

Study	Intervention area	Study design	Selection bias and confounding	Spill-overs, cross-overs and contamination	Outcome reporting	Analysis reporting	Performance bias	Other risks of bias
Perez-Ribas et al., 2011	Cash transfers	Natural experiment	Unclear	Low	Low	Unclear	High	Low
Chaudhury et al., 2013	Cash transfers	Cluster RCT	Unclear	Low	Low	Low	High	Low
Eyal & Woolard, 2014	Cash transfers	Natural Experiment/ Fuzzy RDD	High	Unclear	Low	Unclear	Low	Unclear
Santana, 2008	Cash transfers	Natural Experiment	Unclear	Unclear	Low	Low	Low	Low
Khandker et al. 2003	Cash transfers	CBA	High	Unclear	Low	Low	Low	Unclear
Benhassine et al., 2014	Cash transfers	Cluster RCT	Low	High	Low	Unclear	Unclear	High
Galsasso, 2010	Cash transfers	RDD	Unclear	Unclear	Low	Unclear	Unclear	Low
Kassouf, 2012	Cash transfers	RDD	Unclear	Unclear	Low	Unclear	Unclear	Low
Ahmed et al., 2007	Cash transfers	RDD	Unclear	High	Low	Low	High	High
De Janvry et al., 2006	Cash transfers	CBA	High	High	Low	Low	Unclear	High

Study	Intervention area	Study design	Selection bias and confounding	Spill-overs, cross-overs and contamination	Outcome reporting	Analysis reporting	Performance bias	Other risks of bias
Glewwe & Kassouf, 2010	Cash transfers	CBA	High	High	Unclear	Low	Unclear	Low
Luseno, 2013	Cash transfers	Cluster RCT	Unclear	Unclear	Low	Low	Unclear	Low
Baulch, 2011	Cash transfers	CBA	Unclear	Low	Low	Unclear	Unclear	High
De Brauw & Gilligan, 2011	Cash transfers	RDD	Unclear	Low	Low	Low	Unclear	High
Fusades, 2008	Cash transfers	NA	NA	NA	NA	NA	NA	NA
Mo et al., 2011	Cash transfers	Cluster RCT	Low	Unclear	Low	Low	Unclear	Low
Robertson et al., 2013	Cash transfers	Cluster RCT	High	High	Low	Unclear	High	High
Amarante et al., 2012	Cash transfers	RDD	Unclear	Unclear	Low	Low	Unclear	High
Akresh et al., 2013	Cash transfers	Cluster RCT	Low	Low	Low	Low	Unclear	Unclear
De Brauw et al., 2014	Cash transfers	RDD	Unclear	Unclear	Low	Low	Unclear	Low
Schultz, 2004	Cash transfers	Cluster RCT	Unclear	Low	Low	Unclear	High	Low

Study	Intervention area	Study design	Selection bias and confounding	Spill-overs, cross-overs and contamination	Outcome reporting	Analysis reporting	Performance bias	Other risks of bias
Dubois, de Janvry and Sadoulet, 2011,	Cash transfers	Cluster RCT	Unclear	Low	Low	Unclear	High	Low
Behrman et al., 2000	Cash transfers	Cluster RCT	Unclear	Low	Low	Unclear	High	High
Skoufius, 2001	Cash transfers	Cluster RCT	Unclear	Low	Low	Unclear	High	Low
Raymond & Sadoulet, 2003	Cash transfers	Cluster RCT	NA	NA	NA	NA	NA	NA
Barrera-Osorio & Filmer, 2012	Cash transfers	Cluster RCT	High	Unclear	High	Low	Unclear	Low
Ferre & Sharif, 2014	Cash transfers	CBA	Unclear	Unclear	Low	Low	Unclear	High
Edmonds, 2014	Cash transfers	RCT	Low	Unclear	Low	Low	Unclear	High
Evans et al., 2014	Cash transfers	Cluster RCT	Low	High	Low	Unclear	High	Low
Pellerano et al., 2014	Cash transfers	Cluster RCT	Low	Low	Low	Low	High	Low
Benedetti et al., 2015	Cash transfers	Cluster RCT	Low	Unclear	Low	Low	Unclear	High
Heinrich et al., 2005	Cash transfers	RCT	Unclear	Unclear	Low	High	High	High

Study	Intervention area	Study design	Selection bias and confounding	Spill-overs, cross-overs and contamination	Outcome reporting	Analysis reporting	Performance bias	Other risks of bias
Barrera-Osorio et al. 2007	Reducing fees	RDD	Low	Unclear	Low	Low	High	Unclear
Evans et al., 2012	Reducing fees	Natural experiment	Low	Unclear	Unclear	Low	High	Unclear
Garlick, 2013	Reducing fees	CBA	Unclear	Low	Low	Low	High	Unclear
Gajigo, 2012	Reducing fees	CBA	High	Unclear	Low	Low	Low	Unclear
Grogan 2008	Reducing fees	RDD	Low	Unclear	Low	Low	Low	Unclear
Hidalgo et al., 2013	Reducing fees	Cluster RCT	Low	Unclear	Low	Low	High	Unclear
Kharisma, N.D.	Reducing fees	CBA	Unclear	Unclear	Low	Low	High	Unclear
Hau, 2014	Reducing fees	CBA	Unclear	Low	Low	Low	Low	Low
Edmonds, 2014	Reducing fees	RCT	Low	Unclear	Low	Low	Unclear	High
Adelman, 2015	Reducing fees	Cluster RCT	Unclear	High	Low	Unclear	Low	Unclear
Banerjee et al. 2007	Remedial education	Cluster RCT	Low	Unclear	Low	Low	High	Unclear
Cabezas et al., 2011	Remedial education	Cluster RCT	Low	Unclear	Low	Low	High	Unclear

Study	Intervention area	Study design	Selection bias and confounding	Spill-overs, cross-overs and contamination	Outcome reporting	Analysis reporting	Performance bias	Other risks of bias
Gutiérrez & Rodrigo, 2014	Remedial education	CBA	High	Unclear	Low	Low	High	Unclear
Lakshminarayana, 2013	Remedial education	Cluster RCT	High	Low	Low	High	Unclear	Low
Kazianga, 2012	School feeding	Cluster RCT	Unclear	Unclear	High	Low	Unclear	Unclear
Powell, 1998	School feeding	RCT	Unclear	Low	Low	Unclear	Unclear	Low
Jacoby, 1996	School feeding	Cluster RCT	Unclear	Low	High	Low	Unclear	Low
Buttenheim et al., 2013	School feeding	CBA	Unclear	Low	Low	Low	High	Unclear
Jayaraman & Simroth, 2015	School feeding	CBA	Unclear	Unclear	Low	Low	High	Unclear
Afridi, 2014	School feeding	CBA	High	Low	Low	Low	Low	High
Diagne et al., 2014	School feeding	Cluster RCT	High	High	Low	Low	Unclear	Unclear
Kleiman-Weiner et al., 2013	School feeding	Cluster RCT	Low	Low	Low	Low	Unclear	Low
McEwan, 2013	School feeding	RDD	Unclear	Low	Low	Low	Low	Low
Omwami et al., 2011	School feeding	Cluster RCT	High	Low	Unclear	High	Unclear	Low
Androgue, 2012	School feeding	CBA	Unclear	Unclear	Low	Low	High	Unclear

Study	Intervention area	Study design	Selection bias and confounding	Spill-overs, cross-overs and contamination	Outcome reporting	Analysis reporting	Performance bias	Other risks of bias
Tan, 1999	School feeding	Cluster RCT	Low	Unclear	Low	Low	Unclear	Unclear
Cheung & Berlin, 2014	School feeding	CBA	Unclear	Low	Low	Low	Low	High
Ismail et al., 2012	School feeding	CBA	Unclear	Low	Low	Low	Unclear	High
He, 2010	School feeding	CBA	High	Low	Low	Unclear	Unclear	High
He, 2010	School feeding	CBA	High	Unclear	Low	Unclear	Unclear	Low
Glewwe, 2014	School-based health	RCT	Low	Low	Low	Low	High	Unclear
Simeon, 1995	School-based health	RCT	Low	Unclear	Low	Low	Low	Unclear
Miguel, 2004	School-based health	Cluster RCT	Unclear	High	Low	Unclear	High	Unclear
Simwaka et al., 2009	School-based health	CBA	Unclear	Low	Low	Low	Unclear	Low
Ebenezer, 2013	School-based health	RCT	Low	Low	Low	Low	Low	Unclear
Mahawithanage, 2007	School-based health	RCT	Low	Low	Low	Low	Low	Unclear
Fernando et al., 2006	School-based health	RCT	Low	Unclear	Low	Low	Low	Unclear

Study	Intervention area	Study design	Selection bias and confounding	Spill-overs, cross-overs and contamination	Outcome reporting	Analysis reporting	Performance bias	Other risks of bias
Grigorenko et al., 2007	School-based health	Cluster RCT	High	Unclear	Low	High	Unclear	Low
Kleiman-Weiner et al., 2013	School-based health	Cluster RCT	Low	Unclear	Low	Low	High	Unclear
Luo et al., 2012	School-based health	Cluster RCT	Low	Low	Low	Low	High	Unclear
Sylvia et al., 2013	School-based health	Cluster RCT	Low	Unclear	Low	Low	High	Unclear
Brooker et al., 2015	School-based health	Cluster RCT	Low	Unclear	Low	Low	High	Unclear
Clarke et al., 2008	School-based health	Cluster RCT	High	Unclear	Low	Low	Low	Low
Watkins et al., 1996	School-based health	RCT	Low	Unclear	Low	Low	Low	Unclear
Wong et al., 2014	School-based health	Cluster RCT	Unclear	Low	Low	Low	Unclear	Unclear
Jukes et al., 2014	School-based health	Cluster RCT	Low	Unclear	Unclear	Low	Unclear	Unclear
Gertler et al. 2008, Gertler et al. 2012	School-based management (SBM)	CBA	Low	Unclear	Low	Low	High	High

Study	Intervention area	Study design	Selection bias and confounding	Spill-overs, cross-overs and contamination	Outcome reporting	Analysis reporting	Performance bias	Other risks of bias
Bando, 2010	School-based management (SBM)	RCT	Unclear	Unclear	Low	Low	High	High
Skoufias & Shapiro 2006	School-based management (SBM)	CBA	Low	Unclear	Low	Low	High	High
Murnane et al. 2006	School-based management (SBM)	CBA	Low	Unclear	Low	Low	High	High
Khattri et al. 2010	School-based management (SBM)	Natural experiment	High	Unclear	Unclear	Unclear	Unclear	Unclear
Yamauchi, 2014	School-based management (SBM)	CBA	Unclear	Unclear	Low	Low	High	High
Beasly and Huillery 2014	School-based management (SBM)	Cluster RCT	Low	Unclear	Low	Low	Low	High
Pradhan et al. 2014	School-based management (SBM)	Cluster RCT	Low	Low	Low	Low	Unclear	Low

Study	Intervention area	Study design	Selection bias and confounding	Spill-overs, cross-overs and contamination	Outcome reporting	Analysis reporting	Performance bias	Other risks of bias
Carnoy et al. 2008	School-based management (SBM)	CBA	Unclear	Unclear	Low	Unclear	High	High
Blimpo et al. 2015	School-based management (SBM)	Cluster RCT	Unclear	Low	Low	Low	Unclear	Low
San Antonio	School-based management (SBM)	RCT	Low	Unclear	Low	Low	High	High
Carneiro et al. 2015	School-based management (SBM)	Cluster RCT	Unclear	Low	Low	Low	Low	High
Aturupane et al. 2014	School-based management (SBM)	Cluster RCT	Unclear	Low	Low	Low	Low	Low
Santibanez et al. 2014	School-based management (SBM)	CBA	Unclear	Low	Low	Unclear	Unclear	High
Barrera-Osorio & Raju, 2015	Teacher Incentives	Cluster RCT	Low	Low	Low	High	High	High
Duflo et al. 2008	Teacher incentives	Cluster RCT	Unclear	Unclear	Low	Low	High	High

Study	Intervention area	Study design	Selection bias and confounding	Spill-overs, cross-overs and contamination	Outcome reporting	Analysis reporting	Performance bias	Other risks of bias
Glewwe, 2010	Teacher incentives	Cluster RCT	High	Unclear	Unclear	High	High	Unclear
Gallego et al., 2008,	Teacher incentives	RDD	NA	NA	NA	NA	NA	NA
Contreras, 2012	Teacher incentives	Natural experiment	Unclear	Low	Low	Unclear	Unclear	Low
Santibanez, 2007	Teacher incentives	RDD	Unclear	Low	Low	Unclear	Low	High
Muralidharan & Sundararaman, 2011	Teacher incentives	Cluster RCT	Low	High	Unclear	Low	Low	Unclear
Behrman, 2012	Teacher incentives	Cluster RCT	Low	Unclear	Unclear	Low	High	Low
Cueto, 2008	Teacher incentives	CBA	NA	NA	NA	NA	NA	NA
Loyalka et al., 2015	Teacher incentives	Cluster RCT	Unclear	Low	Low	Low	Unclear	Unclear
Halpin et al., 2014	Teacher Training	Cluster RCT	Unclear	Low	High	Low	High	Low

Appendix G: Targeted search guidance

For each education programme covered by one or more included full-text, a targeted search was undertaken. The aim of the targeted search was twofold.

(1) Firstly, it served as a forwards and backwards citation-tracking tool to search for new, potentially includable **impact evaluations**. These newly identified studies might evaluate programmes covered by study or set of studies that had already been included in the review. Alternatively, they may be impact evaluations of a completely new education programme that had not yet been included in the review. All newly identified impact evaluations were systematically screened for inclusion.

(2) The second purpose of the targeted search was to identify **additional data-sources** containing new information relating to an education programme described in one or more included full-texts. To be included as an 'additional data-source', documents found through the targeted search needed to provide new information relating to process, implementation or cost.

For example, if one of our included full-text studies were to describe the PROGRESA/Oportunidades cash transfer programme in Mexico, we will systematically search for additional data-sources relating specifically to that programme. These data-sources were then used, in conjunction with the included full-text studies, to address Questions 2 and 3 of the review regarding barriers and facilitators and cost effectiveness.

Targeted search methodology:

The search employed the following strategies,

1. **Contacts:** This involved identifying the study's contact/lead author and, where possible, contacts for funders and implementers of the programme described in the included study, and then emailing them using an email template to request additional data-sources.
2. **Citation tracking:** This involved conducting forward and backward citation-tracking of included studies to identify any relevant sister papers or other documentation covering the programme in question.
   Forward citation-tracking: Google Scholar searched for all articles which cited the study in question – if you find any papers which cited the study, assess them for relevance (for any papers found to be relevant, screen their references for relevance as well)
   Backward citation-tracking: all references cited in the study in question screened

3. **Search by programme name:** internet and database searches conducted using the names of programmes described in our included full-text studies.

4. **Targeted searches of funder & implementer websites:** searches conducted of the databases and websites of agencies that have implemented or funded an intervention described in one of our included full-text studies.

To be included in the review, data-sources found through the targeted search needed to be,

(1) associated with an education programme included in the review
(2) be one of the 'data-source types' listed below

1. **Impact evaluation**

   Additional documents that are linked to an included impact evaluation, for example if they describe the methodology of the impact evaluation in more detail than the main report, should also be categorised as an impact evaluation. They should be included if they provide additional information not provided in the main IE report, and/or may provide useful information for the critical appraisal and effect size calculations.
2. A **qualitative study** collecting primary data using qualitative methods of data collection and analysis, and report some information on all of the following: the research question, procedures for collecting data, sampling and recruitment, and at least two sample characteristics.
3. A **descriptive quantitative** study collecting primary data using quantitative methods of data collection and descriptive quantitative analysis and report some information on all of the following: the research question, procedures for collecting data, sampling and recruitment, and at least two sample characteristics;
4. A **mixed-methods** study that has characteristics corresponding to both qualitative and descriptive quantitative studies.
5. A **process evaluation** assessing whether a policy is being implemented as intended and what is felt to be working more or less well, and why (HM Treasury, 2011). Process evaluations may include the collection of qualitative and quantitative data from different stakeholders to cover subjective issues, such as perceptions of intervention success or more objective issues, such as how an intervention was operationalised. They might also be used to collect organisational information;
6. A **project document** providing information about planned, ongoing or completed interventions. They may describe the background and design of an intervention, or the resources available for a project for instance. As such, these documents do not typically include much analysis of primary evidence, but they provide factual information about interventions. The purpose of including them in our review is to ensure we have sufficient information about the context and interventions in included studies.

Web pages or internet pages identified through the targeted search and judged to be relevant were included with converted into a Word/PDF document where possible.

Recording the targeted search

All additional data-sources identified through the targeted search were uploaded to EPPI-reviewer, classified as one of the data-source types listed above, and linked to the studies/education programmes that they described.

Appendix H: Detailed technical content from meta-analysis

Below are the results of our meta-analysis for all interventions including assessment of heterogeneity and sensitivity analysis. Results of the sensitivity analysis are not present in the same order as the forest plots so should not be read as such. Anyone wishing more information on sensitivity analysis should contact the corresponding author.

Where results of sub-group meta-analysis were not sufficiently different from the main results for that outcome we have presented the forest plots in this appendix rather than the main report.

Also presented here are individual effect sizes not included in the meta-analysis.

4. 1. School based health (SBH)

Malaria Control

Nutrition

Random-Effects Model (k = 4; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0549 (SE = 0.0477)

tau (square root of estimated tau^2 value):   0.2342

I^2 (total heterogeneity / total variability):   97.98%

H^2 (total variability / sampling variability): 49.49

Test for Heterogeneity:

Q(df = 3) = 244.2458, p-val < .0001

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub	
0.2814	0.1208	2.3292	0.0198	0.0446	0.5182	*

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #---

> # Sensitivity Analysis

> #---

> leave1out(fit1)

	estimate	se	zval	pval	ci.lb	ci.ub	Q	Qp	tau2	I2	H2
1	0.2151	0.1385	1.5525	0.1205	-0.0564	0.4866	232.8360	0.0000	0.0551	98.5401	68.4974
2	0.2879	0.1673	1.7205	0.0853	-0.0401	0.6159	243.8640	0.0000	0.0815	99.0056	100.5582
3	0.2341	0.1571	1.4900	0.1362	-0.0738	0.5420	47.7707	0.0000	0.0694	94.7728	19.1306
4	0.4025	0.0487	8.2577	0.0000	0.3070	0.4980	4.1307	0.1268	0.0038	50.2970	2.0120

Table 1: Malaria Nutrition– Other effect sizes

Study ID	N	Unit	Country	ES	Variance	Lower CI	Upper CI
Halliday – 2nd Follow- up	3962	Student	Kenya	-0.04584	0.000505	-0.08988	-0.0018

Math

Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.1490 (SE = 0.1507)

tau (square root of estimated tau^2 value):   0.3860

I^2 (total heterogeneity / total variability):   99.13%

H^2 (total variability / sampling variability): 115.49

Test for Heterogeneity:

Q(df = 2) = 115.1079, p-val < .0001

Model Results:

estimate   se   zval   pval   ci.lb   ci.ub

  0.1623   0.2242   0.7242   0.4689 -0.2770   0.6017

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> #---

> # Sensitivity Analysis

> #---

>   leave1out(fit1)

  estimate   se   zval   pval   ci.lb   ci.ub   Q   Qp   tau2   I2   H2

1   0.2674 0.3462 0.7723 0.4399 -0.4112 0.9460 109.0769 0.0000 0.2376 99.0832 109.0769

2   0.2839 0.3296 0.8613 0.3891 -0.3622 0.9300 97.4011 0.0000 0.2151 98.9733 97.4011

3 -0.0609 0.0207 -2.9386 0.0033 -0.1016 -0.0203   0.6383 0.4243 0.0000 0.0000   1.0000

Table 2: Malaria Math– Other effect sizes

Study ID	N	Unit	Country	ES	Variance	Lower CI	Upper CI
Halliday – Grade 1 2nd Follow- up	2027	Student	Kenya	-0.18976	0.00201	-0.27764	-0.10188
Halliday- Grade 5 2nd Follow Up	2079	Student	Kenya	-0.18097	0.001956	-0.26766	-0.09429

Language

Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.2086 (SE = 0.2095)

tau (square root of estimated tau^2 value): 0.4567

I^2 (total heterogeneity / total variability): 99.55%

H^2 (total variability / sampling variability): 219.95

Test for Heterogeneity:

Q(df = 2) = 423.7535, p-val < .0001

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
  0.0283   0.2643   0.1069   0.9149  -0.4897   0.5462
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
  estimate    se    zval   pval   ci.lb   ci.ub      Q     Qp   tau2     I2      H2
1  -0.2355 0.0262 -8.9914 0.0000 -0.2868 -0.1842   1.4941 0.2216 0.0005 33.0689   1.4941
2   0.1733 0.3827  0.4528 0.6507 -0.5768  0.9234 301.7386 0.0000 0.2920 99.6686
301.7386
3   0.1471 0.4089  0.3598 0.7190 -0.6543  0.9486 343.9543 0.0000 0.3335 99.7093
343.9543
> #-----------------------------------------------------
```

Table 3: Malaria Language– Other effect sizes

Study ID	N	Unit	Country	ES	Variance	Lower CI	Upper CI
Halliday – Grade 1 2nd Follow- up	2027	Student	Kenya	-0.29769	0.000998	-0.3596	-0.23579
Halliday- Grade 5 2nd Follow Up	2079	Student	Kenya	-0.10079	0.000963	-0.16162	-0.03996

Cognitive

Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0037 (SE = 0.0054)

tau (square root of estimated tau^2 value): 0.0611

I^2 (total heterogeneity / total variability): 69.17%

H^2 (total variability / sampling variability): 3.24

Test for Heterogeneity:

Q(df = 2) = 6.5805, p-val = 0.0372

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
  0.0324   0.0424   0.7646   0.4445  -0.0507   0.1155
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
  estimate     se    zval   pval   ci.lb  ci.ub     Q     Qp   tau2     I2     H2
1   0.0593 0.0547  1.0843 0.2782 -0.0479 0.1664 3.7858 0.0517 0.0044 73.5853 3.7858
2   0.0457 0.0687  0.6661 0.5054 -0.0889 0.1803 5.6576 0.0174 0.0078 82.3247 5.6576
3  -0.0095 0.0296 -0.3215 0.7479 -0.0675 0.0485 0.2236 0.6363 0.0000  0.0000 1.0000
```

Table 4: Malaria Cognitive– Other effect sizes

Study ID	N	Unit	Country	ES	Variance	Lower CI	Upper CI
Halliday – Grade 1 2nd Follow- up	1964	Student	Kenya	-0.06138	0.002045	-0.15002	0.02725

Study ID	N	Unit	Country	ES	Variance	Lower CI	Upper CI
Halliday- Grade 5 2nd Follow Up	2050	Student	Kenya	0.103381	0.001954	0.01673	0.190032

Micronutrient

Nutrition

Random-Effects Model (k = 5; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0432 (SE = 0.0319)

tau (square root of estimated tau^2 value):   0.2079

I^2 (total heterogeneity / total variability):   95.86%

H^2 (total variability / sampling variability): 24.15

Test for Heterogeneity:

Q(df = 4) = 102.3580, p-val < .0001

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
 -0.0059   0.0950  -0.0621   0.9505  -0.1921   0.1803
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #------------------------------------------------------
> # Sensitivity Analysis
> #------------------------------------------------------
>   leave1out(fit1)
  estimate     se    zval   pval   ci.lb  ci.ub       Q     Qp   tau2      I2      H2
1  -0.0467 0.1108 -0.4215 0.6734 -0.2639 0.1705 82.5620 0.0000 0.0472 96.0918 25.5876
2  -0.0277 0.1191 -0.2327 0.8160 -0.2612 0.2058 98.1381 0.0000 0.0550 96.8176 31.4232
3  -0.0520 0.1073 -0.4852 0.6276 -0.2623 0.1582 77.9044 0.0000 0.0441 95.8609 24.1601
4   0.0726 0.0700  1.0382 0.2992 -0.0645 0.2098 29.4654 0.0000 0.0177 90.2148 10.2195
5   0.0249 0.1158  0.2152 0.8296 -0.2020 0.2518 93.5487 0.0000 0.0518 96.6311 29.6836
```

Math

Random-Effects Model (k = 4; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0005 (SE = 0.0013)

tau (square root of estimated tau^2 value):   0.0213

I^2 (total heterogeneity / total variability):   29.07%

H^2 (total variability / sampling variability): 1.41

Test for Heterogeneity:

Q(df = 3) = 4.2227, p-val = 0.2384

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub	
0.0571	0.0198	2.8882	0.0039	0.0183	0.0958	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> leave1out(fit1)

	estimate	se	zval	pval	ci.lb	ci.ub	Q	Qp	tau2	I2	H2
1	0.0533	0.0282	1.8934	0.0583	-0.0019	0.1085	4.0456	0.1323	0.0012	50.5350	2.0216
2	0.0754	0.0192	3.9190	0.0001	0.0377	0.1132	0.5977	0.7417	0.0000	0.0000	1.0000
3	0.0445	0.0211	2.1092	0.0349	0.0032	0.0859	2.4299	0.2967	0.0003	19.0451	1.2353
4	0.0556	0.0272	2.0414	0.0412	0.0022	0.1090	4.1918	0.1230	0.0012	52.3900	2.1004

Cognitive

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0011)

tau (square root of estimated tau^2 value):   0

I^2 (total heterogeneity / total variability):   0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 1) = 0.4248, p-val = 0.5146

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub
0.0073	0.0193	0.3791	0.7046	-0.0305	0.0451

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Deworming

Nutrition

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0084 (SE = 0.0245)

tau (square root of estimated tau^2 value): 0.0915

I^2 (total heterogeneity / total variability): 48.38%

H^2 (total variability / sampling variability): 1.94

Test for Heterogeneity:

Q(df = 1) = 1.9373, p-val = 0.1640

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub	
-0.2619	0.0845	-3.0988	0.0019	-0.4275	-0.0963	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Math

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0081)

tau (square root of estimated tau^2 value): 0

I^2 (total heterogeneity / total variability): 0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 1) = 0.3157, p-val = 0.5742

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub
0.0537	0.0379	1.4162	0.1567	-0.0206	0.1279

Language

Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0038)

tau (square root of estimated tau^2 value): 0

I^2 (total heterogeneity / total variability): 0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 2) = 1.0281, p-val = 0.5981

Model Results:

estimate se zval pval ci.lb ci.ub

```
 -0.0413   0.0327  -1.2633   0.2065  -0.1053   0.0228
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
>   leave1out(fit1)
  estimate     se    zval   pval   ci.lb  ci.ub      Q     Qp   tau2     I2     H2
1   0.0024 0.0541  0.0441 0.9648 -0.1036 0.1084 0.0008 0.9778 0.0000 0.0000 1.0000
2  -0.0492 0.0354 -1.3893 0.1647 -0.1186 0.0202 0.6915 0.4057 0.0000 0.0000 1.0000
3  -0.0530 0.0369 -1.4364 0.1509 -0.1254 0.0193 0.5604 0.4541 0.0000 0.0000 1.0000
```

Cognitive

```
Random-Effects Model (k = 3; tau^2 estimator: REML)
  logLik  deviance       AIC       BIC      AICc
  4.3200   -8.6400   -4.6400   -7.2537    7.3600
tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0011)
tau (square root of estimated tau^2 value):      0
I^2 (total heterogeneity / total variability):   0.00%
H^2 (total variability / sampling variability):  1.00
Test for Heterogeneity:
Q(df = 2) = 0.3891, p-val = 0.8232
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.0082   0.0195   0.4211   0.6737  -0.0301   0.0465
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> leave1out(fit1)
  estimate     se    zval   pval   ci.lb  ci.ub      Q     Qp   tau2 I2     H2
1   0.0186 0.0265  0.7035 0.4818 -0.0332 0.0704 0.0502 0.8228 0.0000   0.0000 1.0000
2   0.0086 0.0203  0.4229 0.6724 -0.0312 0.0484 0.3848 0.5350 0.0000   0.0000 1.0000
3  -0.0031 0.0269 -0.1158 0.9078 -0.0558 0.0495 0.0107 0.9175 0.0000   0.0000 1.0000
```

Table 5: Deworming Cognitive- Other Effect sizes

Study ID	N	Unit	Country	ES	Variance	Lower CI	Upper CI
Jukes – 2nd Follow-up- countingsounds	1460	Student	Philippines	-0.03387	0.00137	-0.10642	0.038679
Jukes – 2nd Follow-up- Code	1460	Student	Philippines	0.001913	0.00137	-0.07063	0.074456
Jukes – 2nd Follow- up Teacheratten	1460	Student	Philippines	0.018312	0.00137	-0.05423	0.090857
Jukes – 2nd Follow- up Dividedatten	1460	Student	Philippines	-0.03814	0.00137	-0.11068	0.034414

Attendance

Random-Effects Model (k = 4; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0231 (SE = 0.0242)

tau (square root of estimated tau^2 value): 0.1521

I^2 (total heterogeneity / total variability): 87.27%

H^2 (total variability / sampling variability): 7.86

Test for Heterogeneity:

Q(df = 3) = 18.9626, p-val = 0.0003

Model Results:

```
estimate      se    zval    pval   ci.lb   ci.ub
  0.0389  0.0861  0.4523  0.6510 -0.1298  0.2077
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
>  leave1out(fit1)
  estimate    se   zval   pval   ci.lb  ci.ub      Q     Qp   tau2      I2      H2
1   0.0384 0.1317 0.2913 0.7708 -0.2198 0.2965 11.7931 0.0027 0.0421 82.6849  5.7753
2   0.0912 0.0818 1.1149 0.2649 -0.0691 0.2515 16.3671 0.0003 0.0164 86.5486  7.4342
3   0.0131 0.0062 2.1105 0.0348  0.0009 0.0253  2.5479 0.2797 0.0000  0.0550  1.0006
4   0.0410 0.1215 0.3375 0.7358 -0.1971 0.2790 18.9491 0.0001 0.0382 92.9231 14.1305
```

Table 6: Deworming Attendance– Other effect sizes

Study ID	N	Unit	Country	ES	Variance	Lower CI	Upper CI
Miguel et al. – Y1 Follow- up	56487	Student	Kenya	0.03366	3.54E-05	0.021997	0.045324

Other Health Intervention

Attendance- Other Effect sizes

Study ID	N	Unit	Country	ES	Variance	Lower CI	Upper CI
Mahawithanage et al. – School Absenteeism (all causes)	613	Student	Sri Lanka	0	0.00653156	-0.15840	0.158403

Math- Other Effect sizes

Study ID	N	Unit	Country	ES	Variance	Lower CI	Upper CI
Glewwe- Vision Correction	9709	Student	China	-0.03438	0.000206	-0.06251	-0.00625
Sylvia- Health Incentive	5656	Student	China	-0.03611	0.000354	-0.07297	0.000745
Sylvia- Health subsidy	5656	Student	China	0.015643	0.000354	-0.02121	0.052501

Language Arts- Other Effect sizes

Study ID	N	Unit	Country	ES	Variance	Lower CI	Upper CI
Glewwe- Vision Correction	9785	Student	China	0.002527	0.000204	-0.02549	0.030549

Composite- Other Effect sizes

Study ID	N	Unit	Country	ES	Variance	Lower CI	Upper CI
Glewwe- Vision Correction	28271	Student	China	-0.0269	7.52E-05	-0.04387	-0.0988

4.2 School feeding programs

Enrolment

Random-Effects Model (k = 7; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0624 (SE = 0.0366)

tau (square root of estimated tau^2 value):   0.2498

I^2 (total heterogeneity / total variability): 98.83%

H^2 (total variability / sampling variability): 85.31

Test for Heterogeneity:

Q(df = 6) = 320.4760, p-val < .0001

Model Results:

```
estimate      se    zval     pval    ci.lb    ci.ub
  0.1393   0.0951   1.4653   0.1428  -0.0470   0.3257
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
-----------------------------------------------------
Sensitivity Analysis
-----------------------------------------------------
  estimate    se   zval   pval   ci.lb  ci.ub       Q     Qp   tau2     I2      H2
1   0.1482 0.1122 1.3205 0.1867 -0.0718 0.3681 320.3816 0.0000 0.0746 99.0099 100.9992
2   0.1501 0.1120 1.3401 0.1802 -0.0694 0.3695 320.4539 0.0000 0.0743 99.0051 100.5084
3   0.1620 0.1095 1.4786 0.1393 -0.0527 0.3767 306.4533 0.0000 0.0710 98.8797  89.2604
4   0.1650 0.1086 1.5193 0.1287 -0.0478 0.3778 302.4477 0.0000 0.0698 98.8954  90.5312
5   0.1408 0.1124 1.2527 0.2103 -0.0795 0.3610 318.9803 0.0000 0.0750 99.1215 113.8284
6   0.1639 0.1090 1.5042 0.1325 -0.0497 0.3774 301.7783 0.0000 0.0703 98.8676  88.3081
7   0.0412 0.0227 1.8175 0.0691 -0.0032 0.0857  19.9517 0.0013 0.0023 77.2642   4.3984
-----------------------------------------------------
```

Table A4.2 a: Enrolment – Other effect sizes

Study ID	N	Unit	Country	ES	Variance	Lower CI	Upper CI
Kazianga – Take home rations arm	2743	Student	Burkina Faso	0.114561	0.00073	0.061593	0.167529
Kazianga – Take home rations arm girls	1372	Student	Burkina Faso	0.178858	0.001464	0.103876	0.253841
Kazianga – Take home rations arm boys	1372	Student	Burkina Faso	0.115703	0.00146	0.040808	0.190599
Buttenheim – In school and take-	3016	Student	Laos	0.044511	0.000663	-0.00597	0.09499

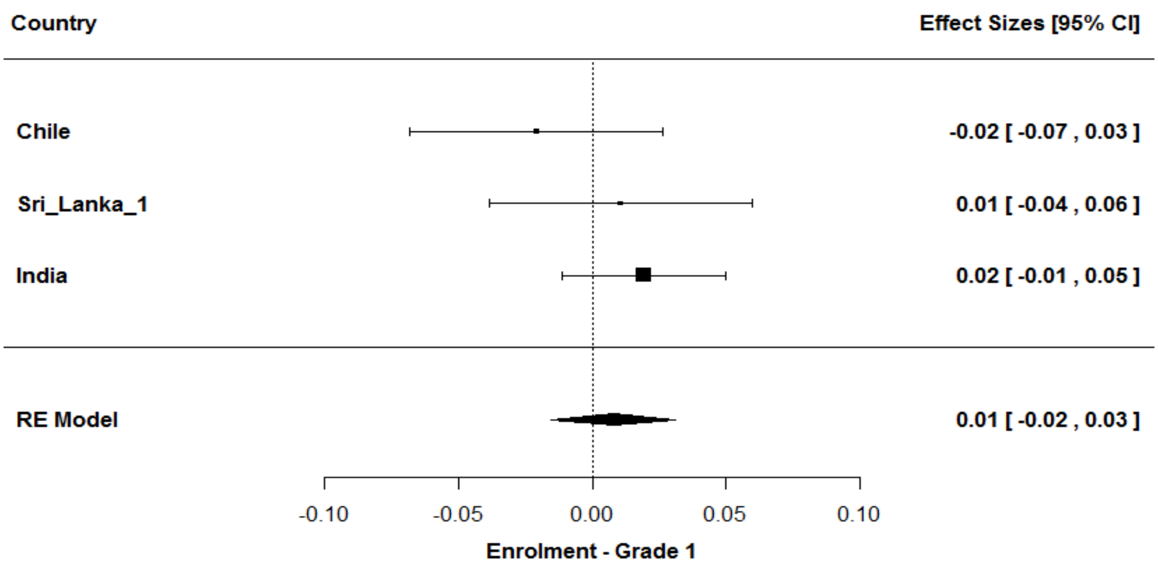
home rations arm							
Buttenheim – take-home rations arm	282 5	Student	Laos	0.10420 3	0.00070 9	0.05201 7	0.156389
Buttenheim – in-school feeding, aged 6-10 children	164 2	Student	Laos	0.10170 4	0.00122	0.03325 5	0.170153
Buttenheim – In school and take-home rations arm, aged 6-10 children	181 2	Student	Laos	0.07492 1	0.00110 5	0.00978 1	0.14006
Buttenheim – take-home rations arm, aged 6-10 children	170 1	Student	Laos	0.09310 6	0.00117 7	0.02586 2	0.16035
Buttenheim – in-school feeding, aged 11-14 children	103 9	Student	Laos	0.07608 2	0.00192 6	-0.00994	0.162107
Buttenheim – In school and take-home rations arm, aged 11-14 children	108 5	Student	Laos	0.01156 5	0.00184 3	-0.07259	0.095716
Buttenheim – take-home rations arm, aged 11-14 children	103 9	Student	Laos	0.08261 8	0.00192 7	-0.00341	0.168648
Buttenheim – In school and take-home rations arm, aged 6-10 boys	103 9	Student	Laos	0.12604	0.00192 9	0.03996 2	0.212118
Buttenheim – take-home rations arm, aged 6-10 boys	103 9	Student	Laos	0.11551 8	0.00192 8	0.02945 4	0.201583

Buttenheim – In school and take-home rations arm, aged 11-14 boys	1039	Student	Laos	0.00715	0.001925	-0.07884	0.093143
Buttenheim – take-home rations arm, aged 11-14 boys	1039	Student	Laos	-0.00958	0.001925	-0.09557	0.076417
Buttenheim – In school and take-home rations arm, aged 6-10 girls	1039	Student	Laos	-0.00932	0.001925	-0.09532	0.076671
Buttenheim – take-home rations arm, aged 6-10 girls	1039	Student	Laos	0.105293	0.001928	0.019241	0.191346
Buttenheim – In school and take-home rations arm, aged 11-14 girls	1039	Student	Laos	0.018315	0.001925	-0.06768	0.10431
Buttenheim – take-home rations arm, aged 11-14 girls	1039	Student	Laos	0.15774	0.001931	0.071613	0.243867
Afridi – full sample	1039	School	India	4.116535	0.006002	3.964684	4.268386
Cheung – group 1, year 2	1053	School	Cambodia	-0.04453	0.0019	-0.12996	0.040903
Cheung – group 1, year 3	1053	School	Cambodia	-0.09307	0.001901	-0.17853	-0.0076
Cheung – group 1, year 4	1053	School	Cambodia	-0.14429	0.001904	-0.22982	-0.05876
Cheung – group 2, year 1	1706	School	Cambodia	0.073175	0.001173	0.006044	0.140307
Cheung – group 2, year 2	1706	School	Cambodia	0.097647	0.001174	0.030498	0.164796
Cheung – group 2, year 3	1706	School	Cambodia	0.083603	0.001173	0.016465	0.150742

Cheung – group 3, year 1	1934	School	Cambodia	0.188777	0.001039	0.125607	0.251946
Cheung – group 3, year 2	1934	School	Cambodia	0.180613	0.001038	0.117455	0.243771
He – standard treatment grade 6	3995	Grade	Sri_Lanka_1	-0.01799	0.000501	-0.06184	0.025866
He – standard treatment grade 7	3995	Grade	Sri_Lanka_1	-0.04667	0.000501	-0.09053	-0.00281
He – standard treatment grade 8	3995	Grade	Sri_Lanka_1	-0.00838	0.000501	-0.05223	0.035477
He – standard treatment grade 9	3995	Grade	Sri_Lanka_1	0.006952	0.000501	-0.0369	0.050807
He – welfare treatment arm	3995	Grade	Sri_Lanka_2	0.101728	0.000501	0.057846	0.145611
He – welfare treatment grade 6	3995	Grade	Sri_Lanka_2	0.096919	0.000501	0.053039	0.140799
He – welfare treatment grade 7	3995	Grade	Sri_Lanka_2	0.0666	0.000501	0.022734	0.110467
He – welfare treatment grade 8	3995	Grade	Sri_Lanka_2	0.09076	0.000501	0.046883	0.134637
He – welfare treatment grade 9	3995	Grade	Sri_Lanka_2	0.090322	0.000501	0.046445	0.134199

Enrolment - Sub-group analysis

Figure A4.2 a: Enrolment Grade 1 children



Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0000 (SE = 0.0005)

tau (square root of estimated tau^2 value):   0.0041

I^2 (total heterogeneity / total variability):   3.57%

H^2 (total variability / sampling variability): 1.04

Test for Heterogeneity:

Q(df = 2) = 1.9692, p-val = 0.3736

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub
0.0079	0.0119	0.6608	0.5087	-0.0155	0.0312

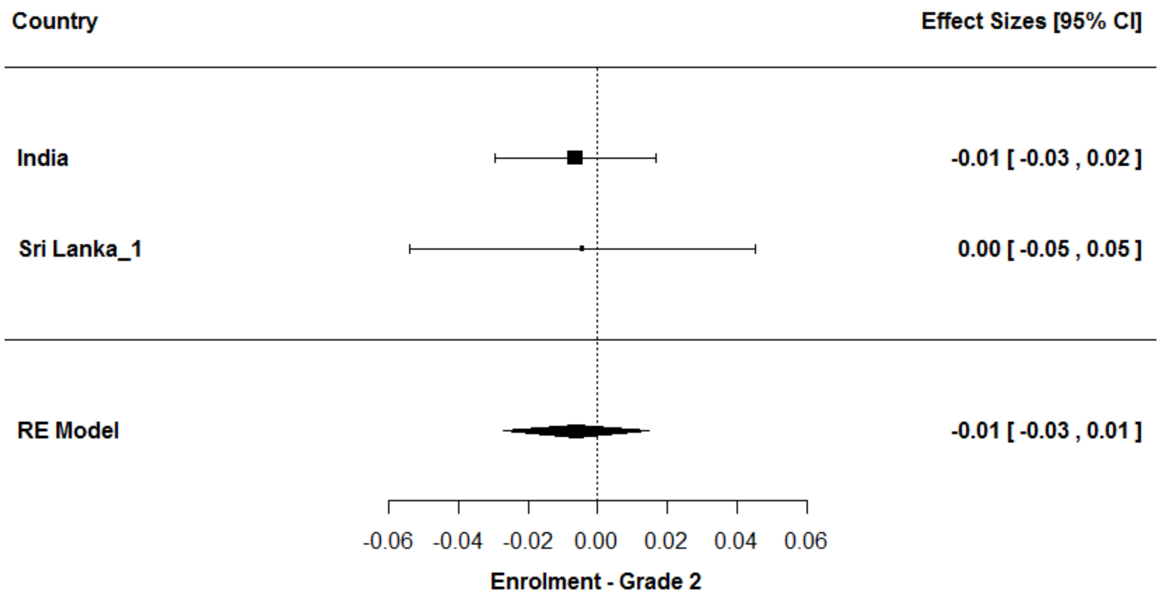
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sensitivity Analysis

----

	estimate	se	zval	pval	ci.lb	ci.ub	Q	Qp	tau2	I2	H2
1	-0.0057	0.0174	-0.3264	0.7441	-0.0397	0.0284	0.8310	0.3620	0.0000	0.0000	1.0000
2	0.0169	0.0133	1.2730	0.2030	-0.0091	0.0429	0.0825	0.7739	0.0000	0.0000	1.0000
3	0.0034	0.0197	0.1714	0.8639	-0.0352	0.0419	1.9546	0.1621	0.0004	48.8384	1.9546

Figure A4.2 b: Enrolment Grade 2 children



Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0005)

tau (square root of estimated tau^2 value): 0

I^2 (total heterogeneity / total variability): 0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

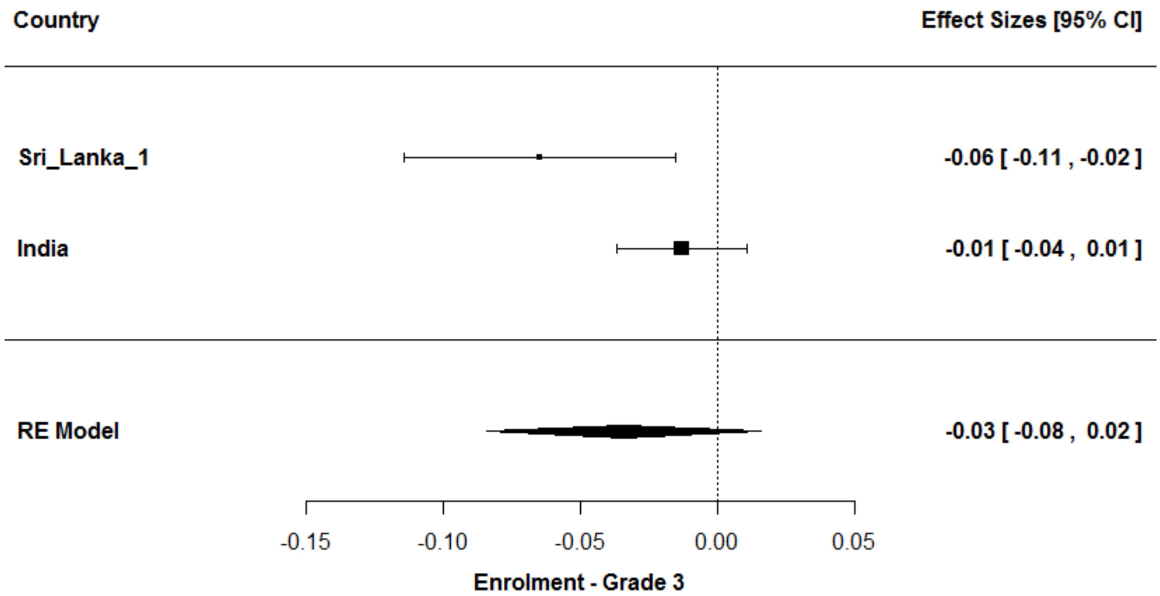
Q(df = 1) = 0.0048, p-val = 0.9447

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub
-0.0061	0.0106	-0.5695	0.5690	-0.0269	0.0148

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Figure A4.2 c: Enrolment Grade 3 children



Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0010 (SE = 0.0019)

tau (square root of estimated tau^2 value): 0.0309

I^2 (total heterogeneity / total variability): 70.85%

H^2 (total variability / sampling variability): 3.43

Test for Heterogeneity:

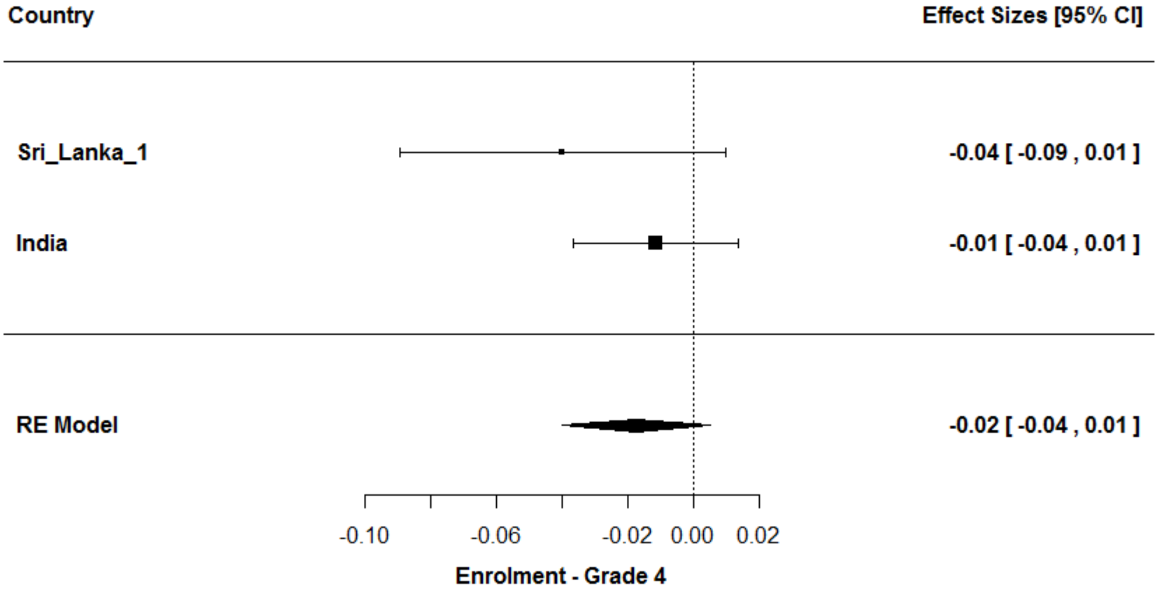
Q(df = 1) = 3.4303, p-val = 0.0640

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub
-0.0342	0.0255	-1.3401	0.1802	-0.0841	0.0158

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figure A4.2 d: Enrolment Grade 4 children



Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0000 (SE = 0.0006)

tau (square root of estimated tau^2 value): 0.0016

I^2 (total heterogeneity / total variability): 0.65%

H^2 (total variability / sampling variability): 1.01

Test for Heterogeneity:

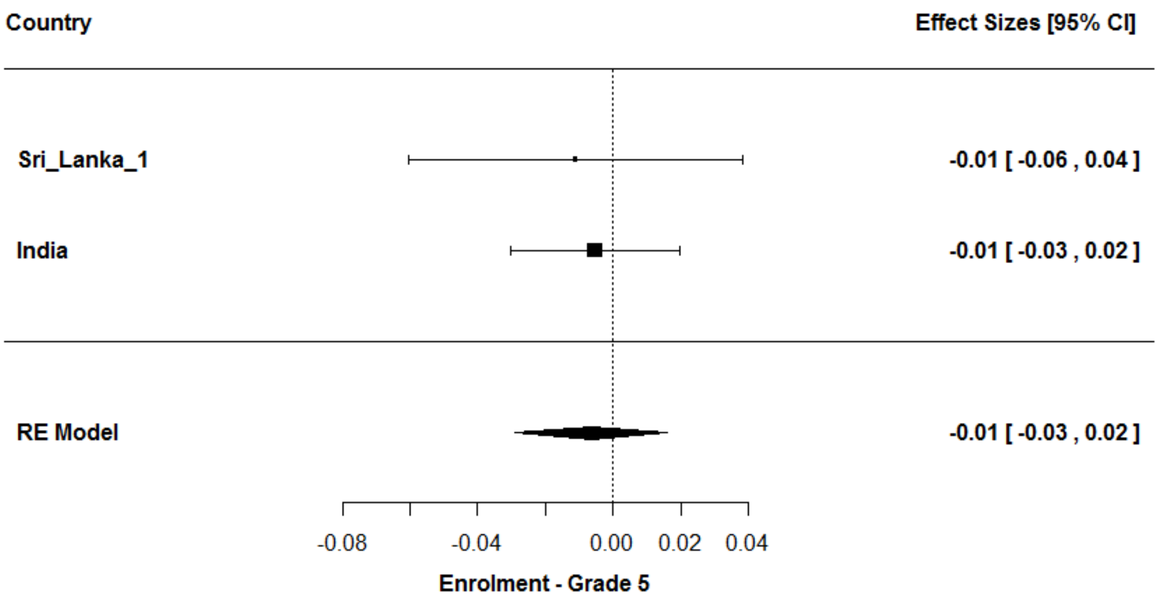
Q(df = 1) = 1.0066, p-val = 0.3157

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub
-0.0175	0.0115	-1.5185	0.1289	-0.0400	0.0051

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Figure A4.2 e: Enrolment Grade 5 children



Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0006)

tau (square root of estimated tau^2 value):   0

I^2 (total heterogeneity / total variability):   0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

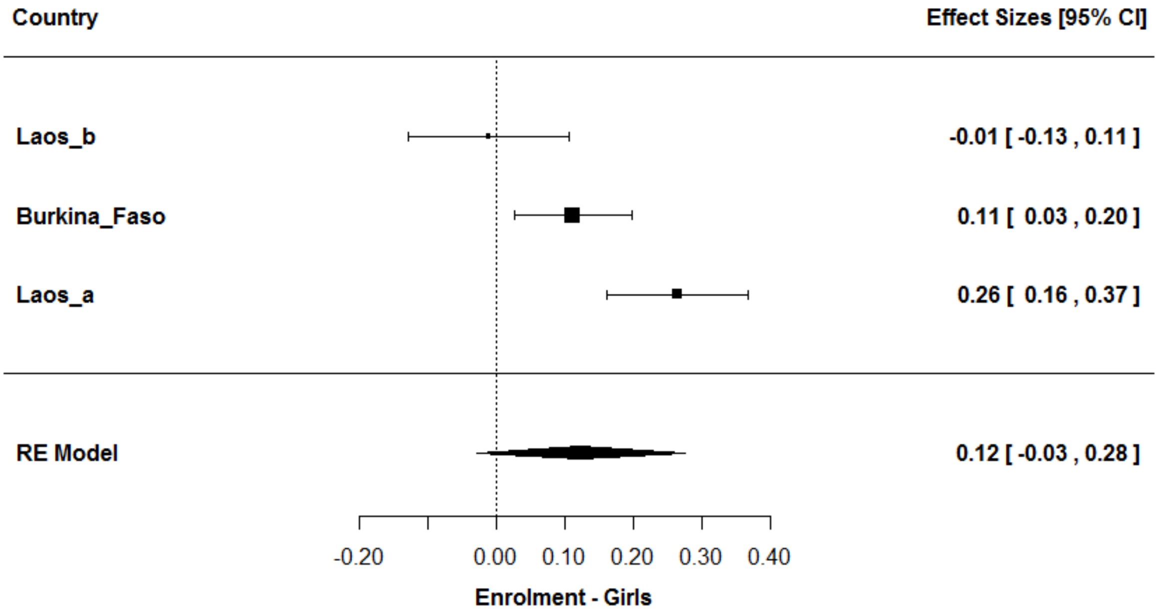
Q(df = 1) = 0.0410, p-val = 0.8395

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub
-0.0064	0.0114	-0.5631	0.5734	-0.0288	0.0160

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Figure A4.2 f: Enrolment - Girls



Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0155 (SE = 0.0182)

tau (square root of estimated tau^2 value): 0.1244

I^2 (total heterogeneity / total variability): 85.32%

H^2 (total variability / sampling variability): 6.81

Test for Heterogeneity:

Q(df = 2) = 12.2632, p-val = 0.0022

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub
0.1227	0.0779	1.5752	0.1152	-0.0300	0.2753

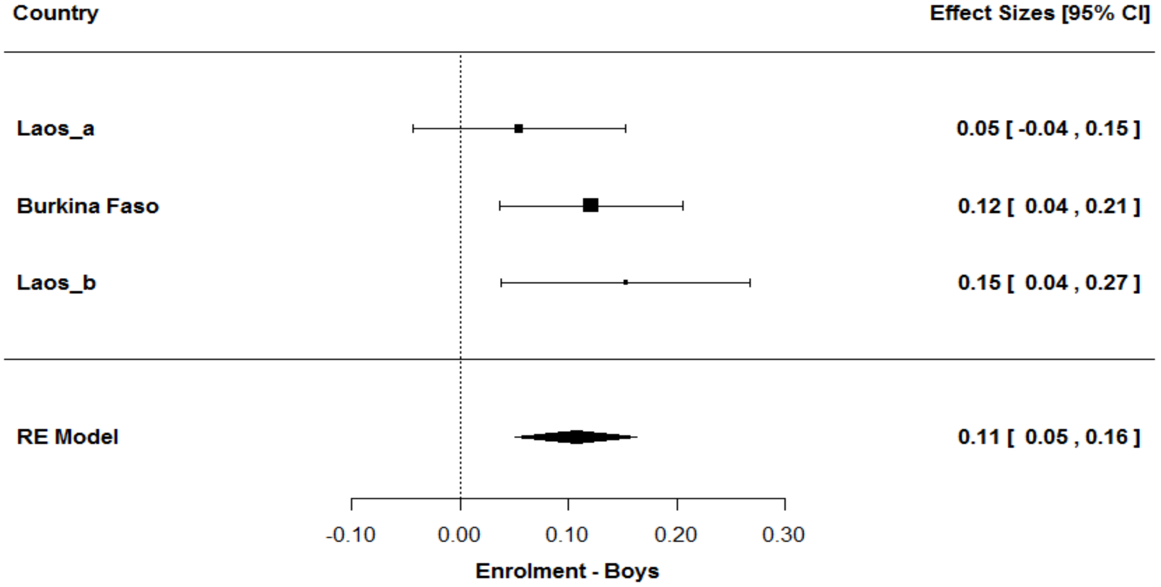
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Sensitivity Analysis

----

	estimate	se	zval	pval	ci.lb	ci.ub	Q	Qp	tau2	I2	H2
1	0.1273	0.1376	0.9249	0.3550	-0.1424	0.3970	11.9647	0.0005	0.0347	91.6421	11.9647
2	0.0565	0.0612	0.9234	0.3558	-0.0634	0.1764	2.7657	0.0963	0.0048	63.8433	2.7657
3	0.1846	0.0760	2.4279	0.0152	0.0356	0.3335	4.9583	0.0260	0.0092	79.8320	4.9583

Figure A4.2 g: Enrolment - Boys



Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0025)

tau (square root of estimated tau^2 value): 0

I^2 (total heterogeneity / total variability): 0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 2) = 1.8072, p-val = 0.4051

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub	
0.1071	0.0285	3.7533	0.0002	0.0512	0.1631	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sensitivity Analysis

----

	estimate	se	zval	pval	ci.lb	ci.ub	Q	Qp	tau2	I2	H2
1	0.0992	0.0489	2.0265	0.0427	0.0033	0.1951	1.6228	0.2027	0.0019	38.3783	1.6228
2	0.1322	0.0347	3.8103	0.0001	0.0642	0.2002	0.1932	0.6602	0.0000	0.0000	1.0000
3	0.0929	0.0328	2.8319	0.0046	0.0286	0.1571	1.0058	0.3159	0.0000	0.5731	1.0058

Attendance

Random-Effects Model (k = 6; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0048 (SE = 0.0038)

tau (square root of estimated tau^2 value): 0.0690

I^2 (total heterogeneity / total variability): 84.58%

H^2 (total variability / sampling variability): 6.49

Test for Heterogeneity:

Q(df = 5) = 23.0011, p-val = 0.0003

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
  0.0932   0.0318   2.9286   0.0034   0.0308   0.1555      **
```

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Sensitivity Analysis

```
  estimate     se   zval   pval  ci.lb  ci.ub      Q     Qp   tau2     I2     H2
1   0.0997 0.0397 2.5108 0.0120 0.0219 0.1775 22.9926 0.0001 0.0064 86.4854 7.3994
2   0.0729 0.0286 2.5488 0.0108 0.0168 0.1289 14.6679 0.0054 0.0030 79.7822 4.9461
3   0.1130 0.0284 3.9807 0.0001 0.0573 0.1686 14.7556 0.0052 0.0027 75.1571 4.0253
4   0.1032 0.0396 2.6059 0.0092 0.0256 0.1808 21.5610 0.0002 0.0062 81.8202 5.5006
5   0.0796 0.0345 2.3047 0.0212 0.0119 0.1472 16.0602 0.0029 0.0047 84.6470 6.5134
6   0.0911 0.0375 2.4315 0.0150 0.0177 0.1646 22.3075 0.0002 0.0060 88.7711 8.9056
```

Table A4.2 b: Attendance– Other effect sizes

ID	N	Unit	Country	ES	Variance	Lower CI	Upper
Kazianga – Take home rations arm	1916	Student	Burkina Faso	0.103814	0.00073	0.050854	0.156774
Kazianga – In-school meals Girls	1369	Student	Burkina Faso	0.068279	0.001462	-0.00666	0.143216
Kazianga – Take home rations Girls	1372	Student	Burkina Faso	0.116154	0.00146	0.041257	0.19105

Kazianga – In-school meals Boys	1369	Student	Burkina Faso	0.11796	0.001463	0.042979	0.19294
Kazianga – Take home rations Boys	1372	Student	Burkina Faso	0.139187	0.001461	0.064263	0.214111
Afridi – full sample	1591	Student	India	0.111425	0.001259	0.041879	0.180971

Dropout

Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0053 (SE = 0.0063)

tau (square root of estimated tau^2 value):   0.0725

I^2 (total heterogeneity / total variability):   83.52%

H^2 (total variability / sampling variability): 6.07

Test for Heterogeneity:

Q(df = 2) = 11.2250, p-val = 0.0037

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
 -0.0566   0.0459  -1.2343   0.2171  -0.1465   0.0333
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
-----------------------------------------------------
Sensitivity Analysis
----------------------------------------------------
  estimate     se    zval   pval   ci.lb  ci.ub      Q     Qp   tau2      I2      H2
1  -0.0738 0.0755 -0.9774 0.3284 -0.2219 0.0742 10.5454 0.0012 0.0103 90.5172 10.5454
2  -0.0869 0.0629 -1.3799 0.1676 -0.2102 0.0365  6.7278 0.0095 0.0067 85.1363  6.7278
3  -0.0113 0.0213 -0.5301 0.5961 -0.0531 0.0305  0.3462 0.5563 0.0000  0.0000  1.0000
-----------------------------------------------------
```

Table A4.2 c: Dropout – Other effect sizes

Study ID	N	Unit	Country	ES	Variance
Tan – School feeding + Parent-Teacher partnership arm	2071	Student	Philippines	-0.06153	0.000966

Completion (repetition)

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0004)

tau (square root of estimated tau^2 value): 0

I^2 (total heterogeneity / total variability): 0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 1) = 0.0298, p-val = 0.8629

Model Results:

```
estimate      se    zval     pval    ci.lb    ci.ub
 -0.0127   0.0099  -1.2816   0.2000  -0.0321   0.0067
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
  estimate     se    zval   pval   ci.lb  ci.ub     Q     Qp   tau2    I2    H2
1  -0.0156 0.0197 -0.7933 0.4276 -0.0543 0.0230 0.0000 1.0000 0.0000   NaN   NaN
2  -0.0117 0.0115 -1.0213 0.3071 -0.0342 0.0108 0.0000 1.0000 0.0000   NaN   NaN
```

Cognitive Scores

Random-Effects Model (k = 7; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0163 (SE = 0.0118)

tau (square root of estimated tau^2 value): 0.1277

I^2 (total heterogeneity / total variability): 83.34%

H^2 (total variability / sampling variability): 6.00

Test for Heterogeneity:

Q(df = 6) = 29.5257, p-val < .0001

Model Results:

```
estimate      se    zval     pval    ci.lb    ci.ub
  0.1099   0.0541   2.0312   0.0422   0.0039   0.2159       *
---
```

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Sensitivity Analysis

----

	estimate	se	zval	pval	ci.lb	ci.ub	Q	Qp	tau2	I2	H2
1	0.1379	0.0532	2.5926	0.0095	0.0337	0.2422	24.1614	0.0002	0.0131	81.2944	5.3460
2	0.1252	0.0635	1.9705	0.0488	0.0007	0.2497	26.6780	0.0001	0.0195	83.0904	5.9138
3	0.0721	0.0430	1.6750	0.0939	-0.0123	0.1565	16.3646	0.0059	0.0075	71.2830	3.4823
4	0.0884	0.0593	1.4900	0.1362	-0.0279	0.2047	20.7096	0.0009	0.0166	82.7369	5.7927
5	0.1041	0.0632	1.6481	0.0993	-0.0197	0.2280	28.8664	0.0000	0.0201	86.9510	7.6634
6	0.1179	0.0631	1.8674	0.0618	-0.0058	0.2416	29.3890	0.0000	0.0200	86.9256	7.6485
7	0.1253	0.0635	1.9749	0.0483	0.0009	0.2497	26.5903	0.0001	0.0194	83.0584	5.9026

> #---

Maths test scores

Random-Effects Model (k = 10; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0210 (SE = 0.0110)

tau (square root of estimated tau^2 value): 0.1450

I^2 (total heterogeneity / total variability): 92.63%

H^2 (total variability / sampling variability): 13.58

Test for Heterogeneity:

Q(df = 9) = 110.7828, p-val < .0001

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub	
0.0973	0.0484	2.0109	0.0443	0.0025	0.1921	*

\---

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Sensitivity Analysis

----

	estimate	se	zval	pval	ci.lb	ci.ub	Q	Qp	tau2	I2	H2
1	0.0965	0.0543	1.7763	0.0757	-0.0100	0.2029	110.2191	0.0000	0.0240	93.1575	14.6144
2	0.1197	0.0480	2.4933	0.0127	0.0256	0.2137	95.9834	0.0000	0.0184	91.9693	12.4522

3	0.0902	0.0534	1.6897	0.0911	-0.0144	0.1948	108.0414	0.0000	0.0233	93.5207	15.4338
4	0.0873	0.0528	1.6529	0.0983	-0.0162	0.1909	105.7795	0.0000	0.0227	93.3517	15.0415
5	0.1047	0.0538	1.9470	0.0515	-0.0007	0.2100	108.6111	0.0000	0.0235	93.0114	14.3090
6	0.1137	0.0513	2.2177	0.0266	0.0132	0.2142	95.2732	0.0000	0.0211	92.3010	12.9886
7	0.1110	0.0522	2.1250	0.0336	0.0086	0.2134	100.5319	0.0000	0.0220	92.5889	13.4932
8	0.0963	0.0543	1.7734	0.0762	-0.0101	0.2027	110.1668	0.0000	0.0240	93.1916	14.6877
9	0.0572	0.0333	1.7172	0.0859	-0.0081	0.1225	40.0861	0.0000	0.0078	82.4579	5.7006
10	0.0949	0.0524	1.8098	0.0703	-0.0079	0.1976	110.5844	0.0000	0.0230	93.8361	16.2235

Language

Random-Effects Model (k = 8; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0112 (SE = 0.0075)

tau (square root of estimated tau^2 value): 0.1057

I^2 (total heterogeneity / total variability): 80.49%

H^2 (total variability / sampling variability): 5.13

Test for Heterogeneity:

Q(df = 7) = 36.8236, p-val < .0001

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub	
0.0912	0.0421	2.1662	0.0303	0.0087	0.1737	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sensitivity Analysis

----

	estimate	se	zval	pval	ci.lb	ci.ub	Q	Qp	tau2	I2	H2
1	0.0915	0.0488	1.8738	0.0610	-0.0042	0.1873	36.8057	0.0000	0.0136	82.9660	5.8706
2	0.1110	0.0434	2.5539	0.0107	0.0258	0.1961	28.5802	0.0001	0.0101	78.4137	4.6326
3	0.0865	0.0484	1.7865	0.0740	-0.0084	0.1813	36.0666	0.0000	0.0133	82.7657	5.8024
4	0.0943	0.0488	1.9327	0.0533	-0.0013	0.1899	36.7447	0.0000	0.0135	82.9180	5.8541
5	0.0593	0.0344	1.7240	0.0847	-0.0081	0.1267	17.2100	0.0085	0.0053	65.5338	2.9014
6	0.1157	0.0396	2.9198	0.0035	0.0380	0.1934	24.6182	0.0004	0.0080	74.5568	3.9303
7	0.0892	0.0484	1.8423	0.0654	-0.0057	0.1840	36.6428	0.0000	0.0134	83.3897	6.0204

8   0.0805 0.0450 1.7881 0.0738 -0.0077 0.1688 35.2831 0.0000 0.0117 82.7964 5.8127

Composite test scores

Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0219 (SE = 0.0268)

tau (square root of estimated tau^2 value):   0.1480

I^2 (total heterogeneity / total variability):   92.15%

H^2 (total variability / sampling variability): 12.73

Test for Heterogeneity:

Q(df = 2) = 20.2284, p-val < .0001

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub
0.1421	0.0941	1.5100	0.1311	-0.0424	0.3266

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Sensitivity Analysis

----

	estimate	se	zval	pval	ci.lb	ci.ub	Q	Qp	tau2	I2	H2
1	0.1474	0.1776	0.8299	0.4066	-0.2008	0.4956	7.0701	0.0078	0.0551	85.8559	7.0701
2	0.2150	0.0759	2.8312	0.0046	0.0661	0.3638	1.7275	0.1887	0.0066	42.1132	1.7275
3	0.0814	0.0905	0.8989	0.3687	-0.0961	0.2589	16.1652	0.0001	0.0154	93.8139	16.1652

4.3 Merit based scholarships

Teacher Attendance

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0130 (SE = 0.0238)

tau (square root of estimated tau^2 value):   0.1141

I^2 (total heterogeneity / total variability):   77.33%

H^2 (total variability / sampling variability): 4.41

Test for Heterogeneity:

Q(df = 1) = 4.4110, p-val = 0.0357

```
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.1390   0.0917   1.5156   0.1296  -0.0407   0.3187
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
>   leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub     Q     Qp   tau2    I2    H2
1   0.0450 0.0650 0.6916 0.4892 -0.0825 0.1725 0.0000 1.0000 0.0000   NaN   NaN
2   0.2285 0.0583 3.9183 0.0001  0.1142 0.3427 0.0000 1.0000 0.0000   NaN   NaN
```

Student Attendance

```
Random-Effects Model (k = 4; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.0037 (SE = 0.0045)
tau (square root of estimated tau^2 value):      0.0608
I^2 (total heterogeneity / total variability):   67.02%
H^2 (total variability / sampling variability):  3.03
Test for Heterogeneity:
Q(df = 3) = 9.5368, p-val = 0.0229
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.0089   0.0373   0.2382   0.8117  -0.0642   0.0820
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
> leave1out(fit1)
  estimate     se    zval   pval   ci.lb  ci.ub     Q     Qp   tau2     I2     H2
1  -0.0192 0.0377 -0.5102 0.6099 -0.0930 0.0546 4.1469 0.1258 0.0022 51.8642 2.0775
2   0.0383 0.0352  1.0889 0.2762 -0.0307 0.1073 3.7796 0.1511 0.0018 47.2588 1.8961
```

3 -0.0062 0.0461 -0.1336 0.8938 -0.0966 0.0843 8.3071 0.0157 0.0047 74.4243 3.9100

4 0.0222 0.0484 0.4586 0.6465 -0.0727 0.1171 8.4903 0.0143 0.0052 74.9361 3.9898

Maths

Random-Effects Model (k = 10; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0162 (SE = 0.0087)

tau (square root of estimated tau^2 value): 0.1272

I^2 (total heterogeneity / total variability): 87.66%

H^2 (total variability / sampling variability): 8.10

Test for Heterogeneity:

Q(df = 9) = 67.9728, p-val < .0001

Model Results:

estimate se zval pval ci.lb ci.ub

0.1137 0.0430 2.6457 0.0082 0.0295 0.1979 **

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #---

> # Sensitivity Analysis

> #---

> leave1out(fit1)

	estimate	se	zval	pval	ci.lb	ci.ub	Q	Qp	tau2	I2	H2
1	0.1180	0.0479	2.4631	0.0138	0.0241	0.2118	67.0672	0.0000	0.0183	88.8716	8.9860
2	0.0962	0.0439	2.1915	0.0284	0.0102	0.1823	55.7357	0.0000	0.0150	86.7721	7.5598
3	0.1201	0.0475	2.5260	0.0115	0.0269	0.2132	66.2473	0.0000	0.0180	88.8361	8.9574
4	0.1012	0.0461	2.1978	0.0280	0.0110	0.1915	61.8292	0.0000	0.0168	87.9718	8.3138
5	0.1080	0.0477	2.2625	0.0237	0.0144	0.2015	66.7790	0.0000	0.0182	88.7972	8.9263
6	0.1161	0.0481	2.4156	0.0157	0.0219	0.2103	67.6475	0.0000	0.0185	88.9509	9.0505
7	0.0919	0.0411	2.2350	0.0254	0.0113	0.1725	51.9120	0.0000	0.0130	85.2969	6.8013
8	0.1288	0.0449	2.8672	0.0041	0.0407	0.2168	58.9906	0.0000	0.0159	87.5143	8.0092
9	0.1414	0.0363	3.8927	0.0001	0.0702	0.2125	40.8157	0.0000	0.0096	81.1348	5.3008
10	0.1150	0.0481	2.3890	0.0169	0.0206	0.2093	67.8593	0.0000	0.0185	88.9782	9.0729

Language

Random-Effects Model (k = 3; tau^2 estimator: REML)

```
tau^2 (estimated amount of total heterogeneity): 0.0079 (SE = 0.0099)
tau (square root of estimated tau^2 value):      0.0889
I^2 (total heterogeneity / total variability):   79.58%
H^2 (total variability / sampling variability):  4.90
Test for Heterogeneity:
Q(df = 2) = 9.6847, p-val = 0.0079
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.0412   0.0576   0.7150   0.4746  -0.0717   0.1540
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
>   leave1out(fit1)
  estimate     se    zval   pval   ci.lb  ci.ub      Q     Qp   tau2     I2     H2
1   0.0775 0.0779  0.9948 0.3198 -0.0751 0.2300 5.9442 0.0148 0.0101 83.1769 5.9442
2  -0.0154 0.0316 -0.4872 0.6261 -0.0773 0.0465 0.2419 0.6228 0.0000  0.0000 1.0000
3   0.0621 0.0934  0.6648 0.5062 -0.1210 0.2451 8.4742 0.0036 0.0154 88.1995 8.4742
```

Composite Score

```
Random-Effects Model (k = 7; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.0043 (SE = 0.0046)
tau (square root of estimated tau^2 value):      0.0658
I^2 (total heterogeneity / total variability):   54.61%
H^2 (total variability / sampling variability):  2.20
Test for Heterogeneity:
Q(df = 6) = 13.2394, p-val = 0.0394
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.0995   0.0337   2.9526   0.0032   0.0334   0.1655      **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> #----------------------------------------------------

> # Sensitivity Analysis

> #----------------------------------------------------

>  leave1out(fit1)

  estimate     se   zval   pval  ci.lb  ci.ub       Q     Qp   tau2     I2     H2

1   0.1134 0.0364 3.1185 0.0018 0.0421 0.1847 10.9415 0.0526 0.0043 54.2358 2.1851

2   0.0731 0.0255 2.8659 0.0042 0.0231 0.1232  5.2006 0.3919 0.0003  7.4674 1.0807

3   0.1068 0.0389 2.7475 0.0060 0.0306 0.1830 12.5789 0.0277 0.0054 60.0191 2.5012

4   0.1038 0.0395 2.6254 0.0087 0.0263 0.1813 13.0037 0.0233 0.0057 61.2733 2.5822

5   0.1107 0.0376 2.9435 0.0032 0.0370 0.1844 11.7288 0.0387 0.0048 57.2131 2.3372

6   0.0845 0.0359 2.3549 0.0185 0.0142 0.1548 10.6505 0.0588 0.0041 53.0226 2.1287

7   0.1035 0.0388 2.6677 0.0076 0.0275 0.1796 13.0420 0.0230 0.0056 61.6766 2.6094
```

Completion

```
Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0011 (SE = 0.0143)

tau (square root of estimated tau^2 value):      0.0331

I^2 (total heterogeneity / total variability):   10.87%

H^2 (total variability / sampling variability):  1.12

Test for Heterogeneity:

Q(df = 1) = 1.1219, p-val = 0.2895

Model Results:

estimate      se     zval     pval    ci.lb    ci.ub

  0.3227   0.0711   4.5406   <.0001   0.1834   0.4620      ***

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

5.1. Programmes to reduce or eliminate user fees

5.1.1 Sensitivity analysis

Enrolment

```
> #----------------------------------------------------

> # Sensitivity Analysis

> #----------------------------------------------------

>  leave1out(fit1)
```

```
  estimate     se   zval   pval   ci.lb  ci.ub        Q      Qp   tau2      I2       H2
1   0.0957 0.0729 1.3119 0.1896 -0.0473 0.2386 161.5263 0.0000 0.0414 99.0112 101.1347
2   0.0876 0.0733 1.1952 0.2320 -0.0560 0.2312 143.2490 0.0000 0.0418 99.0211 102.1549
3   0.0979 0.0726 1.3475 0.1778 -0.0445 0.2402 159.6389 0.0000 0.0410 99.0025 100.2483
4   0.0996 0.0723 1.3786 0.1680 -0.0420 0.2413 155.8633 0.0000 0.0407 98.9931  99.3164
5   0.1081 0.0698 1.5490 0.1214 -0.0287 0.2448 125.0957 0.0000 0.0378 98.9699  97.0733
6   0.0916 0.0733 1.2504 0.2112 -0.0520 0.2352 157.2717 0.0000 0.0418 99.0207 102.1144
7   0.0281 0.0191 1.4713 0.1412 -0.0093 0.0655  56.6486 0.0000 0.0023 87.2863   7.8655
8   0.0933 0.0728 1.2814 0.2001 -0.0494 0.2360 161.4356 0.0000 0.0415 99.1770 121.5137
9   0.1002 0.0715 1.4010 0.1612 -0.0400 0.2404 160.8609 0.0000 0.0401 99.1603 119.0858
```

Enrolment boys

```
> # Sensitivity Analysis
> #---------------------------------------------------
>  leave1out(fit1a)
  estimate     se   zval   pval   ci.lb  ci.ub        Q      Qp   tau2      I2       H2
1   0.1850 0.1618 1.1439 0.2527 -0.1320 0.5021 138.5983 0.0000 0.1283 99.8676 755.2382
2   0.2074 0.1576 1.3166 0.1880 -0.1014 0.5162 164.4926 0.0000 0.1215 99.3614 156.5895
3   0.2107 0.1565 1.3464 0.1782 -0.0960 0.5174 152.4751 0.0000 0.1198 99.3525 154.4461
4   0.2056 0.1573 1.3071 0.1912 -0.1027 0.5139 164.5118 0.0000 0.1217 99.8736 791.3427
5   0.2012 0.1581 1.2726 0.2032 -0.1087 0.5110 164.3877 0.0000 0.1233 99.8760 806.3975
6   0.0454 0.0271 1.6762 0.0937 -0.0077 0.0984  30.7613 0.0000 0.0027 94.5871  18.4745
> #---------------------------------------------------
```

Enrolment girls

```
> #---------------------------------------------------
> # Sensitivity Analysis
> #---------------------------------------------------
>  leave1out(fit1b)
  estimate     se   zval   pval   ci.lb  ci.ub        Q      Qp   tau2      I2       H2
1   0.1640 0.1344 1.2202 0.2224 -0.0994 0.4274 102.5845 0.0000 0.0878 99.8067 517.2398
2   0.1785 0.1311 1.3620 0.1732 -0.0784 0.4354 112.6508 0.0000 0.0833 99.0617 106.5729
3   0.1806 0.1303 1.3855 0.1659 -0.0749 0.4361 106.6789 0.0000 0.0824 99.0510 105.3731
4   0.1591 0.1342 1.1858 0.2357 -0.1039 0.4221 110.0004 0.0000 0.0880 99.8251 571.6891
```

```
5   0.1796 0.1292 1.3900 0.1645 -0.0736 0.4329 112.7112 0.0000 0.0818 99.8130 534.7055
6   0.0363 0.0205 1.7679 0.0771 -0.0039 0.0765  15.4465 0.0039 0.0013 89.7779   9.7827
```

Enrolment low SES

```
> #---------------------------------------------------
> # Sensitivity Analysis
> #---------------------------------------------------
>  leave1out(fit1a)
  estimate     se    zval   pval   ci.lb  ci.ub       Q     Qp   tau2      I2      H2
1  -0.0164 0.0134 -1.2236 0.2211 -0.0427 0.0099  9.1674 0.0025 0.0003 89.0918  9.1674
2   0.0327 0.0638  0.5132 0.6078 -0.0923 0.1578 69.8473 0.0000 0.0080 98.5683 69.8473
3   0.0459 0.0504  0.9107 0.3625 -0.0528 0.1446 52.7429 0.0000 0.0050 98.1040 52.7429
> #---------------------------------------------------
```

Enrolment high SES

```
> #---------------------------------------------------
> # Sensitivity Analysis
> #---------------------------------------------------
>  leave1out(fit1b)
  estimate     se    zval   pval   ci.lb  ci.ub       Q     Qp   tau2      I2      H2
1  -0.0054 0.0032 -1.6730 0.0943 -0.0118 0.0009  0.3090 0.5783 0.0000  0.0000  1.0000
2   0.0501 0.0582  0.8612 0.3891 -0.0639 0.1641 68.7114 0.0000 0.0067 98.5446 68.7114
3   0.0519 0.0563  0.9204 0.3574 -0.0586 0.1623 65.9997 0.0000 0.0063 98.4848 65.9997
> #---------------------------------------------------
```

Drop out

```
> #---------------------------------------------------
> # Sensitivity Analysis
> #---------------------------------------------------
>  leave1out(fit1)
  estimate     se    zval   pval   ci.lb  ci.ub        Q     Qp   tau2      I2       H2
1  -0.0405 0.0244 -1.6588 0.0971 -0.0884 0.0074  24.2517 0.0000 0.0016 90.9317  11.0274
2  -0.1291 0.0834 -1.5481 0.1216 -0.2925 0.0343 232.5573 0.0000 0.0207 99.2210
128.3654
3  -0.1331 0.0805 -1.6522 0.0985 -0.2909 0.0248 235.3439 0.0000 0.0193 99.2338
130.5134
```

```
4  -0.1072 0.0914 -1.1734 0.2406 -0.2864 0.0719 282.5834 0.0000 0.0249 99.2676
136.5358
> #-----------------------------------------------------
```

Completion

```
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub      Q     Qp   tau2      I2      H2
1   0.1967 0.1797 1.0944 0.2738 -0.1555 0.5488 62.1839 0.0000 0.0938 96.8072 31.3205
2   0.2296 0.1552 1.4792 0.1391 -0.0746 0.5339 46.3614 0.0000 0.0692 95.7239 23.3861
3   0.1540 0.1908 0.8073 0.4195 -0.2199 0.5279 69.9361 0.0000 0.1061 97.1700 35.3361
4   0.0246 0.0657 0.3743 0.7082 -0.1042 0.1533  8.4658 0.0145 0.0099 76.3835  4.2343
```

5.2. Cash Transfers

Enrolment

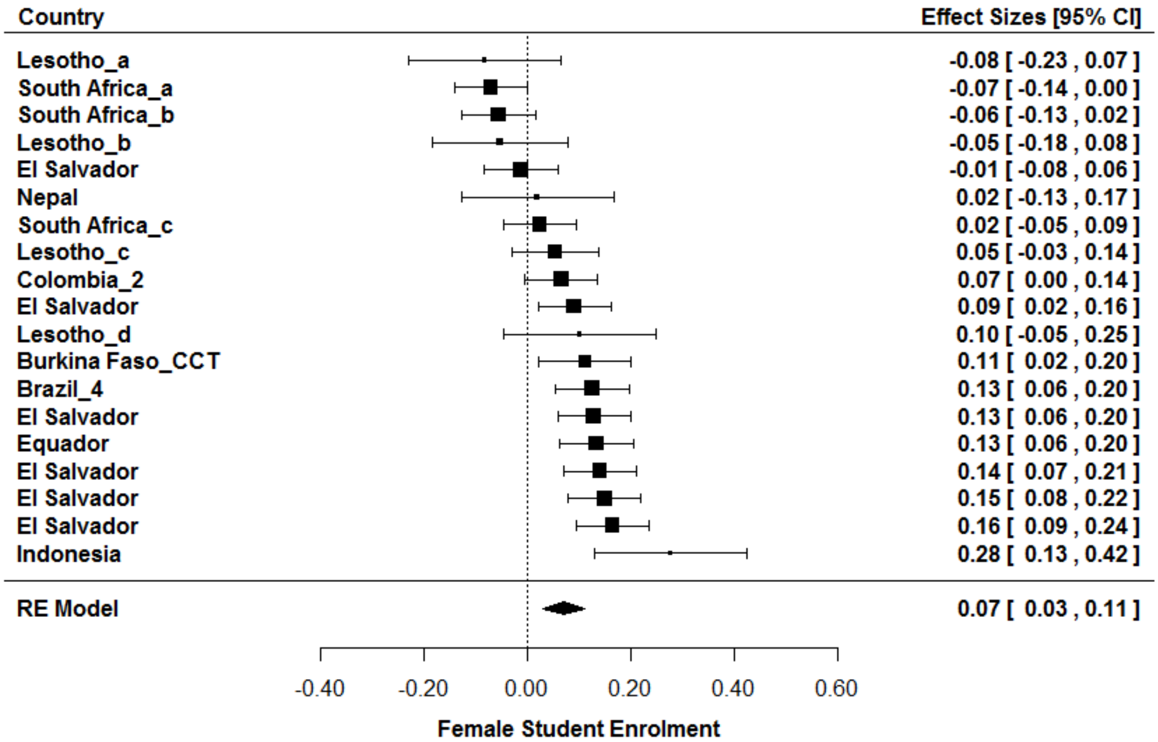
```
RobustResults(fit1)
t test of coefficients:
        Estimate Std. Error t value  Pr(>|t|)
intrcpt 0.114698   0.020466  5.6043 1.283e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Test for Heterogeneity:
Q(df = 48) = 671.7829, p-val < .0001
```

Moderator analysis - enrolment

```
t test of coefficients:
                             Estimate  Std. Error t value Pr(>|t|)
intrcpt                     3.8429e-03  2.3411e-02  0.1642  0.87048
as.numeric(Size)           -1.1252e-05  2.1518e-05 -0.5229  0.60407
factor(Recipient)household  8.2666e-02  3.6226e-02  2.2819  0.02818 *
factor(Recipient)mother     3.6630e-02  1.9927e-02  1.8382  0.07385 .
factor(Recipient)student    8.4677e-02  4.1873e-02  2.0222  0.05023 .
as.numeric(Intensity)       1.3147e-02  8.2243e-03  1.5985  0.11821
---
```

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Enrolment – girls



Random-Effects Model (k = 19; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0056 (SE = 0.0025)

tau (square root of estimated tau^2 value): 0.0745

I^2 (total heterogeneity / total variability): 76.49%

H^2 (total variability / sampling variability): 4.25

Test for Heterogeneity:

Q(df = 18) = 74.8414, p-val < .0001

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub	
0.0704	0.0201	3.5011	0.0005	0.0310	0.1098	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

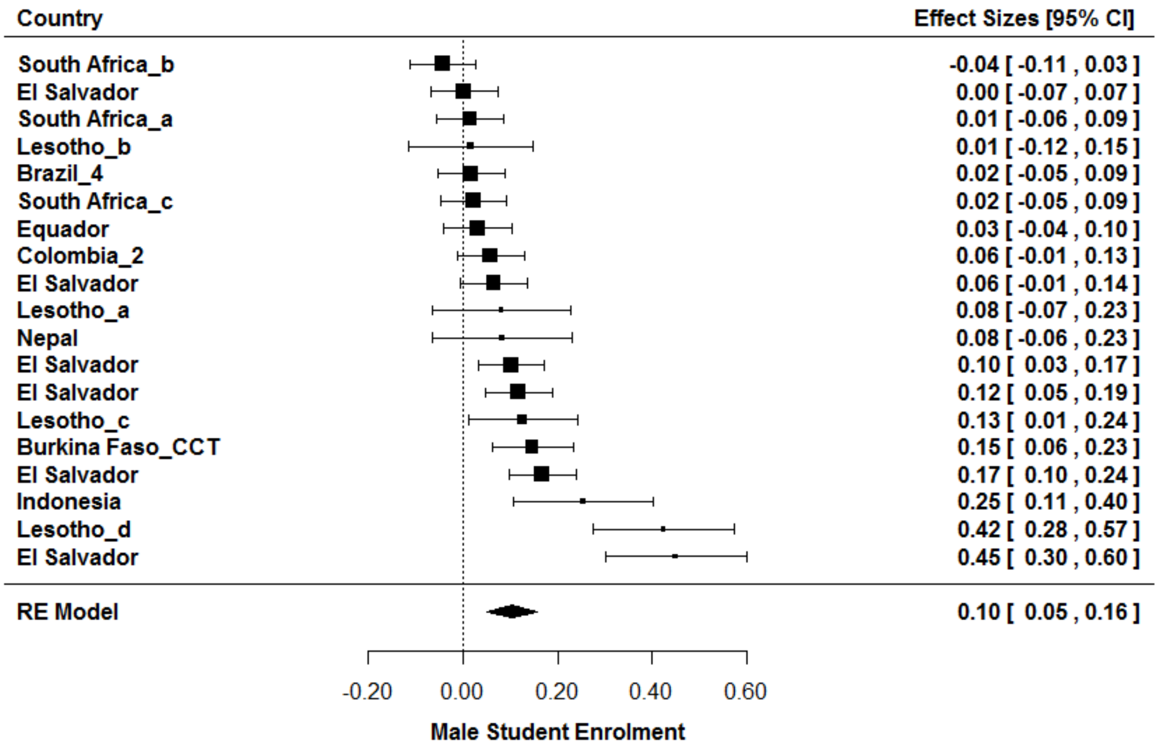
> #---

> # Sensitivity Analysis

> #---

```
>  leave1out(fit1b)
  estimate     se   zval   pval  ci.lb  ci.ub       Q     Qp  tau2      I2     H2
1    0.0706 0.0214 3.2916 0.0010 0.0286 0.1126 74.8092 0.0000 0.0061 77.7632 4.4971
2    0.0645 0.0204 3.1535 0.0016 0.0244 0.1046 67.7727 0.0000 0.0054 75.5038 4.0823
3    0.0654 0.0208 3.1522 0.0016 0.0247 0.1061 69.9270 0.0000 0.0056 76.2429 4.2093
4    0.0660 0.0209 3.1556 0.0016 0.0250 0.1069 70.9829 0.0000 0.0057 76.5872 4.2712
5    0.0755 0.0207 3.6572 0.0003 0.0351 0.1160 68.9360 0.0000 0.0055 75.9994 4.1666
6    0.0667 0.0211 3.1643 0.0016 0.0254 0.1080 72.1732 0.0000 0.0058 76.9623 4.3407
7    0.0690 0.0214 3.2255 0.0013 0.0271 0.1110 74.5528 0.0000 0.0061 77.6792 4.4801
8    0.0680 0.0212 3.2136 0.0013 0.0265 0.1095 74.0788 0.0000 0.0059 77.7992 4.5043
9    0.0669 0.0211 3.1678 0.0015 0.0255 0.1083 72.4807 0.0000 0.0058 77.0572 4.3587
10   0.0664 0.0210 3.1602 0.0016 0.0252 0.1075 71.6919 0.0000 0.0058 76.8154 4.3132
11   0.0631 0.0193 3.2649 0.0011 0.0252 0.1010 67.3767 0.0000 0.0048 74.3709 3.9018
12   0.0760 0.0200 3.8043 0.0001 0.0369 0.1152 70.5553 0.0000 0.0052 76.0416 4.1739
13   0.0755 0.0202 3.7321 0.0002 0.0359 0.1152 71.2876 0.0000 0.0054 76.4617 4.2484
14   0.0712 0.0213 3.3423 0.0008 0.0295 0.1130 74.6584 0.0000 0.0060 77.9971 4.5449
15   0.0691 0.0209 3.3091 0.0009 0.0282 0.1101 74.6835 0.0000 0.0059 78.1146 4.5692
16   0.0723 0.0208 3.4722 0.0005 0.0315 0.1131 74.3331 0.0000 0.0058 77.9487 4.5349
17   0.0795 0.0190 4.1883 0.0000 0.0423 0.1167 57.9703 0.0000 0.0044 71.4986 3.5086
18   0.0785 0.0195 4.0178 0.0001 0.0402 0.1167 61.3269 0.0000 0.0047 73.0845 3.7153
19   0.0732 0.0212 3.4527 0.0006 0.0316 0.1147 72.9342 0.0000 0.0059 77.2429 4.3942
> #-------------------------------------------------
```

Enrolment – boys



Random-Effects Model (k = 19; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0114 (SE = 0.0046)

tau (square root of estimated tau^2 value): 0.1068

I^2 (total heterogeneity / total variability): 86.03%

H^2 (total variability / sampling variability): 7.16

Test for Heterogeneity:

Q(df = 18) = 88.1725, p-val < .0001

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub	
0.1022	0.0271	3.7730	0.0002	0.0491	0.1553	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #---

> # Sensitivity Analysis

> #---

> leave1out(fit1a)

estimate se zval pval ci.lb ci.ub Q Qp tau2 I2 H2

1 0.1055 0.0289 3.6490 0.0003 0.0488 0.1622 87.9683 0.0000 0.0124 86.7358 7.5391

2 0.0833 0.0216 3.8546 0.0001 0.0409 0.1256 63.2851 0.0000 0.0061 77.4482 4.4342

3 0.1020 0.0290 3.5167 0.0004 0.0451 0.1588 86.5926 0.0000 0.0125 86.8109 7.5820

4 0.0986 0.0285 3.4546 0.0006 0.0427 0.1546 80.8569 0.0000 0.0121 86.3832 7.3439

5 0.1085 0.0282 3.8423 0.0001 0.0532 0.1639 83.8814 0.0000 0.0117 86.0740 7.1808

6 0.1029 0.0290 3.5459 0.0004 0.0460 0.1599 87.5355 0.0000 0.0126 86.8492 7.6041

7 0.1052 0.0290 3.6318 0.0003 0.0484 0.1619 88.0994 0.0000 0.0125 86.7722 7.5598

8 0.1002 0.0287 3.4870 0.0005 0.0439 0.1565 85.2843 0.0000 0.0123 86.8944 7.6303

9 0.1078 0.0285 3.7856 0.0002 0.0520 0.1636 85.4229 0.0000 0.0120 86.3059 7.3024

10 0.1070 0.0287 3.7330 0.0002 0.0508 0.1632 86.7327 0.0000 0.0122 86.5438 7.4315

11 0.0951 0.0272 3.4996 0.0005 0.0418 0.1483 82.3497 0.0000 0.0109 85.9393 7.1120

12 0.1037 0.0286 3.6321 0.0003 0.0477 0.1597 88.1640 0.0000 0.0123 87.3054 7.8774

13 0.1068 0.0283 3.7736 0.0002 0.0513 0.1623 87.3829 0.0000 0.0120 86.9765 7.6784

14 0.1015 0.0287 3.5384 0.0004 0.0453 0.1578 87.3747 0.0000 0.0123 87.2226 7.8263

15 0.0850 0.0226 3.7657 0.0002 0.0407 0.1292 66.5334 0.0000 0.0069 79.3834 4.8505

16 0.1036 0.0286 3.6286 0.0003 0.0476 0.1596 88.1580 0.0000 0.0123 87.3075 7.8787

17 0.1079 0.0284 3.7933 0.0001 0.0521 0.1636 85.2246 0.0000 0.0119 86.2758 7.2864

18 0.1107 0.0272 4.0674 0.0000 0.0573 0.1640 76.6946 0.0000 0.0107 84.9505 6.6447

19 0.1075 0.0285 3.7662 0.0002 0.0516 0.1635 85.9011 0.0000 0.0121 86.3790 7.3416

> #---

Attendance

Multivariate Meta-Analysis Model (k = 38; method: REML)

Variance Components:

estimate sqrt nlvls fixed factor

sigma^2 0.0195 0.1396 33 no IDnum

Test for Heterogeneity:

Q(df = 37) = 941.3885, p-val < .0001

I^2= 96.069%

Moderator analysis: attendance

t test of coefficients:

Estimate Std. Error t value Pr(>|t|)

intrcpt -3.6132e-02 1.4731e-02 -2.4528 0.021196 * as.numeric(Size) 4.4327e-05 7.2462e-05 0.6117 0.546029

factor(Recipient)grandparent -2.3759e-02 1.3102e-02 -1.8134 0.081331 .

factor(Recipient)household 1.5731e-01 4.8074e-02 3.2723 0.003010 **

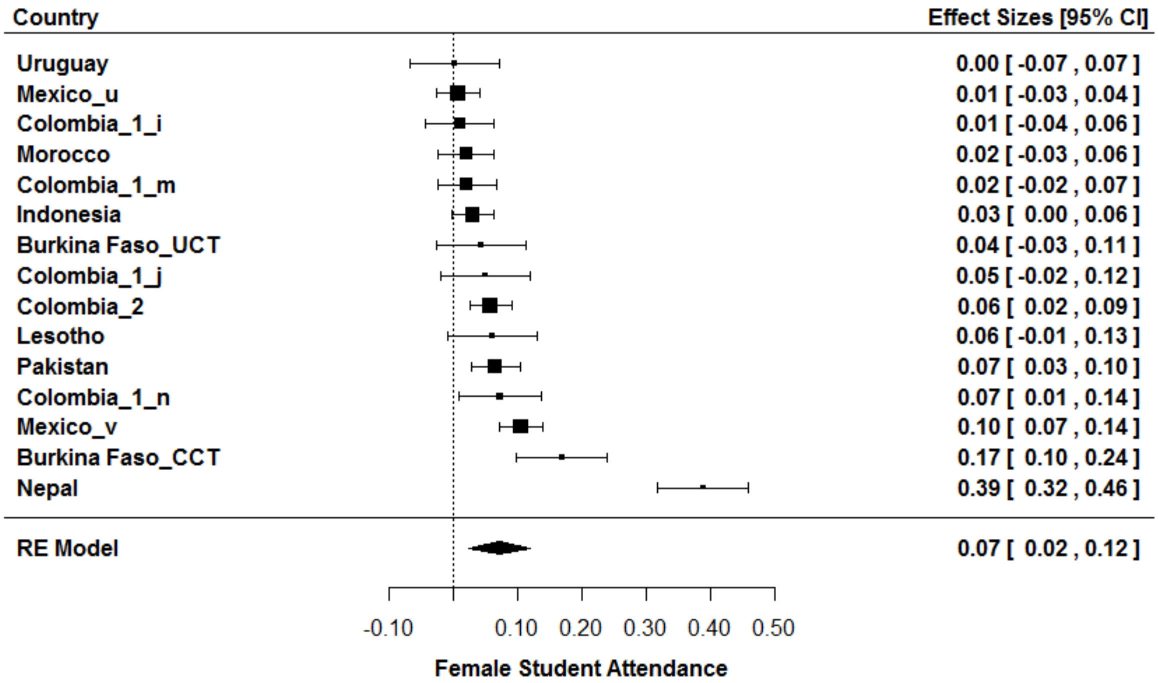
factor(Recipient)mother 5.5468e-02 2.8671e-02 1.9346 0.063987 .

factor(Recipient)student 1.1203e-01 7.4068e-02 1.5125 0.142473

as.numeric(Intensity) 1.1976e-02 3.9036e-03 3.0679 0.004987 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Attendance – girls



Random-Effects Model (k = 15; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0079 (SE = 0.0033)

tau (square root of estimated tau^2 value): 0.0888

I^2 (total heterogeneity / total variability): 93.33%

H^2 (total variability / sampling variability): 14.99

Test for Heterogeneity:

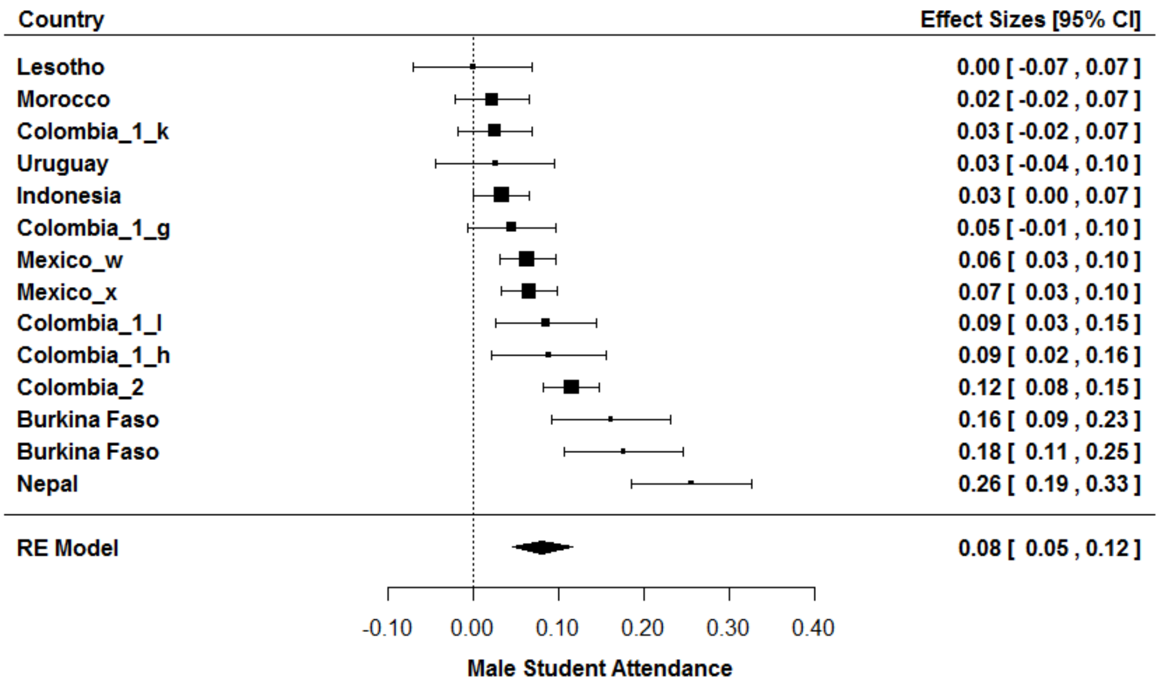
Q(df = 14) = 124.6363, p-val < .0001

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub	
0.0712	0.0240	2.9625	0.0031	0.0241	0.1184	**

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
>  leave1out(fit1b)
  estimate     se   zval   pval  ci.lb  ci.ub        Q     Qp   tau2      I2      H2
1    0.0724 0.0260 2.7850 0.0054 0.0215 0.1234 124.6362 0.0000 0.0086 93.4338 15.2295
2    0.0752 0.0256 2.9325 0.0034 0.0249 0.1254 121.5654 0.0000 0.0084 93.6442 15.7338
3    0.0761 0.0253 3.0081 0.0026 0.0265 0.1256 122.0427 0.0000 0.0082 93.8014 16.1328
4    0.0647 0.0247 2.6174 0.0089 0.0162 0.1131 114.7003 0.0000 0.0078 93.5067 15.4005
5    0.0718 0.0260 2.7627 0.0057 0.0209 0.1228 124.4558 0.0000 0.0087 93.6729 15.8051
6    0.0757 0.0255 2.9730 0.0029 0.0258 0.1257 121.3995 0.0000 0.0083 93.7236 15.9328
7    0.0729 0.0258 2.8286 0.0047 0.0224 0.1234 124.5742 0.0000 0.0085 94.0378 16.7724
8    0.0750 0.0257 2.9253 0.0034 0.0248 0.1253 121.9286 0.0000 0.0084 93.6716 15.8018
9    0.0713 0.0259 2.7566 0.0058 0.0206 0.1220 124.4133 0.0000 0.0086 94.0384 16.7740
10   0.0761 0.0255 2.9893 0.0028 0.0262 0.1261 115.0797 0.0000 0.0083 93.2044 14.7153
11   0.0688 0.0259 2.6617 0.0078 0.0181 0.1195 115.9474 0.0000 0.0085 93.4086 15.1712
12   0.0745 0.0258 2.8884 0.0039 0.0240 0.1251 121.4552 0.0000 0.0085 93.3257 14.9828
13   0.0722 0.0258 2.7959 0.0052 0.0216 0.1228 124.6324 0.0000 0.0086 94.0561 16.8239
14   0.0486 0.0110 4.4162 0.0000 0.0270 0.0701  37.5317 0.0003 0.0010 65.4589  2.8951
15   0.0733 0.0257 2.8475 0.0044 0.0228 0.1237 124.4620 0.0000 0.0085 94.0226 16.7297
> #----------------------------------------------------
```

Attendance – boys



Random-Effects Model (k = 14; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0038 (SE = 0.0018)

tau (square root of estimated tau^2 value): 0.0615

I^2 (total heterogeneity / total variability): 87.02%

H^2 (total variability / sampling variability): 7.70

Test for Heterogeneity:

Q(df = 13) = 70.3775, p-val < .0001

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub	
0.0809	0.0180	4.4888	<.0001	0.0456	0.1162	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

\> #---

\> # Sensitivity Analysis

\> #---

\> leave1out(fit1a)

estimate se zval pval ci.lb ci.ub Q Qp tau2 I2 H2

1 0.0780 0.0194 4.0203 0.0001 0.0400 0.1161 62.0317 0.0000 0.0041 87.0016 7.6933

2 0.0857 0.0189 4.5287 0.0000 0.0486 0.1228 65.3636 0.0000 0.0039 87.1562 7.7858

3 0.0847 0.0190 4.4647 0.0000 0.0475 0.1219 68.7659 0.0000 0.0039 87.9747 8.3158

4 0.0741 0.0177 4.1745 0.0000 0.0393 0.1089 61.2700 0.0000 0.0034 86.1842 7.2381

5 0.0838 0.0194 4.3289 0.0000 0.0459 0.1218 69.3699 0.0000 0.0041 88.0475 8.3664

6 0.0805 0.0194 4.1437 0.0000 0.0424 0.1186 70.1011 0.0000 0.0042 88.5097 8.7030

7 0.0855 0.0190 4.4947 0.0000 0.0482 0.1227 65.8717 0.0000 0.0039 87.2412 7.8377

8 0.0807 0.0195 4.1349 0.0000 0.0425 0.1190 70.1207 0.0000 0.0042 88.4571 8.6634

9 0.0826 0.0197 4.1934 0.0000 0.0440 0.1212 70.1684 0.0000 0.0042 87.3807 7.9244

10 0.0824 0.0197 4.1798 0.0000 0.0438 0.1210 70.2846 0.0000 0.0042 87.4028 7.9383

11 0.0851 0.0192 4.4247 0.0000 0.0474 0.1228 64.6098 0.0000 0.0040 86.7378 7.5402

12 0.0863 0.0184 4.6911 0.0000 0.0503 0.1224 66.3099 0.0000 0.0037 87.1876 7.8050

13 0.0678 0.0135 5.0338 0.0000 0.0414 0.0942 42.6243 0.0000 0.0017 75.4926 4.0804

14 0.0752 0.0182 4.1246 0.0000 0.0395 0.1110 63.6818 0.0000 0.0036 86.9446 7.6596

> #---

Dropout

Random-Effects Model (k = 16; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0091 (SE = 0.0036)

tau (square root of estimated tau^2 value): 0.0955

I^2 (total heterogeneity / total variability): 92.70%

H^2 (total variability / sampling variability): 13.70

Test for Heterogeneity:

Q(df = 15) = 192.1435, p-val < .0001

Model Results:

estimate se zval pval ci.lb ci.ub

-0.1229 0.0248 -4.9562 <.0001 -0.1715 -0.0743 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> leave1out(fit1)

estimate se zval pval ci.lb ci.ub Q Qp tau2 I2 H2

1 -0.1139 0.0247 -4.6158 0.0000 -0.1623 -0.0656 167.3506 0.0000 0.0084 92.2214 12.8559

2   -0.1218 0.0265 -4.5969 0.0000 -0.1737 -0.0699 191.4225 0.0000 0.0098 93.1597 14.6193

3   -0.1286 0.0258 -4.9818 0.0000 -0.1792 -0.0780 181.9996 0.0000 0.0093 92.7919 13.8733

4   -0.1265 0.0262 -4.8277 0.0000 -0.1779 -0.0752 188.9192 0.0000 0.0096 93.1000 14.4928

5   -0.1180 0.0260 -4.5449 0.0000 -0.1689 -0.0671 184.2789 0.0000 0.0094 92.9700 14.2247

6   -0.1261 0.0263 -4.7936 0.0000 -0.1776 -0.0745 189.3271 0.0000 0.0097 93.0684 14.4267

7   -0.1348 0.0232 -5.8183 0.0000 -0.1802 -0.0894 144.2556 0.0000 0.0073 91.0536 11.1777

8   -0.1059 0.0194 -5.4674 0.0000 -0.1439 -0.0679 107.5612 0.0000 0.0049 87.3709 7.9182

9   -0.1242 0.0265 -4.6885 0.0000 -0.1761 -0.0723 191.8186 0.0000 0.0098 93.1562 14.6118

10 -0.1225 0.0265 -4.6198 0.0000 -0.1745 -0.0705 191.9637 0.0000 0.0098 93.1723 14.6461

11 -0.1254 0.0264 -4.7504 0.0000 -0.1771 -0.0736 190.5255 0.0000 0.0097 93.1062 14.5057

12 -0.1222 0.0265 -4.6107 0.0000 -0.1742 -0.0703 191.8015 0.0000 0.0098 93.1689 14.6389

13 -0.1191 0.0262 -4.5457 0.0000 -0.1705 -0.0678 186.2695 0.0000 0.0096 93.0082 14.3025

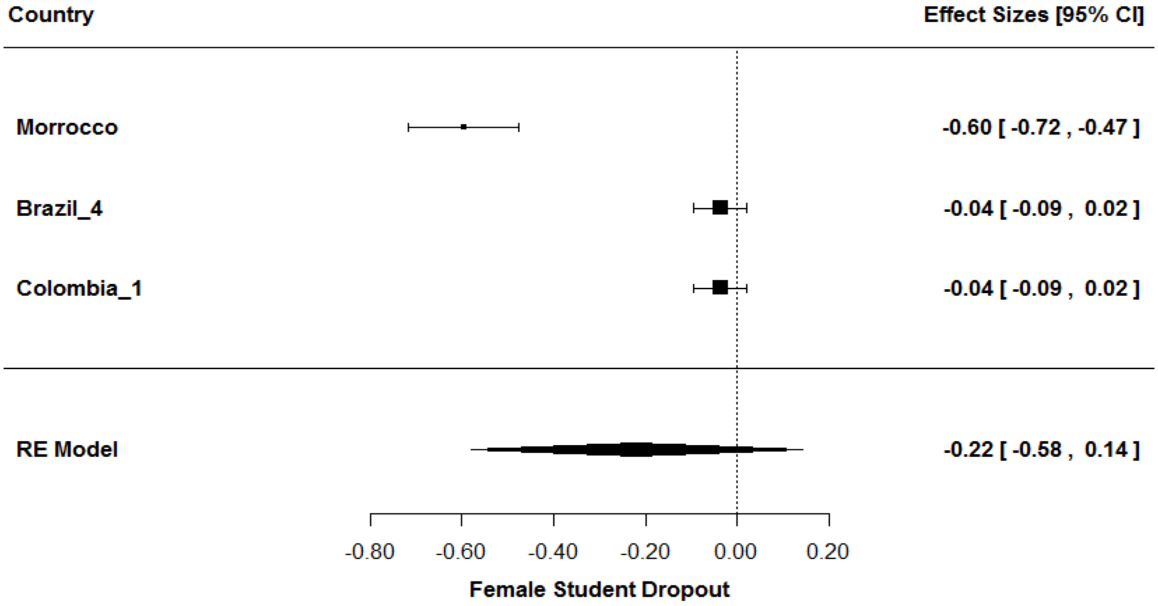
14 -0.1245 0.0265 -4.7017 0.0000 -0.1764 -0.0726 191.6094 0.0000 0.0098 93.1476 14.5934

15 -0.1233 0.0265 -4.6495 0.0000 -0.1753 -0.0713 192.1416 0.0000 0.0098 93.1726 14.6469

16 -0.1293 0.0256 -5.0491 0.0000 -0.1795 -0.0791 179.0374 0.0000 0.0091 92.6789 13.6591

> #---

Dropout – girls



Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.1006 (SE = 0.1024)

tau (square root of estimated tau^2 value): 0.3172

I^2 (total heterogeneity / total variability): 98.69%

H^2 (total variability / sampling variability): 76.46

Test for Heterogeneity:

Q(df = 2) = 74.3173, p-val < .0001

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub
-0.2188	0.1848	-1.1841	0.2364	-0.5809	0.1434

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #---

> # Sensitivity Analysis

> #---

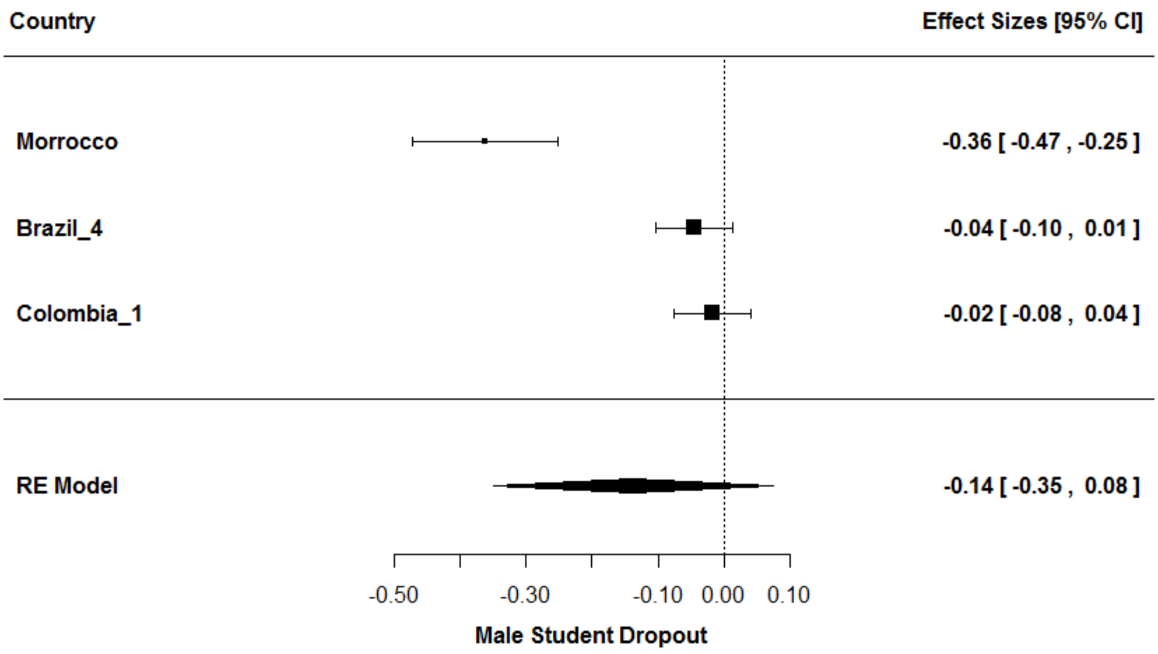
> leave1out(fit1b)

	estimate	se	zval	pval	ci.lb	ci.ub	Q	Qp	tau2	I2	H2
1	-0.0357	0.0209	-1.7095	0.0874	-0.0767	0.0052	0.0001	0.9934	0.0000	0.0000	1.0000
2	-0.3129	0.2799	-1.1178	0.2636	-0.8614	0.2357	67.4320	0.0000	0.1544	98.5170	67.4320

3 -0.3130 0.2797 -1.1192 0.2631 -0.8612 0.2352 67.2375 0.0000 0.1542 98.5127 67.2375

> #--

Dropout- boys



Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0335 (SE = 0.0351)

tau (square root of estimated tau^2 value):   0.1829

I^2 (total heterogeneity / total variability):   96.33%

H^2 (total variability / sampling variability): 27.24

Test for Heterogeneity:

Q(df = 2) = 30.9517, p-val < .0001

Model Results:

estimate   se   zval   pval   ci.lb   ci.ub

 -0.1369   0.1081 -1.2660   0.2055 -0.3488   0.0750

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #--

> # Sensitivity Analysis

> #--

>   leave1out(fit1a)

  estimate   se   zval   pval   ci.lb ci.ub   Q   Qp   tau2   I2   H2

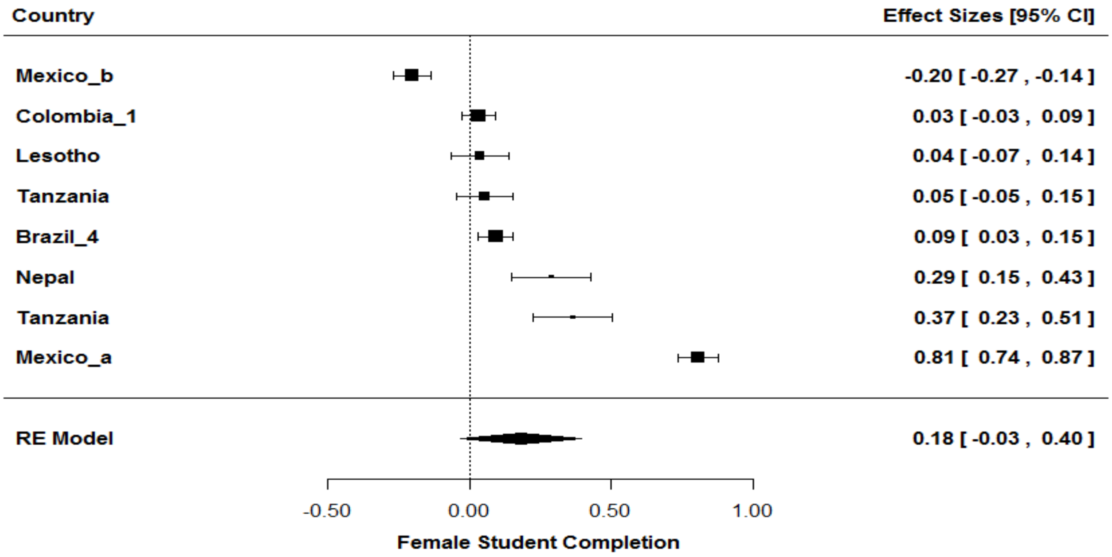
1 -0.0313 0.0209 -1.5012 0.1333 -0.0722 0.0096 0.4258 0.5141 0.0000 0.0000 1.0000

2 -0.1865 0.1721 -1.0838 0.2784 -0.5238 0.1508 29.5070 0.0000 0.0573 96.6110 29.5070

3 -0.1999 0.1585 -1.2610 0.2073 -0.5105 0.1108 25.0244 0.0000 0.0483 96.0039 25.0244

> #---

Completion – Girls



Random-Effects Model (k = 8; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0921 (SE = 0.0505)

tau (square root of estimated tau^2 value):   0.3035

I^2 (total heterogeneity / total variability):   98.19%

H^2 (total variability / sampling variability): 55.33

Test for Heterogeneity:

Q(df = 7) = 492.1081, p-val < .0001

Model Results:

estimate   se   zval   pval   ci.lb   ci.ub

  0.1834   0.1087   1.6869   0.0916   -0.0297   0.3965   .

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #--

> # Sensitivity Analysis

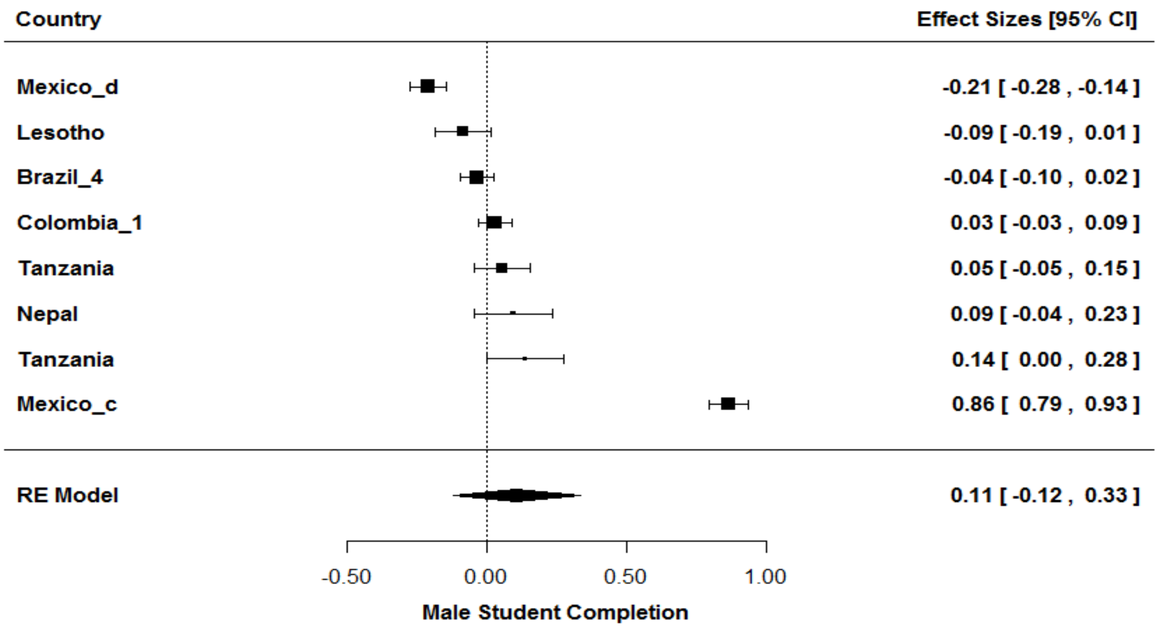
> #--

>   leave1out(fit1b)

  estimate   se   zval   pval   ci.lb   ci.ub   Q   Qp   tau2   I2   H2

1	0.1688	0.1238	1.3637	0.1727	-0.0738	0.4114	488.4492	0.0000	0.1052	98.5609	69.4889
2	0.2020	0.1236	1.6344	0.1022	-0.0402	0.4442	487.8188	0.0000	0.1045	98.4750	65.5731
3	0.1582	0.1215	1.3021	0.1929	-0.0799	0.3964	483.1502	0.0000	0.1013	98.5069	66.9742
4	0.0887	0.0691	1.2840	0.1991	-0.0467	0.2241	86.2094	0.0000	0.0308	94.4673	18.0745
5	0.2393	0.1082	2.2121	0.0270	0.0273	0.4513	358.0662	0.0000	0.0793	97.7403	44.2534
6	0.2054	0.1231	1.6689	0.0951	-0.0358	0.4467	471.3148	0.0000	0.1034	98.1788	54.9090
7	0.1967	0.1247	1.5772	0.1147	-0.0477	0.4412	486.6943	0.0000	0.1062	98.2264	56.3816
8	0.2043	0.1231	1.6597	0.0970	-0.0370	0.4455	486.7270	0.0000	0.1036	98.4731	65.4916

Completion – Boys



Random-Effects Model (k = 8; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.1059 (SE = 0.0579)

tau (square root of estimated tau^2 value): 0.3254

I^2 (total heterogeneity / total variability): 98.44%

H^2 (total variability / sampling variability): 63.90

Test for Heterogeneity:

Q(df = 7) = 587.8609, p-val < .0001

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub
0.1058	0.1163	0.9091	0.3633	-0.1223	0.3338

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

```
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
>  leave1out(fit1a)
  estimate     se   zval   pval   ci.lb  ci.ub       Q    Qp   tau2     I2     H2
1   0.1073 0.1337  0.8024 0.4223 -0.1548 0.3695 587.8307 0.0000 0.1232 98.7772 81.7792
2   0.1131 0.1338  0.8450 0.3981 -0.1492 0.3754 586.7205 0.0000 0.1230 98.7126 77.6729
3   0.1014 0.1337  0.7584 0.4482 -0.1606 0.3633 587.6680 0.0000 0.1230 98.7758 81.6860
4  -0.0111 0.0448 -0.2469 0.8050 -0.0989 0.0768  44.3004 0.0000 0.0116 86.6970  7.5171
5   0.1516 0.1237  1.2262 0.2201 -0.0907 0.3940 482.3502 0.0000 0.1045 98.2919 58.5450
6   0.1170 0.1337  0.8747 0.3817 -0.1451 0.3791 579.4493 0.0000 0.1226 98.4738 65.5242
7   0.1262 0.1323  0.9538 0.3402 -0.1331 0.3855 560.3376 0.0000 0.1199 98.4403 64.1158
8   0.1332 0.1304  1.0214 0.3071 -0.1224 0.3887 572.0635 0.0000 0.1166 98.6404 73.5497
```

Cognitive scores

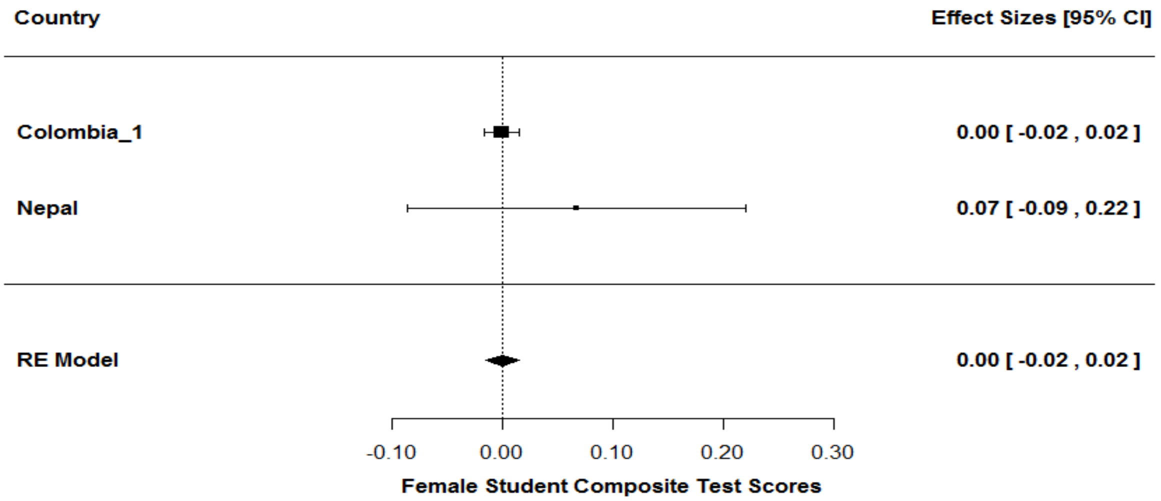
```
Random-Effects Model (k = 2; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.0152 (SE = 0.0238)
tau (square root of estimated tau^2 value):      0.1233
I^2 (total heterogeneity / total variability):   90.37%
H^2 (total variability / sampling variability):  10.38
Test for Heterogeneity:
Q(df = 1) = 10.3807, p-val = 0.0013
Model Results:
estimate      se    zval    pval    ci.lb    ci.ub
  0.0717   0.0916   0.7820   0.4342  -0.1079   0.2513
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Composite test scores

```
Random-Effects Model (k = 3; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.0003 (SE = 0.0003)
tau (square root of estimated tau^2 value):      0.0169
I^2 (total heterogeneity / total variability):   84.99%
H^2 (total variability / sampling variability):  6.66
```

```
Test for Heterogeneity:
Q(df = 2) = 13.9518, p-val = 0.0009
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.0089   0.0106   0.8343   0.4041  -0.0120   0.0297
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #---------------------------------------------------
> # Sensitivity Analysis
> #---------------------------------------------------
>   leave1out(fit1)
  estimate     se    zval   pval   ci.lb  ci.ub       Q      Qp   tau2      I2      H2
1   0.0179 0.0111  1.6187 0.1055 -0.0038 0.0396  3.7010 0.0544 0.0002 72.9802  3.7010
2  -0.0015 0.0069 -0.2213 0.8248 -0.0151 0.0120  2.0969 0.1476 0.0001 52.3115  2.0969
3   0.0103 0.0181  0.5711 0.5679 -0.0251 0.0458 13.8375 0.0002 0.0006 92.7733 13.8375
> #---------------------------------------------------
```

Composite – girls




```
Random-Effects Model (k = 2; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0044)
tau (square root of estimated tau^2 value):      0
I^2 (total heterogeneity / total variability):   0.00%
H^2 (total variability / sampling variability):  1.00
Test for Heterogeneity:
```

Q(df = 1) = 0.7374, p-val = 0.3905

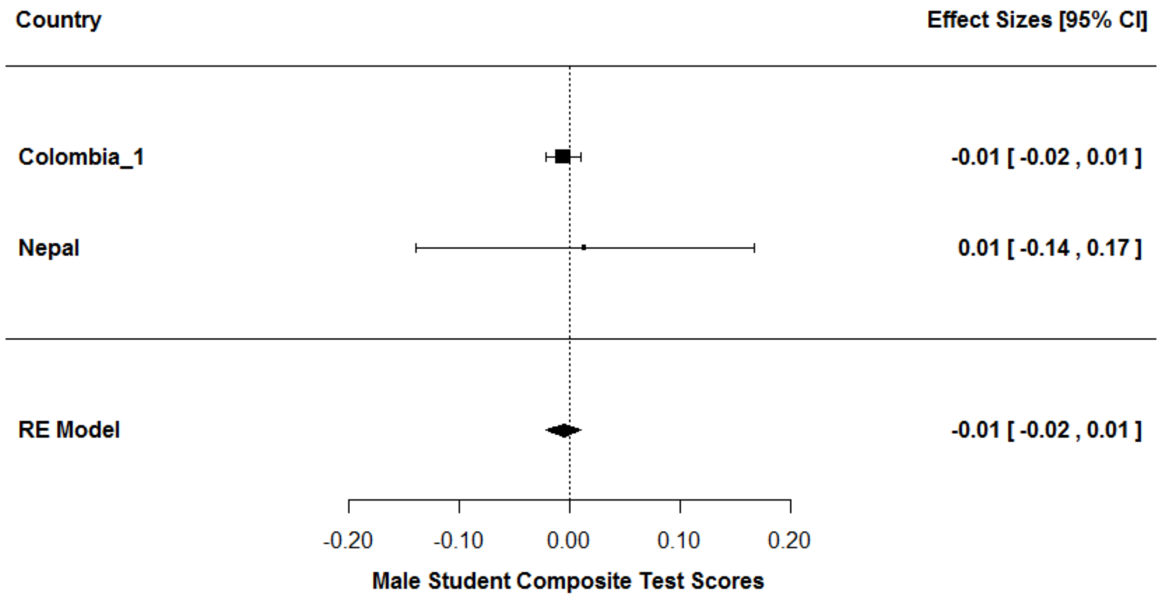
Model Results:

estimate se zval pval ci.lb ci.ub

0.0006 0.0080 0.0703 0.9440 -0.0151 0.0163

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Composite – boys



Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0044)

tau (square root of estimated tau^2 value): 0

I^2 (total heterogeneity / total variability): 0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 1) = 0.0596, p-val = 0.8071

Model Results:

estimate se zval pval ci.lb ci.ub

-0.0057 0.0083 -0.6851 0.4933 -0.0219 0.0106

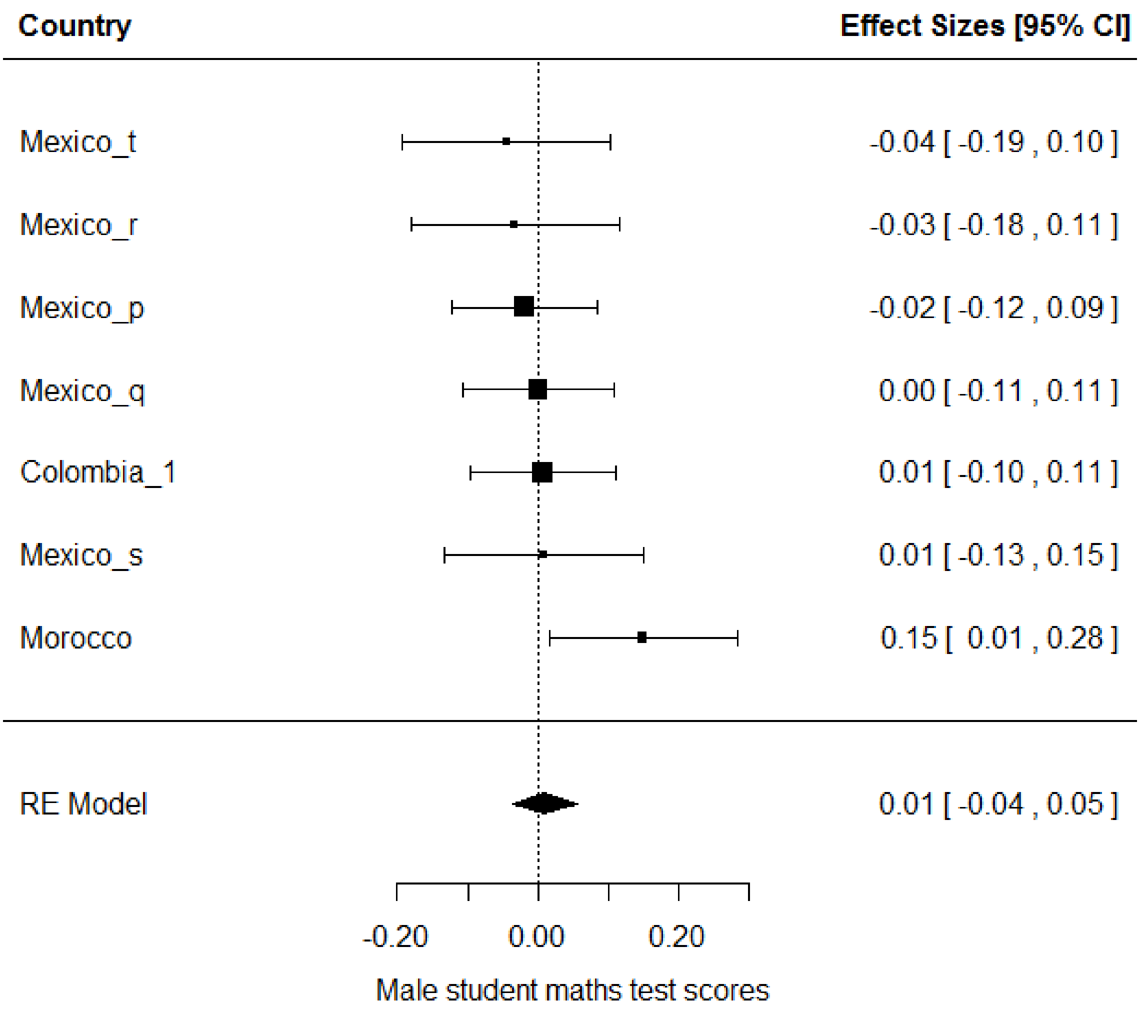
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Maths

Random-Effects Model (k = 14; tau^2 estimator: REML)

```
tau^2 (estimated amount of total heterogeneity): 0.0118 (SE = 0.0056)
tau (square root of estimated tau^2 value):      0.1088
I^2 (total heterogeneity / total variability):   86.43%
H^2 (total variability / sampling variability):  7.37
Test for Heterogeneity:
Q(df = 13) = 65.5800, p-val < .0001
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
 -0.0125   0.0318  -0.3918   0.6952  -0.0749   0.0499
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
   estimate     se    zval   pval   ci.lb  ci.ub      Q     Qp   tau2     I2     H2
1   -0.0173 0.0340 -0.5094 0.6105 -0.0839 0.0493 64.9299 0.0000 0.0128 87.7878 8.1885
2   -0.0206 0.0336 -0.6140 0.5392 -0.0865 0.0452 60.6798 0.0000 0.0122 86.6020 7.4638
3   -0.0164 0.0348 -0.4705 0.6380 -0.0845 0.0518 64.8804 0.0000 0.0132 87.0705 7.7343
4   -0.0208 0.0337 -0.6168 0.5374 -0.0868 0.0452 58.8740 0.0000 0.0123 86.1997 7.2462
5   -0.0175 0.0346 -0.5048 0.6137 -0.0852 0.0503 63.9720 0.0000 0.0131 86.9328 7.6528
6   -0.0225 0.0322 -0.6980 0.4852 -0.0857 0.0407 61.1012 0.0000 0.0113 86.3978 7.3517
7    0.0150 0.0168  0.8923 0.3722 -0.0180 0.0480 23.6905 0.0224 0.0017 48.9167 1.9576
8   -0.0096 0.0347 -0.2762 0.7824 -0.0775 0.0584 63.2343 0.0000 0.0132 87.3502 7.9052
9   -0.0147 0.0348 -0.4219 0.6731 -0.0830 0.0536 65.5465 0.0000 0.0133 87.5330 8.0212
10  -0.0100 0.0342 -0.2911 0.7710 -0.0770 0.0571 64.7976 0.0000 0.0130 87.9352 8.2886
11  -0.0136 0.0347 -0.3929 0.6944 -0.0816 0.0543 65.5681 0.0000 0.0133 87.9360 8.2891
12  -0.0103 0.0343 -0.2994 0.7646 -0.0775 0.0569 64.8654 0.0000 0.0130 87.9469 8.2966
13  -0.0102 0.0346 -0.2953 0.7677 -0.0780 0.0575 64.4408 0.0000 0.0131 87.7567 8.1677
14  -0.0095 0.0347 -0.2723 0.7854 -0.0775 0.0586 62.3692 0.0000 0.0132 87.0379 7.7148
> #-----------------------------------------------------
```

Maths – boys



Random-Effects Model (k = 7; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0022)

tau (square root of estimated tau^2 value):   0

I^2 (total heterogeneity / total variability):   0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 6) = 5.3136, p-val = 0.5043

Model Results:

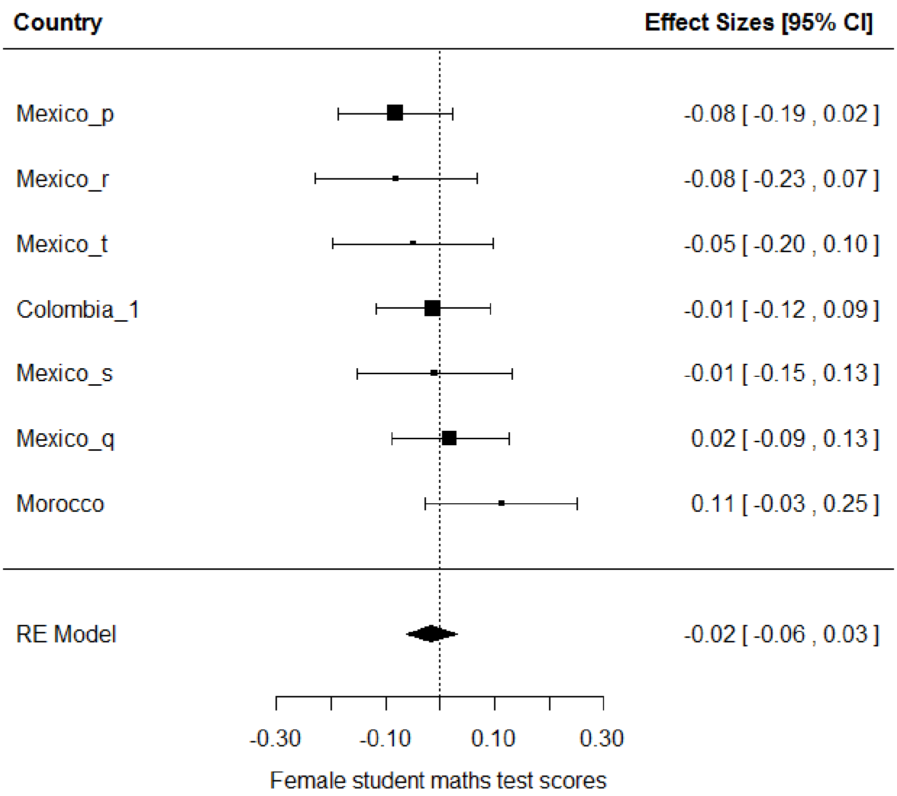
estimate	se	zval	pval	ci.lb	ci.ub
0.0087	0.0236	0.3710	0.7106	-0.0374	0.0549

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #---

```
> # Sensitivity Analysis
> #-----------------------------------------------------
>  leave1out(fit1a)
  estimate     se    zval   pval   ci.lb  ci.ub     Q     Qp   tau2     I2     H2
1  -0.0102 0.0251 -0.4071 0.6840 -0.0594 0.0390 0.5354 0.9908 0.0000 0.0000 1.0000
2   0.0091 0.0263  0.3472 0.7284 -0.0424 0.0606 5.3126 0.3789 0.0000 0.0031 1.0000
3   0.0155 0.0263  0.5890 0.5559 -0.0360 0.0670 4.9760 0.4188 0.0000 0.1399 1.0014
4   0.0106 0.0261  0.4062 0.6846 -0.0406 0.0619 5.2861 0.3820 0.0000 0.0483 1.0005
5   0.0133 0.0248  0.5345 0.5930 -0.0354 0.0619 4.9750 0.4189 0.0000 0.0900 1.0009
6   0.0089 0.0249  0.3562 0.7217 -0.0400 0.0578 5.3133 0.3788 0.0000 0.0372 1.0004
7   0.0146 0.0248  0.5866 0.5575 -0.0341 0.0632 4.7538 0.4467 0.0000 0.0141 1.0001
> #-----------------------------------------------------
```

Maths – girls




```
Random-Effects Model (k = 7; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.0000 (SE = 0.0022)
tau (square root of estimated tau^2 value):      0.0027
I^2 (total heterogeneity / total variability):   0.18%
H^2 (total variability / sampling variability):  1.00
```

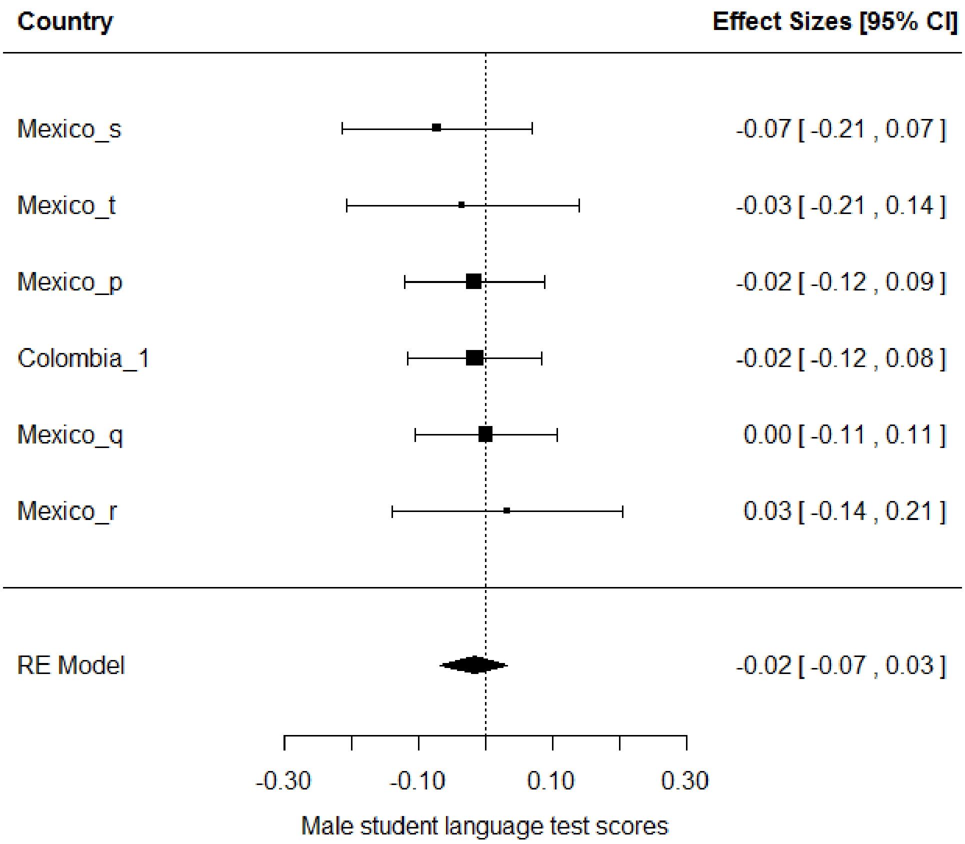
```
Test for Heterogeneity:
Q(df = 6) = 6.1685, p-val = 0.4046
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
 -0.0166   0.0237  -0.6989   0.4846  -0.0630   0.0299
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #---------------------------------------------------
> # Sensitivity Analysis
> #---------------------------------------------------
>   leave1out(fit1b)
  estimate     se    zval   pval   ci.lb  ci.ub      Q     Qp   tau2      I2     H2
1  -0.0328 0.0251 -1.3055 0.1917 -0.0820 0.0164 2.4311 0.7868 0.0000  0.0000 1.0000
2  -0.0166 0.0298 -0.5583 0.5767 -0.0749 0.0417 6.1640 0.2906 0.0010 19.6255 1.2442
3  -0.0005 0.0264 -0.0193 0.9846 -0.0523 0.0513 4.2876 0.5088 0.0000  0.0000 1.0000
4  -0.0242 0.0277 -0.8754 0.3814 -0.0785 0.0300 5.6523 0.3415 0.0004  8.8946 1.0976
5  -0.0093 0.0256 -0.3623 0.7172 -0.0594 0.0409 5.3691 0.3725 0.0002  4.6447 1.0487
6  -0.0170 0.0272 -0.6255 0.5316 -0.0703 0.0363 6.1588 0.2911 0.0006 13.7887 1.1599
7  -0.0125 0.0266 -0.4699 0.6384 -0.0646 0.0396 5.9624 0.3099 0.0005 11.1281 1.1252
> #---------------------------------------------------
> #---------------------------------------------------
```

Language Arts

```
Random-Effects Model (k = 14; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.0039 (SE = 0.0023)
tau (square root of estimated tau^2 value):      0.0627
I^2 (total heterogeneity / total variability):   72.19%
H^2 (total variability / sampling variability):  3.60
Test for Heterogeneity:
Q(df = 13) = 46.8803, p-val < .0001
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
 -0.0002   0.0205  -0.0074   0.9941  -0.0404   0.0401
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
  estimate     se    zval   pval   ci.lb  ci.ub       Q      Qp   tau2      I2     H2
1   -0.0030 0.0224 -0.1343 0.8932 -0.0469 0.0409 45.7715 0.0000 0.0044 73.0603 3.7120
2   -0.0114 0.0186 -0.6113 0.5410 -0.0479 0.0251 32.3688 0.0012 0.0026 62.0089 2.6322
3   -0.0007 0.0226 -0.0291 0.9768 -0.0450 0.0436 46.8690 0.0000 0.0045 73.5442 3.7799
4    0.0050 0.0210  0.2354 0.8139 -0.0363 0.0462 44.6811 0.0000 0.0039 72.9965 3.7032
5   -0.0095 0.0192 -0.4946 0.6209 -0.0471 0.0281 38.6479 0.0001 0.0030 67.3448 3.0623
6    0.0003 0.0220  0.0133 0.9894 -0.0428 0.0434 46.8498 0.0000 0.0044 74.9046 3.9848
7   -0.0019 0.0219 -0.0849 0.9324 -0.0448 0.0411 46.6872 0.0000 0.0043 74.7630 3.9624
8    0.0050 0.0216  0.2295 0.8185 -0.0374 0.0473 43.9913 0.0000 0.0041 72.6833 3.6608
9   -0.0008 0.0223 -0.0354 0.9718 -0.0446 0.0430 46.8614 0.0000 0.0045 74.5955 3.9363
10  -0.0003 0.0218 -0.0121 0.9903 -0.0430 0.0424 46.8802 0.0000 0.0043 74.9025 3.9845
11   0.0036 0.0216  0.1645 0.8694 -0.0389 0.0460 45.6412 0.0000 0.0042 73.9225 3.8347
12   0.0057 0.0208  0.2720 0.7856 -0.0351 0.0465 44.0196 0.0000 0.0038 72.4022 3.6235
13  -0.0039 0.0222 -0.1776 0.8590 -0.0475 0.0396 44.8333 0.0000 0.0043 72.6406 3.6551
14   0.0107 0.0193  0.5574 0.5772 -0.0270 0.0485 30.7194 0.0022 0.0028 63.2217 2.7190
> #-----------------------------------------------------
```

Language Arts – boys



Random-Effects Model (k = 6; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0024)

tau (square root of estimated tau^2 value): 0

I^2 (total heterogeneity / total variability): 0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 5) = 1.0474, p-val = 0.9586

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub
-0.0172	0.0256	-0.6710	0.5022	-0.0673	0.0330

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #---

> # Sensitivity Analysis

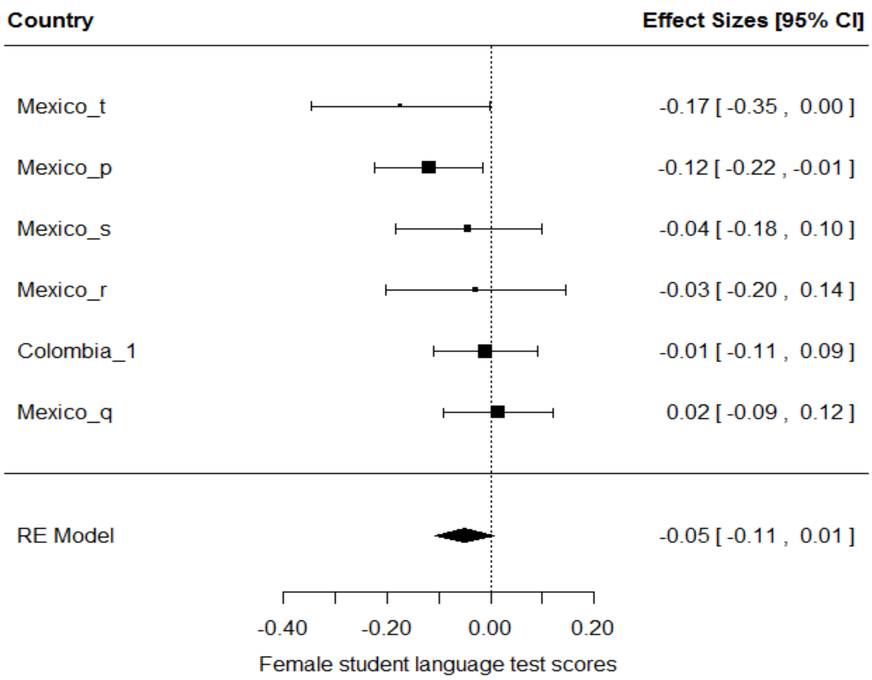
> #---

> leave1out(fit1a)

	estimate	se	zval	pval	ci.lb	ci.ub	Q	Qp	tau2	I2	H2
1	-0.0174	0.0297	-0.5884	0.5563	-0.0756	0.0407	1.0471	0.9026	0.0000	0.0000	1.0000
2	-0.0172	0.0292	-0.5900	0.5552	-0.0744	0.0400	1.0474	0.9025	0.0000	0.0000	1.0000
3	-0.0223	0.0290	-0.7705	0.4410	-0.0792	0.0345	0.9037	0.9240	0.0000	0.0000	1.0000
4	-0.0217	0.0267	-0.8116	0.4170	-0.0741	0.0307	0.7061	0.9506	0.0000	0.0000	1.0000
5	-0.0092	0.0274	-0.3371	0.7360	-0.0629	0.0444	0.3794	0.9841	0.0000	0.0000	1.0000
6	-0.0156	0.0267	-0.5824	0.5603	-0.0680	0.0368	1.0051	0.9090	0.0000	0.0000	1.0000

> #--

Language Arts – girls



Random-Effects Model (k = 6; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0009 (SE = 0.0030)

tau (square root of estimated tau^2 value): 0.0292

I^2 (total heterogeneity / total variability): 17.30%

H^2 (total variability / sampling variability): 1.21

Test for Heterogeneity:

Q(df = 5) = 5.7669, p-val = 0.3296

Model Results:

estimate se zval pval ci.lb ci.ub

```
 -0.0506   0.0286  -1.7672   0.0772  -0.1067   0.0055        .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #---------------------------------------------------
> # Sensitivity Analysis
> #---------------------------------------------------
>   leave1out(fit1b)
  estimate    se    zval   pval   ci.lb   ci.ub     Q     Qp   tau2     I2    H2
1  -0.0640 0.0350 -1.8291 0.0674 -0.1325  0.0046 4.9607 0.2914 0.0015 24.5889 1.3261
2  -0.0282 0.0292 -0.9669 0.3336 -0.0854  0.0290 3.5329 0.4729 0.0000  0.0502 1.0005
3  -0.0681 0.0310 -2.1948 0.0282 -0.1289 -0.0073 3.9871 0.4077 0.0005 10.1685 1.1132
4  -0.0540 0.0323 -1.6693 0.0950 -0.1174  0.0094 5.7106 0.2218 0.0015 28.8383 1.4053
5  -0.0532 0.0340 -1.5638 0.1179 -0.1198  0.0135 5.7598 0.2178 0.0018 31.1977 1.4534
6  -0.0377 0.0283 -1.3291 0.1838 -0.0932  0.0179 3.5822 0.4655 0.0004  9.5824 1.1060
> #---------------------------------------------------
```

6.1 Structured pedagogy

Attendance

```
Random-Effects Model (k = 5; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.0006 (SE = 0.0007)
tau (square root of estimated tau^2 value):      0.0246
I^2 (total heterogeneity / total variability):   61.18%
H^2 (total variability / sampling variability):  2.58
Test for Heterogeneity:
Q(df = 4) = 10.1726, p-val = 0.0376
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.0065   0.0141   0.4625   0.6437  -0.0210   0.0340
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #---------------------------------------------------
> # Sensitivity Analysis
> #---------------------------------------------------
```

```
>  leave1out(fit1)
  estimate     se    zval   pval   ci.lb  ci.ub       Q     Qp   tau2     I2     H2
1  -0.0027 0.0137 -0.1960 0.8446 -0.0295 0.0242  5.7696 0.1234 0.0004 47.7841 1.9151
2   0.0058 0.0184  0.3135 0.7539 -0.0303 0.0418 10.1710 0.0172 0.0010 70.9622 3.4438
3   0.0079 0.0181  0.4348 0.6637 -0.0277 0.0435  9.9762 0.0188 0.0009 70.7564 3.4196
4   0.0025 0.0175  0.1416 0.8874 -0.0319 0.0368  9.5374 0.0229 0.0008 69.1964 3.2464
5   0.0184 0.0096  1.9168 0.0553 -0.0004 0.0372  2.8667 0.4126 0.0000  0.0000 1.0000
```

Table A6.1 a: Effect sizes for studies not included in the meta-analysis for attendance

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
He, Linden and macLeod, 2007: Year 2 Programme	PicTalk Machine only treatment arm	4907	India	0.02855	0.00041	-0.01102	0.06812
He, Linden and macLeod, 2007: Year 2 Programme	Activities only treatment arm	4782	India	0.02025	0.00042	-0.01984	0.06033
Kerwin and Thornton, 2015	NGO treatment arm	4476	Uganda2	0.05783	0.00045	0.01639	0.09927

Completion

```
Random-Effects Model (k = 2; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.0078 (SE = 0.0168)
tau (square root of estimated tau^2 value):      0.0882
I^2 (total heterogeneity / total variability):   65.33%
H^2 (total variability / sampling variability):  2.88
Test for Heterogeneity:
Q(df = 1) = 2.8840, p-val = 0.0895
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.1318   0.0769   1.7139   0.0865  -0.0189   0.2824        .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #-----------------------------------------------------
> # Sensitivity Analysis
```

```
> #----------------------------------------------------
>  leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub      Q     Qp   tau2    I2    H2
1   0.2156 0.0718 3.0024 0.0027  0.0749 0.3563 0.0000 1.0000 0.0000   NaN   NaN
2   0.0612 0.0557 1.0995 0.2716 -0.0479 0.1704 0.0000 1.0000 0.0000   NaN   NaN
> #----------------------------------------------------
```

Cognitive test scores

```
Random-Effects Model (k = 2; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0024)
tau (square root of estimated tau^2 value):      0
I^2 (total heterogeneity / total variability):   0.00%
H^2 (total variability / sampling variability):  1.00
Test for Heterogeneity:
Q(df = 1) = 0.0148, p-val = 0.9030
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.0128   0.0292   0.4370   0.6621  -0.0445   0.0700
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
>  leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub      Q     Qp   tau2    I2    H2
1   0.0094 0.0402 0.2338 0.8152 -0.0694 0.0882 0.0000 1.0000 0.0000   NaN   NaN
2   0.0165 0.0425 0.3887 0.6975 -0.0668 0.0999 0.0000 1.0000 0.0000   NaN   NaN
> #----------------------------------------------------
```

Composite test scores

```
Random-Effects Model (k = 3; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.0001 (SE = 0.0006)
tau (square root of estimated tau^2 value):      0.0112
I^2 (total heterogeneity / total variability):   20.17%
```

H^2 (total variability / sampling variability): 1.25

Test for Heterogeneity:

Q(df = 2) = 2.3084, p-val = 0.3153

Model Results:

estimate   se   zval   pval   ci.lb   ci.ub

0.0557   0.0136   4.0913   <.0001   0.0290   0.0824   ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #--

> # Sensitivity Analysis

> #--

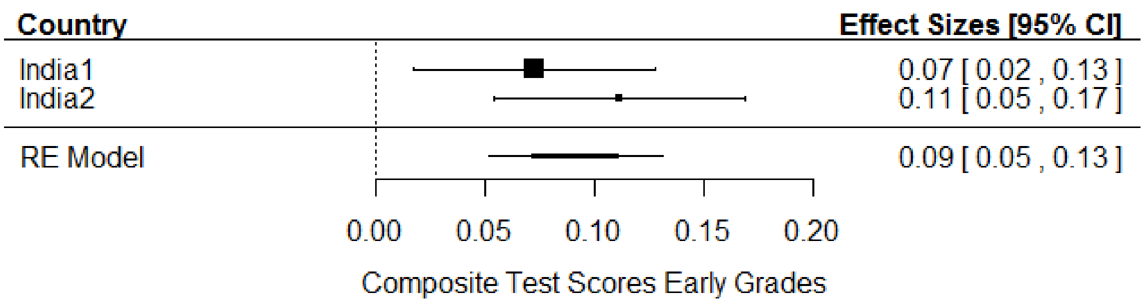
> leave1out(fit1)

	estimate	se	zval	pval	ci.lb	ci.ub	Q	Qp	tau2	I2	H2
1	0.0439	0.0138	3.1899	0.0014	0.0169	0.0709	0.3009	0.5833	0.0000	0.0000	1.0000
2	0.0648	0.0246	2.6367	0.0084	0.0166	0.1129	1.2867	0.2567	0.0004	22.2814	1.2867
3	0.0592	0.0155	3.8089	0.0001	0.0288	0.0897	1.7013	0.1921	0.0002	41.2209	1.7013

> #--

The studies reporting on the Pratham PicTalk year one and year two programmes.[149] also provide composite test scores by grade sub-group. The overall average estimates for grades one to three (0.09, 95% CI [0.05, 0.13]) and four to five are both positive and statistically significant, though there is no real substantive difference from those for the main sample.

Figure 6.1 m: composite test scores for grades 1-3 sub-group[150]



Random-Effects Model (k = 2; tau^2 estimator: REML)

[149] These two programmes shared the same name but took slightly different forms and were implemented in different locations. See Table 6.1a for full details.

[150] India1: He, Linden and MacLeod, 2007: Year 1
India2: He, Linden and MacLeod, 2007: Year 2
Philippines: Tan, Lane and Lassibille, 1999

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0012)

tau (square root of estimated tau^2 value): 0

I^2 (total heterogeneity / total variability): 0.00%

H^2 (total variability / sampling variability): 1.00

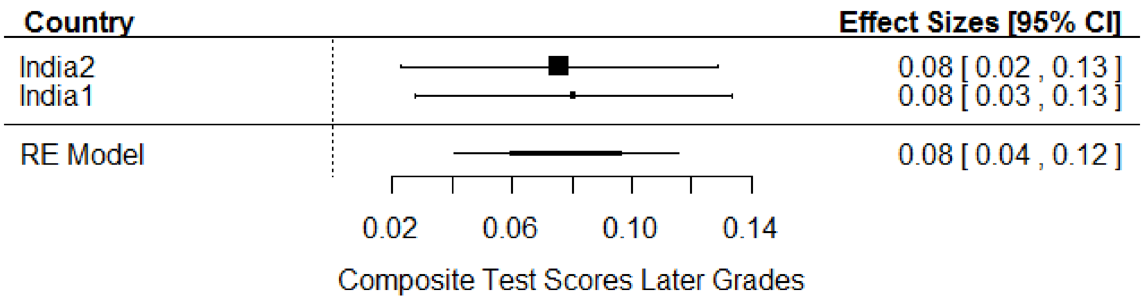
Test for Heterogeneity:

Q(df = 1) = 0.9327, p-val = 0.3342

Model Results:

```
estimate    se     zval     pval    ci.lb    ci.ub
 0.0914   0.0203  4.4952  <.0001  0.0516  0.1313    ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #---------------------------------------------------
> # Sensitivity Analysis
> #---------------------------------------------------
>  leave1out(fit1)
 estimate    se  zval   pval  ci.lb  ci.ub    Q    Qp  tau2    I2   H2
1   0.1119 0.0293 3.8122 0.0001 0.0544 0.1694 0.0000 1.0000 0.0000   NaN   NaN
2   0.0725 0.0282 2.5703 0.0102 0.0172 0.1279 0.0000 1.0000 0.0000   NaN   NaN
> #---------------------------------------------------
```

Figure 6.1 n: composite test scores for grades 4-5 sub-group[151]



Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0010)

tau (square root of estimated tau^2 value): 0

[151] India1: He, Linden and MacLeod, 2007: Year 1
India2: He, Linden and MacLeod, 2007: Year 2
Philippines: Tan, Lane and Lassibille, 1999

```
I^2 (total heterogeneity / total variability):   0.00%
H^2 (total variability / sampling variability):  1.00
Test for Heterogeneity:
Q(df = 1) = 0.0153, p-val = 0.9016
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.0782   0.0192   4.0808   <.0001   0.0406   0.1158      ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
  estimate     se   zval   pval  ci.lb  ci.ub      Q     Qp   tau2     I2    H2
1   0.0758 0.0271 2.8018 0.0051 0.0228 0.1289 0.0000 1.0000 0.0000   NaN   NaN
2   0.0806 0.0271 2.9696 0.0030 0.0274 0.1338 0.0000 1.0000 0.0000   NaN   NaN
> #-----------------------------------------------------
```

Language arts test scores

```
Multivariate Meta-Analysis Model (k = 67; method: REML)
Variance Components:
            estim    sqrt  nlvls  fixed  factor
sigma^2    0.0483  0.2197     17     no   IDnum
Test for Heterogeneity:
Q(df = 66) = 3537.8609, p-val < .0001
> #---------------------------------------------
>    RobustResults(fit1)
t test of coefficients:
        Estimate Std. Error t value  Pr(>|t|)
intrcpt 0.232906   0.053738  4.3341 0.0005129 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table A6.1 b: Effect sizes for studies not included in the meta-analysis for language arts test scores

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
Tan, Lane and Lassibille, 1999	English	1913	Philippines	0.21537	0.00105	0.15182	0.27893
Abeberese, Kumler and Linden, 2011	7 month follow-up	4566	Philippines	0.07192	0.00044	0.03088	0.11295
Nonoyama-Tarumi and Bredenberg, 2009	2 month follow-up	858	Cambodia	0.61092	0.00244	0.51410	0.70773
Jukes and Dubeck, 2015	English word identification (wpm); 24 month follow-up	1992	Kenya	0.16600	0.00101	0.10379	0.22821
Jukes and Dubeck, 2015	English letter knowledge 24 months follow-up	1992	Kenya	-0.01571		-0.08000	0.05000
Jukes and Dubeck, 2015	English passage reading fluency (wpm); 24 month follow-up	1997	Kenya	0.13613	0.00100	0.07403	0.19823
Jukes and Dubeck, 2015	English passage reading comprehension; 24 month follow-up	2002	Kenya	0.10238	0.00100	0.04039	0.16437
Jukes and Dubeck, 2015	English spelling; 24 month follow-up	1990	Kenya	0.22680	0.00101	0.16446	0.28913
Jukes and Dubeck, 2015	Beginning sounds; 9 month follow-up	2211	Kenya	0.14106	0.00091	0.08203	0.20008
Jukes and Dubeck, 2015	Receptive language; 9 month follow-up	2200	Kenya	0.19747	0.00091	0.13823	0.25671
RTI International, 2011	English Letter sound fluency (correct letters per min)	4566	Kenya	0.82258	0.00048		
RTI International, 2011	English Decoding fluency (correct words per min)	4566	Kenya	0.36641	0.00045		
RTI International, 2011	English Segmenting (%)	4566	Kenya	1.21809	0.00052		

RTI International, 2011	English Reading fluency (correct words per min)	4566	Kenya	0.32243	0.00044		
RTI International, 2011	English Vocabulary	4566	Kenya	0.10638	0.00044		
RTI International, 2011	English Reading comprehension	4566	Kenya	0.51402	0.00045		
RTI International, 2011	English Sentence comprehension	4566	Kenya	0.31757	0.00044		
RTI International, 2011	English Readers able to read at greater than 30 cwpm	4566	Kenya	0.13636	0.00044		
RTI International, 2011	English Readers able to read at least at the rate of 30 cwpm	4566	Kenya	0.56621	0.00046		

Language arts test scores sub-groups

Grades 1-3 sub-group

Multivariate Meta-Analysis Model (k = 63; method: REML)

Variance Components:

estim sqrt nlvls fixed factor

sigma^2 0.0488 0.2210 13 no IDnum

Test for Heterogeneity:

Q(df = 62) = 3036.0232, p-val < .0001

> #---

> RobustResults(fit1)

t test of coefficients:

Estimate Std. Error t value Pr(>|t|) intrcpt

0.228195 0.061723 3.6971 0.003052 **

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
Tan, Lane and Lassibille, 1999	English	1913	Philippines	0.21537	0.00105	0.15182	0.27893
Nonoyama-Tarumi and Bredenberg, 2009	2 month follow-up	858	Cambodia	0.61092	0.00244	0.51410	0.70773
Jukes and Dubeck, 2015	English word identification (wpm); 24 month follow-up	1992	Kenya	0.16600	0.00101	0.10379	0.22821
Jukes and Dubeck, 2015	English letter knowledge 24 months follow-up	1992	Kenya	-0.01571		-0.08000	0.05000
Jukes and Dubeck, 2015	English passage reading fluency (wpm); 24 month follow-up	1997	Kenya	0.13613	0.00100	0.07403	0.19823
Jukes and Dubeck, 2015	English passage reading comprehension; 24 month follow-up	2002	Kenya	0.10238	0.00100	0.04039	0.16437
Jukes and Dubeck, 2015	English spelling; 24 month follow-up	1990	Kenya	0.22680	0.00101	0.16446	0.28913
Jukes and Dubeck, 2015	Beginning sounds; 9 month follow-up	2211	Kenya	0.14106	0.00091	0.08203	0.20008
Jukes and Dubeck, 2015	Receptive language; 9 month follow-up	2200	Kenya	0.19747	0.00091	0.13823	0.25671
RTI International, 2011	English Letter sound fluency	4566	Kenya	0.82258	0.00048		

	(correct letters per min)						
RTI International, 2011	English Decoding fluency (correct words per min)	4566	Kenya	0.36641	0.00045		
RTI International, 2011	English Segmenting (%)	4566	Kenya	1.21809	0.00052		
RTI International, 2011	English Reading fluency (correct words per min)	4566	Kenya	0.32243	0.00044		
RTI International, 2011	English Vocabulary	4566	Kenya	0.10638	0.00044		
RTI International, 2011	English Reading comprehension	4566	Kenya	0.51402	0.00045		
RTI International, 2011	English Sentence comprehension	4566	Kenya	0.31757	0.00044		
RTI International, 2011	English Readers able to read at greater than 30 cwpm	4566	Kenya	0.13636	0.00044		
RTI International, 2011	English Readers able to read at least at the rate of 30 cwpm	4566	Kenya	0.56621	0.00046		

Grades 4-6 sub-group

Random-Effects Model (k = 4; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0672 (SE = 0.0555)

tau (square root of estimated tau^2 value): 0.2592

```
I^2 (total heterogeneity / total variability):   98.91%
H^2 (total variability / sampling variability):  91.43
Test for Heterogeneity:
Q(df = 3) = 242.9698, p-val < .0001
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.2143   0.1304   1.6437   0.1002  -0.0412   0.4698
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub       Q     Qp   tau2     I2      H2
1   0.0829 0.0245 3.3805 0.0007  0.0348 0.1309   5.0300 0.0809 0.0011 61.2059   2.5777
2   0.2403 0.1807 1.3296 0.1837 -0.1139 0.5944 239.9793 0.0000 0.0972 99.2770 138.3054
3   0.2669 0.1690 1.5796 0.1142 -0.0643 0.5981 197.2055 0.0000 0.0848 99.0163 101.6564
4   0.2651 0.1700 1.5595 0.1189 -0.0681 0.5982 216.6516 0.0000 0.0859 99.1260 114.4212
> #-----------------------------------------------------
```

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
Abeberese, Kumler and Linden, 2011	7 month follow-up	4566	Philippines2	0.07192	0.00044	0.03088	0.11295
He, Linden and MacLeod, 2007; Year 2 programme	Machines only treatment arm	2629	India	0.10756	0.00076	0.05346	0.16166
He, Linden and MacLeod, 2007; Year 2 programme	Activities only treatment arm	2575	India	0.08723	0.00078	0.03258	0.14188

Maths test scores

```
Multivariate Meta-Analysis Model (k = 38; method: REML)
Variance Components:
           estim    sqrt  nlvls  fixed  factor
sigma^2    0.0137  0.1168     15     no   IDnum
```

Test for Heterogeneity:

Q(df = 37) = 1452.6546, p-val < .0001

> #--

>   RobustResults(fit1)

t test of coefficients:

Estimate Std. Error t value Pr(>|t|) intrcpt

0.138813   0.030818   4.5043 0.0004951 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #---

> # Sensitivity Analysis

Maths test scores sub-groups

Grades 1-3 sub-group

Random-Effects Model (k = 9; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0051 (SE = 0.0031)

tau (square root of estimated tau^2 value):   0.0716

I^2 (total heterogeneity / total variability):   86.19%

H^2 (total variability / sampling variability): 7.24

Test for Heterogeneity:

Q(df = 8) = 50.0119, p-val < .0001

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub	
0.0782	0.0266	2.9356	0.0033	0.0260	0.1304	**

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #---

> # Sensitivity Analysis

> #---

>   leave1out(fit1)

	estimate	se	zval	pval	ci.lb	ci.ub	Q	Qp	tau2	I2	H2
1	0.0872	0.0296	2.9472	0.0032	0.0292	0.1452	46.0371	0.0000	0.0056	85.8954	7.0899
2	0.0770	0.0305	2.5268	0.0115	0.0173	0.1367	48.2581	0.0000	0.0061	88.1309	8.4253

3 0.0773 0.0305 2.5337 0.0113 0.0175 0.1370 48.4588 0.0000 0.0061 88.1712 8.4539

4 0.0901 0.0278 3.2369 0.0012 0.0355 0.1447 38.8522 0.0000 0.0048 84.0275 6.2608

5 0.0839 0.0305 2.7509 0.0059 0.0241 0.1437 49.6729 0.0000 0.0060 87.3006 7.8744

Grades 4-6 sub-group

Random-Effects Model (k = 4; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0275 (SE = 0.0232)

tau (square root of estimated tau^2 value): 0.1659

I^2 (total heterogeneity / total variability): 97.05%

H^2 (total variability / sampling variability): 33.93

Test for Heterogeneity:

Q(df = 3) = 100.7261, p-val < .0001

Model Results:

estimate se zval pval ci.lb ci.ub

0.2063 0.0842 2.4507 0.0143 0.0413 0.3714 *

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #--

> # Sensitivity Analysis

> #--

> leave1out(fit1)

estimate se zval pval ci.lb ci.ub Q Qp tau2 I2 H2

1 0.2313 0.1139 2.0307 0.0423 0.0081 0.4546 94.3612 0.0000 0.0381 97.7581 44.6060

2 0.1832 0.1143 1.6030 0.1089 -0.0408 0.4073 90.6606 0.0000 0.0384 97.9455 48.6738

3 0.2699 0.0783 3.4474 0.0006 0.1164 0.4233 43.6649 0.0000 0.0175 95.2479 21.0431

4 0.1410 0.0751 1.8782 0.0604 -0.0061 0.2882 40.2629 0.0000 0.0161 95.2556 21.0774

Grades 7-11 sub-group

Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0375 (SE = 0.0413)

tau (square root of estimated tau^2 value): 0.1937

I^2 (total heterogeneity / total variability): 92.78%

H^2 (total variability / sampling variability): 13.85

Test for Heterogeneity:

Q(df = 2) = 38.0804, p-val < .0001

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
  0.1252   0.1172   1.0679   0.2856  -0.1046   0.3550
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
>   leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub       Q     Qp   tau2      I2      H2
1   0.2554 0.0477 5.3591 0.0000  0.1620 0.3488  0.3062 0.5800 0.0000  0.0000  1.0000
2   0.0881 0.1811 0.4865 0.6266 -0.2669 0.4431 31.9156 0.0000 0.0636 96.8667 31.9156
3   0.0514 0.1519 0.3383 0.7351 -0.2463 0.3491 10.8748 0.0010 0.0421 90.8044 10.8748
```

Computer Assisted Learning (CAL)

Enrolment

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0024 (SE = 0.0044)

tau (square root of estimated tau^2 value):   0.0490

I^2 (total heterogeneity / total variability):   77.06%

H^2 (total variability / sampling variability): 4.36

Test for Heterogeneity:

Q(df = 1) = 4.3594, p-val = 0.0368

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
 -0.0353   0.0393  -0.8979   0.3693  -0.1122   0.0417
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
>   leave1out(fit1)
```

```
  estimate     se    zval   pval   ci.lb   ci.ub      Q     Qp   tau2    I2    H2
1  -0.0789 0.0323 -2.4441 0.0145 -0.1421 -0.0156 0.0000 1.0000 0.0000   NaN   NaN
2   0.0001 0.0197  0.0049 0.9961 -0.0386  0.0388 0.0000 1.0000 0.0000   NaN   NaN
> #-----------------------------------------------------
```

Attendance

```
Random-Effects Model (k = 2; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0008)
tau (square root of estimated tau^2 value):      0
I^2 (total heterogeneity / total variability):   0.00%
H^2 (total variability / sampling variability):  1.00
Test for Heterogeneity:
Q(df = 1) = 0.0013, p-val = 0.9708
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.0362   0.0163   2.2278   0.0259   0.0044   0.0681       *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub      Q     Qp   tau2    I2    H2
1   0.0358 0.0200 1.7862 0.0741 -0.0035 0.0751 0.0000 1.0000 0.0000   NaN   NaN
2   0.0371 0.0278 1.3318 0.1829 -0.0175 0.0916 0.0000 1.0000 0.0000   NaN   NaN
> #-----------------------------------------------------
```

Completion

```
Random-Effects Model (k = 2; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.0108 (SE = 0.0155)
tau (square root of estimated tau^2 value):      0.1040
I^2 (total heterogeneity / total variability):   98.66%
H^2 (total variability / sampling variability):  74.59
Test for Heterogeneity:
```

```
Q(df = 1) = 74.5943, p-val < .0001
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.0742   0.0741   1.0024   0.3162  -0.0709   0.2194
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #---------------------------------------------------
> # Sensitivity Analysis
> #---------------------------------------------------
>   leave1out(fit1)
  estimate     se    zval   pval   ci.lb  ci.ub      Q     Qp   tau2    I2    H2
1   0.1488 0.0147 10.1292 0.0000  0.1200 0.1776 0.0000 1.0000 0.0000   NaN   NaN
2   0.0006 0.0089  0.0719 0.9427 -0.0167 0.0180 0.0000 1.0000 0.0000   NaN   NaN
> #---------------------------------------------------
```

Dropout

```
Random-Effects Model (k = 2; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.0032 (SE = 0.0048)
tau (square root of estimated tau^2 value):      0.0568
I^2 (total heterogeneity / total variability):   95.77%
H^2 (total variability / sampling variability):  23.62
Test for Heterogeneity:
Q(df = 1) = 23.6240, p-val < .0001
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
 -0.0425   0.0411  -1.0339   0.3012  -0.1229   0.0380
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #---------------------------------------------------
> # Sensitivity Analysis
> #---------------------------------------------------
>   leave1out(fit1)
  estimate     se    zval   pval   ci.lb   ci.ub      Q     Qp   tau2    I2    H2
```

```
1  -0.0843 0.0144 -5.8533 0.0000 -0.1125 -0.0561 0.0000 1.0000 0.0000   NaN   NaN
2  -0.0022 0.0088 -0.2457 0.8059 -0.0195  0.0152 0.0000 1.0000 0.0000   NaN   NaN
> #-----------------------------------------------------
```

Composite scores

```
Random-Effects Model (k = 6; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.0040 (SE = 0.0031)
tau (square root of estimated tau^2 value):      0.0629
I^2 (total heterogeneity / total variability):   81.13%
H^2 (total variability / sampling variability):  5.30
Test for Heterogeneity:
Q(df = 5) = 23.3072, p-val = 0.0003
Model Results:
estimate      se    zval     pval    ci.lb    ci.ub
  0.0123   0.0287   0.4278   0.6688  -0.0440   0.0685
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
  estimate     se    zval   pval   ci.lb  ci.ub      Q     Qp   tau2     I2     H2
1  -0.0008 0.0314 -0.0267 0.9787 -0.0624 0.0608 16.7642 0.0021 0.0039 80.2265 5.0573
2  -0.0034 0.0281 -0.1197 0.9047 -0.0584 0.0517 17.5988 0.0015 0.0031 78.8399 4.7259
3   0.0148 0.0357  0.4132 0.6795 -0.0552 0.0848 23.1666 0.0001 0.0053 84.7203 6.5447
4   0.0314 0.0239  1.3127 0.1893 -0.0155 0.0782 12.4184 0.0145 0.0019 68.8435 3.2096
5   0.0090 0.0355  0.2541 0.7994 -0.0606 0.0786 22.7720 0.0001 0.0053 84.5438 6.4699
6   0.0223 0.0336  0.6645 0.5063 -0.0435 0.0881 19.8812 0.0005 0.0046 82.6833 5.7748
> #-----------------------------------------------------
```

Table A6.2 a: composite scores effect sizes, subgroups and other treatment arms not included in the meta-analysis

Study	Outcome	COUNTRY	N	ESa	vara	lower	upper
Banerjee	Other follow-up period - Year 3	India	2623	0.118728	0.000764	0.064559	0.172898
Banerjee	Schooltype subgroup - Balshaki schools	India	2623	-0.016112	0.000763	-0.070235	0.038010
Barrera-Osorio	Grade subgroup - 4	Colombia	826.8	0.115275	0.002423	0.018797	0.211754
Barrera-Osorio	Grade subgroup - 5	Colombia	826.8	0.048684	0.002420	-0.047729	0.145097
Barrera-Osorio	Grade subgroup- 7	Colombia	826.8	-0.002879	0.002419	-0.099278	0.093519
Barrera-Osorio	Grade subgroup- 8	Colombia	826.8	-0.246623	0.002437	-0.343388	-0.149859
Barrera-Osorio	Grade subgroup -9	Colombia	826.8	0.313030	0.002449	0.216043	0.410017
Cristia (2012)	Baseline subgroup - low	Peru	2079	-0.047623	0.000962	-0.108423	0.013177
Cristia (2012)	Baseline subgroup - high	Peru	2021	0.031088	0.000990	-0.030573	0.092750
Banerjee	Baseline subgroup - bottom third	Indi	1962	0.123062	0.001021	0.060425	0.185700
Banerjee	Baseline subgroup - middle third	India	1844	0.103941	0.001086	0.039348	0.168534
Banerjee	Baseline subgroup - top third	India	1926	0.085302	0.001039	0.022113	0.148491
Sharma	Other follow-up period - Year 2	Nepal	2623	-0.056610	0.000763	-0.110742	-0.002477
Sharma	Baseline subgroup - bottom quintile	Nepal	826.8	-0.133112	0.002424	-0.229617	-0.036606
Sharma	Baseline subgroup - middle quintile	Nepal	826.8	-0.170565	0.002428	-0.267139	-0.073991
Sharma	Baseline subgroup - top quintile	Nepal	826.8	-0.106755	0.002422	-0.203223	-0.010288

Composite Boys

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0012) tau (square root of estimated tau^2 value): 0

I^2 (total heterogeneity / total variability): 0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 1) = 0.4941, p-val = 0.4821

Model Results:

estimate se zval pval ci.lb ci.ub

0.0168 0.0208 0.8077 0.4193 -0.0239 0.0575

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> #---

> # Sensitivity Analysis

> #---

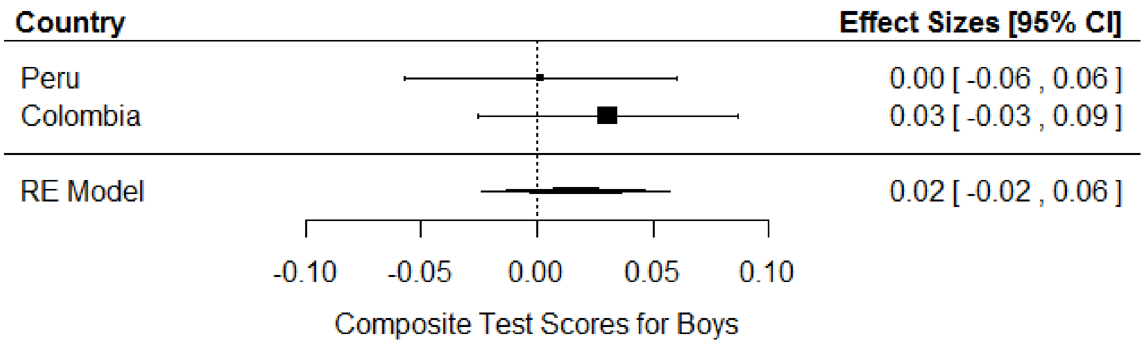
> leave1out(fit1)

estimate se zval pval ci.lb ci.ub Q Qp tau2 I2 H2

1 0.0014 0.0301 0.0469 0.9626 -0.0577 0.0605 0.0000 1.0000 0.0000 NaN NaN

2 0.0306 0.0286 1.0697 0.2848 -0.0255 0.0868 0.0000 1.0000 0.0000 NaN NaN

> #---5 0.0208 0.0443 0.4689 0.6391 -0.0660 0.1075 19.6911 0.0002 0.0067 86.4474 7.3786



Composite Girls

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0002 (SE = 0.0015)

tau (square root of estimated tau^2 value): 0.0140

I^2 (total heterogeneity / total variability): 18.47%

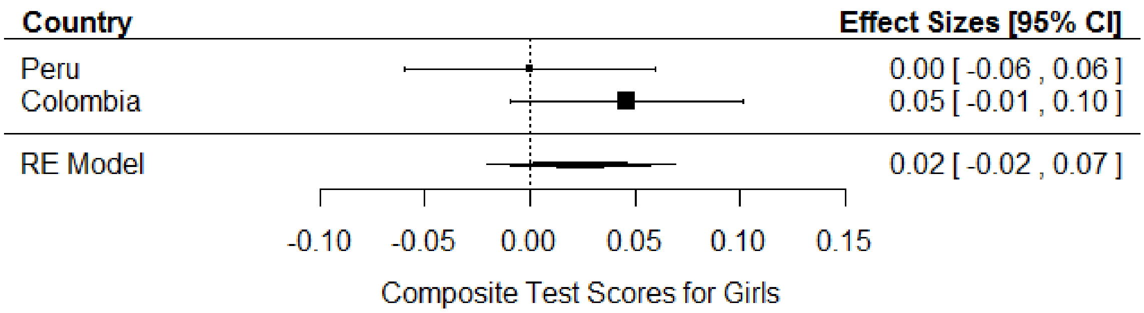
H^2 (total variability / sampling variability): 1.23

Test for Heterogeneity:

Q(df = 1) = 1.2265, p-val = 0.2681

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
 0.0243   0.0230   1.0566   0.2907  -0.0207   0.0693
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #---------------------------------------------------
> # Sensitivity Analysis
> #---------------------------------------------------
>  leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub     Q     Qp   tau2    I2    H2
1   0.0000 0.0303 0.0000 1.0000 -0.0595 0.0595 0.0000 1.0000 0.0000   NaN   NaN
2   0.0460 0.0284 1.6219 0.1048 -0.0096 0.1016 0.0000 1.0000 0.0000   NaN   NaN
```



Composite scores Grade 2

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0015)

tau (square root of estimated tau^2 value):   0

I^2 (total heterogeneity / total variability):   0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 1) = 0.3254, p-val = 0.5684

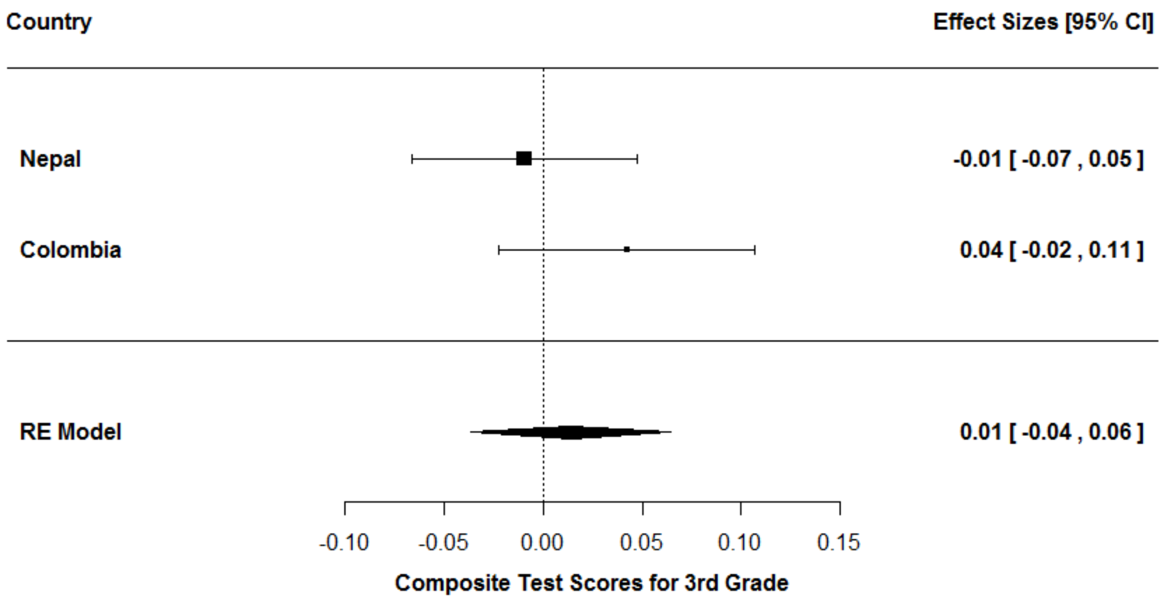
Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
 -0.0628   0.0230  -2.7277   0.0064  -0.1080  -0.0177      **
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
  estimate     se    zval   pval   ci.lb   ci.ub     Q     Qp   tau2    I2    H2
1  -0.0745 0.0308 -2.4208 0.0155 -0.1347 -0.0142 0.0000 1.0000 0.0000   NaN   NaN
2  -0.0480 0.0348 -1.3804 0.1675 -0.1161  0.0201 0.0000 1.0000 0.0000   NaN   NaN
> #-----------------------------------------------------
```

Composite scores Grade 3

```
Random-Effects Model (k = 2; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.0004 (SE = 0.0019)
tau (square root of estimated tau^2 value):      0.0196
I^2 (total heterogeneity / total variability):   28.63%
H^2 (total variability / sampling variability):  1.40
Test for Heterogeneity:
Q(df = 1) = 1.4011, p-val = 0.2365
Model Results:
estimate      se     zval     pval    ci.lb     ci.ub
  0.0141   0.0259   0.5442   0.5863  -0.0366   0.0647
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
  estimate     se    zval   pval   ci.lb  ci.ub     Q     Qp   tau2    I2    H2
1  -0.0095 0.0289 -0.3282 0.7427 -0.0662 0.0472 0.0000 1.0000 0.0000   NaN   NaN
2   0.0424 0.0330  1.2871 0.1981 -0.0222 0.1070 0.0000 1.0000 0.0000   NaN   NaN
> #-----------------------------------------------------
```



Composite scores Grade 6

Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0051 (SE = 0.0065)

tau (square root of estimated tau^2 value): 0.0713

I^2 (total heterogeneity / total variability): 78.33%

H^2 (total variability / sampling variability): 4.62

Test for Heterogeneity:

Q(df = 2) = 9.6159, p-val = 0.0082

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub
0.0253	0.0466	0.5440	0.5865	-0.0659	0.1166

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #--

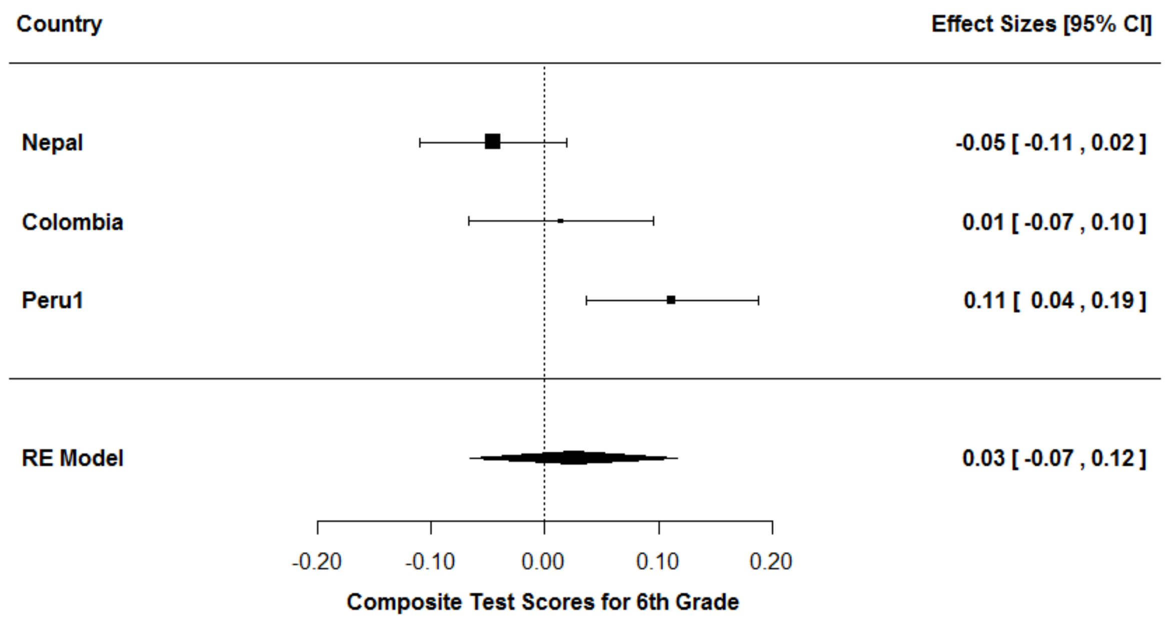
> # Sensitivity Analysis

> #--

> leave1out(fit1)

	estimate	se	zval	pval	ci.lb	ci.ub	Q	Qp	tau2	I2	H2
1	-0.0210	0.0293	-0.7162	0.4738	-0.0784	0.0365	1.2663	0.2605	0.0004	21.0280	1.2663
2	0.0318	0.0786	0.4047	0.6857	-0.1223	0.1859	9.5958	0.0020	0.0111	89.5788	9.5958
3	0.0640	0.0488	1.3109	0.1899	-0.0317	0.1597	2.9816	0.0842	0.0032	66.4614	2.9816

> #---



Language scores

Random-Effects Model (k = 13; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0130 (SE = 0.0060)

tau (square root of estimated tau^2 value):   0.1141

I^2 (total heterogeneity / total variability):   90.08%

H^2 (total variability / sampling variability): 10.08

Test for Heterogeneity:

Q(df = 12) = 102.1851, p-val < .0001

Model Results:

estimate   se   zval   pval   ci.lb   ci.ub

 -0.0134   0.0337   -0.3972   0.6912   -0.0794   0.0527

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #---

> # Sensitivity Analysis

> #---

> leave1out(fit1)

  estimate   se   zval   pval   ci.lb ci.ub   Q   Qp   tau2   I2   H2

1   -0.0190 0.0359 -0.5293 0.5966 -0.0894 0.0514 100.6574 0.0000 0.0138 90.9881 11.0965

2 -0.0171 0.0366 -0.4667 0.6407 -0.0888 0.0547 100.7481 0.0000 0.0143 90.5464 10.5780

3 -0.0154 0.0368 -0.4183 0.6757 -0.0874 0.0567 101.8880 0.0000 0.0144 90.6266 10.6685

4 -0.0064 0.0355 -0.1804 0.8569 -0.0759 0.0631 99.3576 0.0000 0.0135 90.8532 10.9328

5 -0.0127 0.0368 -0.3464 0.7291 -0.0849 0.0594 101.9723 0.0000 0.0145 90.6520 10.6974

6 0.0150 0.0197 0.7648 0.4444 -0.0235 0.0536 34.0421 0.0004 0.0030 68.3056 3.1551

7 -0.0199 0.0359 -0.5541 0.5795 -0.0903 0.0505 99.2598 0.0000 0.0138 90.7467 10.8070

8 -0.0205 0.0358 -0.5715 0.5676 -0.0907 0.0497 98.4510 0.0000 0.0137 90.6413 10.6852

9 -0.0146 0.0368 -0.3967 0.6916 -0.0867 0.0575 102.1246 0.0000 0.0144 90.6464 10.6910

10 -0.0259 0.0341 -0.7602 0.4471 -0.0928 0.0409 84.0264 0.0000 0.0122 89.2605 9.3115

11 -0.0151 0.0368 -0.4095 0.6822 -0.0871 0.0570 102.0061 0.0000 0.0144 90.6359 10.6791

12 -0.0175 0.0361 -0.4859 0.6271 -0.0883 0.0532 101.4499 0.0000 0.0140 91.1241 11.2665

13 -0.0052 0.0356 -0.1468 0.8833 -0.0751 0.0646 92.9780 0.0000 0.0135 90.2098 10.2143

> #--

Table A6.2 b: language scores effect sizes, subgroups and other treatment arms not included in the meta-analysis

Study	Outcome	COUNTRY	N	ESa	vara	lower	upper
Carillo	Other follow-up period-test 2 (earlier)	Ecuador	720	0.153456	0.002786	0.050003	0.256909
Carillo	Other follow-up period - test 1 (earlier)	Ecuador	720	0.188824	0.002790	0.085293	0.292355
Banerjee	Other follow-up period - Year 3	India	2006	-0.003204	0.000997	-0.065092	0.058684
Banerjee	Schooltype subgroup - Balshaki schools	India	2006	-0.003943	0.000997	-0.065831	0.057945
Barrera-Osorio	Grade subgroup - 7	Colombia	772	-0.009112	0.002591	-0.108873	0.090650
Barrera-Osorio	Grade subgroup- 9	Colombia	517	0.228537	0.003894	0.106234	0.350841
Banerjee	Baseline subgroup - bottom third	India	1962	0.043630	0.001020	-0.018955	0.106216
Banerjee	Baseline subgroup - middle third	India	1844	0.002875	0.001085	-0.061674	0.067424
Banerjee	Baseline subgroup - top third	India	1926	-0.018567	0.001038	-0.081728	0.044595
Cristia (2012)	Baseline subgroup - low	Peru	2006	-0.042709	0.000997	-0.104604	0.019186
Cristia (2012)	Baseline subgroup - high	Peru	2006	-0.015015	0.000997	-0.076904	0.046874
Humpage	Grade subgroup - 8	Peru	517	-0.088130	0.003872	-0.210096	0.033835
Humpage	Grade subgroup - 7	Peru	517	0.179124	0.003884	0.056973	0.301274
Sharma	Other follow-up period - Year 2	Nepal	2006	-0.108704	0.000998	-0.170638	-0.046771
Sharma	Baseline subgroup - bottom quintile	Nepal	517	-0.282211	0.003907	-0.404723	-0.159700
Sharma	Baseline subgroup - middle quintile	Nepal	517	-0.231946	0.003894	-0.354261	-0.109631
Sharma	Baseline subgroup - top quintile	Nepal	517	-0.167819	0.003882	-0.289940	-0.045699

Language scores Boys

Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0029 (SE = 0.0039)

tau (square root of estimated tau^2 value):   0.0541

I^2 (total heterogeneity / total variability):   75.53%

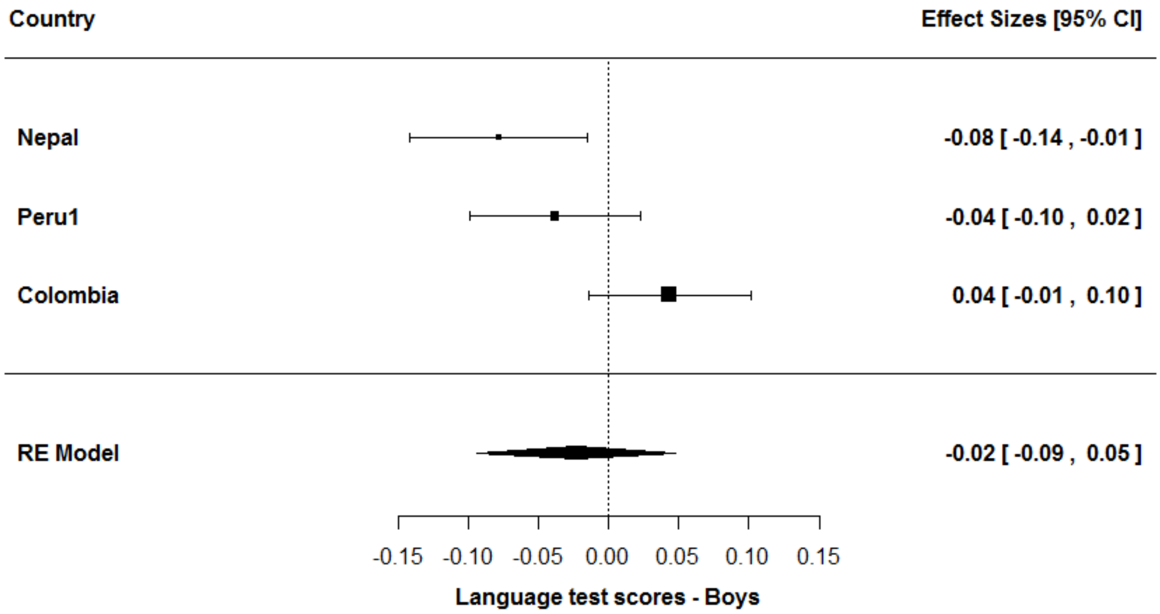
H^2 (total variability / sampling variability): 4.09

Test for Heterogeneity:

Q(df = 2) = 8.2714, p-val = 0.0160

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
 -0.0232   0.0360  -0.6451   0.5189  -0.0937   0.0473
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #------------------------------------------------------
> # Sensitivity Analysis
> #------------------------------------------------------
>  leave1out(fit1)
  estimate     se    zval   pval   ci.lb   ci.ub      Q     Qp   tau2      I2     H2
1  -0.0572 0.0223 -2.5633 0.0104 -0.1010 -0.0135 0.8060 0.3693 0.0000  0.0000 1.0000
2  -0.0165 0.0608 -0.2717 0.7858 -0.1357  0.1026 7.7895 0.0053 0.0064 87.1622 7.7895
3   0.0034 0.0407  0.0839 0.9331 -0.0764  0.0833 3.6445 0.0563 0.0024 72.5615 3.6445
> #------------------------------------------------------
```



Language scores Girls

Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0049 (SE = 0.0058)

tau (square root of estimated tau^2 value): 0.0697

I^2 (total heterogeneity / total variability): 83.47%

H^2 (total variability / sampling variability): 6.05

Test for Heterogeneity:

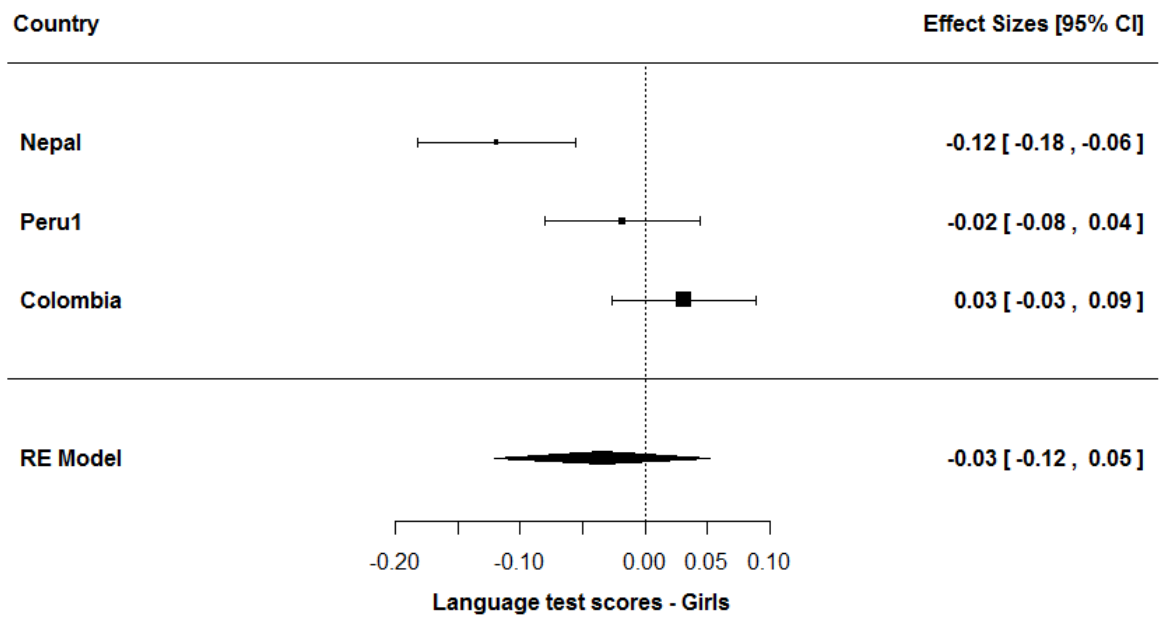
Q(df = 2) = 12.0391, p-val = 0.0024

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
 -0.0346  0.0440  -0.7867   0.4314  -0.1209   0.0517
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
>  leave1out(fit1)
  estimate     se    zval   pval   ci.lb  ci.ub       Q     Qp   tau2      I2      H2
1  -0.0683 0.0504 -1.3540 0.1757 -0.1671 0.0305  4.9972 0.0254 0.0041 79.9887  4.9972
2  -0.0434 0.0749 -0.5800 0.5619 -0.1902 0.1033 11.8001 0.0006 0.0103 91.5255 11.8001
```

3   0.0077 0.0244 0.3168 0.7514 -0.0401 0.0556 1.2942 0.2553 0.0003 22.7298 1.2942

> #---



Language scores Grade 2

Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0019)

tau (square root of estimated tau^2 value):   0

I^2 (total heterogeneity / total variability):   0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 2) = 1.4404, p-val = 0.4867

Model Results:

estimate   se   zval   pval   ci.lb   ci.ub

 -0.0883   0.0252 -3.5004   0.0005 -0.1377 -0.0389   ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #---

> # Sensitivity Analysis

> #---

>   leave1out(fit1)

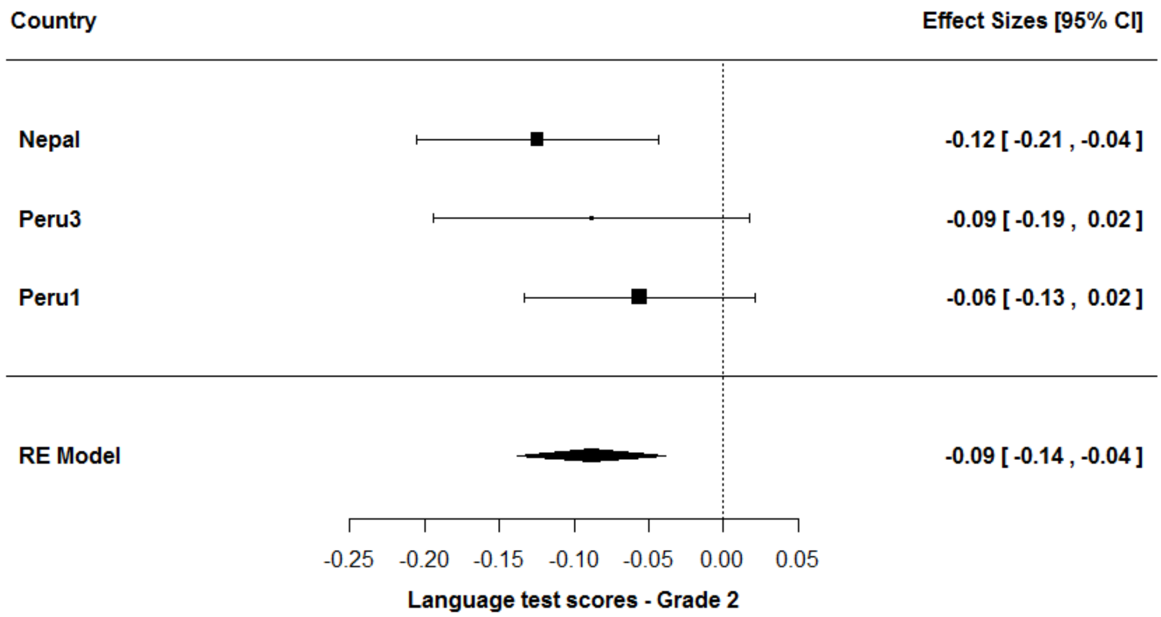
  estimate   se   zval   pval   ci.lb   ci.ub   Q   Qp   tau2   I2   H2

1 -0.1110 0.0329 -3.3740 0.0007 -0.1755 -0.0465 0.2847 0.5936 0.0000 0.0000 1.0000

2 -0.0671 0.0318 -2.1103 0.0348 -0.1293 -0.0048 0.2327 0.6295 0.0000 0.0000 1.0000

3 -0.0889 0.0343 -2.5950 0.0095 -0.1560 -0.0218 1.4404 0.2301 0.0007 30.5743 1.4404

> #---



Language scores Grade 3

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0016 (SE = 0.0043)

tau (square root of estimated tau^2 value):   0.0403

I^2 (total heterogeneity / total variability):   54.11%

H^2 (total variability / sampling variability): 2.18

Test for Heterogeneity:

Q(df = 1) = 2.1790, p-val = 0.1399

Model Results:

estimate   se   zval   pval   ci.lb   ci.ub

 0.0067   0.0388   0.1716   0.8637 -0.0693   0.0826

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #---

> # Sensitivity Analysis

> #---

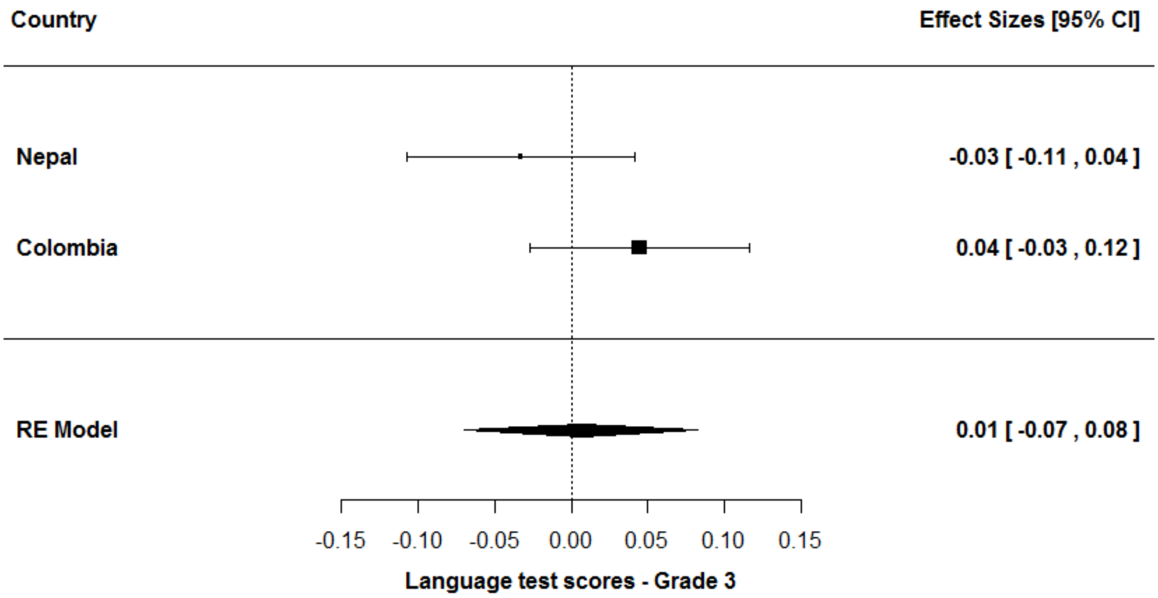
>   leave1out(fit1)

 estimate   se   zval   pval   ci.lb ci.ub   Q   Qp   tau2   I2   H2

1 -0.0328 0.0379 -0.8667 0.3861 -0.1071 0.0414 0.0000 1.0000 0.0000   NaN   NaN

2   0.0447 0.0364 1.2285 0.2192 -0.0266 0.1161 0.0000 1.0000 0.0000   NaN   NaN

> #---



Language scores Grade 4

Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0247 (SE = 0.0291)

tau (square root of estimated tau^2 value):   0.1571

I^2 (total heterogeneity / total variability):   85.24%

H^2 (total variability / sampling variability): 6.77

Test for Heterogeneity:

Q(df = 2) = 13.9303, p-val = 0.0009

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub
0.0285	0.0984	0.2893	0.7724	-0.1645	0.2214

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #---

> # Sensitivity Analysis

> #---

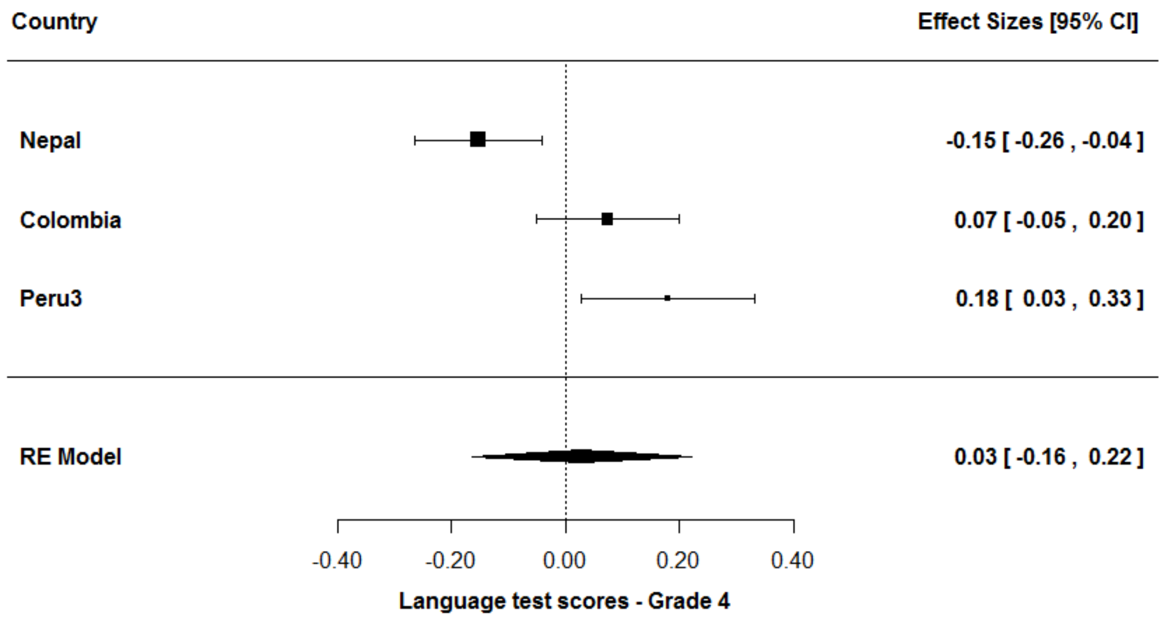
> leave1out(fit1)

	estimate	se	zval	pval	ci.lb	ci.ub	Q	Qp	tau2	I2	H2
1	0.0090	0.1659	0.0544	0.9567	-0.3161	0.3341	11.9362	0.0006	0.0504	91.6221	11.9362

2   0.1173 0.0517 2.2670 0.0234 0.0159 0.2187 1.0984 0.2946 0.0005 8.9600 1.0984

3 -0.0412 0.1133 -0.3636 0.7162 -0.2633 0.1809 7.0579 0.0079 0.0221 85.8314 7.0579

> #---



Language scores Grade 5

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0188 (SE = 0.0308)

tau (square root of estimated tau^2 value):   0.1369

I^2 (total heterogeneity / total variability):   86.09%

H^2 (total variability / sampling variability): 7.19

Test for Heterogeneity:

Q(df = 1) = 7.1894, p-val = 0.0073

Model Results:

estimate   se   zval   pval   ci.lb   ci.ub

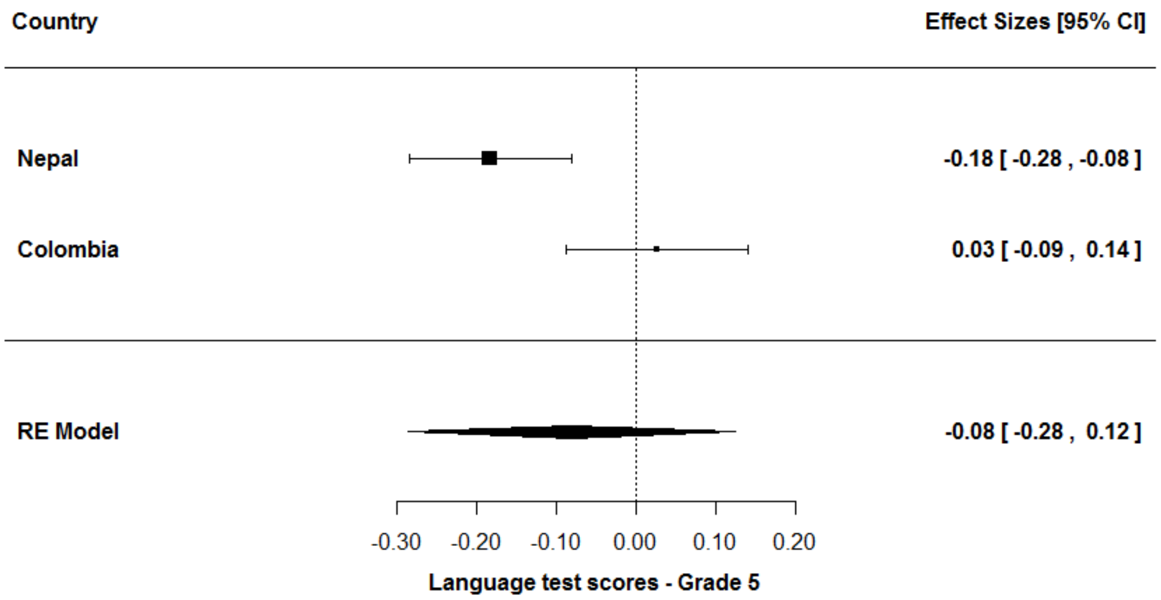
 -0.0801   0.1043 -0.7678   0.4426 -0.2846   0.1244

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #---

> # Sensitivity Analysis

```
> #------------------------------------------------------
>  leave1out(fit1)
  estimate     se    zval   pval   ci.lb   ci.ub     Q     Qp   tau2    I2    H2
1  -0.1828 0.0518 -3.5299 0.0004 -0.2843 -0.0813 0.0000 1.0000 0.0000   NaN   NaN
2   0.0259 0.0581  0.4459 0.6557 -0.0880  0.1398 0.0000 1.0000 0.0000   NaN   NaN
> #------------------------------------------------------
```



Language Grade 6

Random-Effects Model (k = 4; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0043 (SE = 0.0051)

tau (square root of estimated tau^2 value):   0.0657

I^2 (total heterogeneity / total variability):   68.97%

H^2 (total variability / sampling variability): 3.22

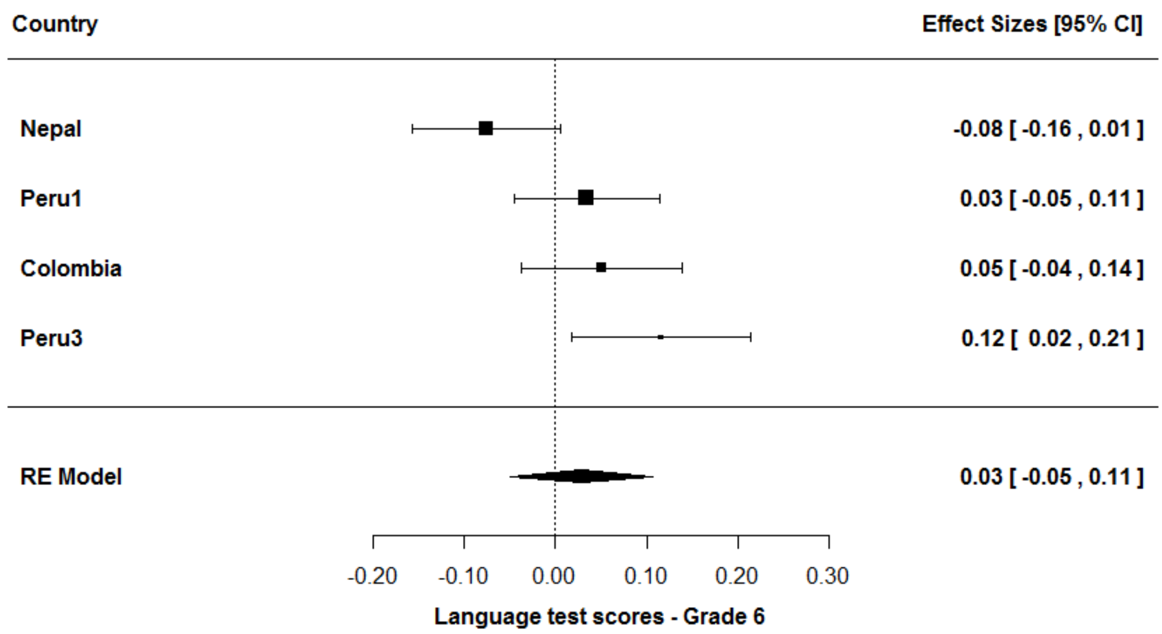
Test for Heterogeneity:

Q(df = 3) = 9.6001, p-val = 0.0223

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
  0.0283   0.0396   0.7138   0.4753  -0.0494   0.1059
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #------------------------------------------------------
> # Sensitivity Analysis
```

```
> #----------------------------------------------------
>  leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub     Q     Qp   tau2     I2     H2
1   0.0275 0.0567 0.4858 0.6271 -0.0836 0.1387 9.4928 0.0087 0.0076 78.6037 4.6737
2   0.0219 0.0549 0.3992 0.6897 -0.0857 0.1296 9.0879 0.0106 0.0071 78.6850 4.6915
3   0.0020 0.0399 0.0509 0.9594 -0.0763 0.0803 5.3575 0.0686 0.0030 62.5930 2.6733
4   0.0612 0.0258 2.3689 0.0178  0.0106 0.1119 1.6841 0.4308 0.0000  0.0000 1.0000
> #----------------------------------------------------
```



Language scores Grade 8

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0120 (SE = 0.0214)

tau (square root of estimated tau^2 value):   0.1093

I^2 (total heterogeneity / total variability):   78.81%

H^2 (total variability / sampling variability): 4.72

Test for Heterogeneity:

Q(df = 1) = 4.7202, p-val = 0.0298

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
 -0.1514   0.0870  -1.7397   0.0819  -0.3220   0.0192       .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> #---------------------------------------------------
> # Sensitivity Analysis
> #---------------------------------------------------
>  leave1out(fit1)
  estimate    se   zval   pval   ci.lb   ci.ub     Q     Qp   tau2    I2    H2
1  -0.0668 0.0529 -1.2629 0.2066 -0.1704  0.0368 0.0000 1.0000 0.0000   NaN   NaN
2  -0.2409 0.0603 -3.9979 0.0001 -0.3590 -0.1228 0.0000 1.0000 0.0000   NaN   NaN
> #---------------------------------------------------
```

Math scores

```
Multivariate Meta-Analysis Model (k = 19; method: REML)
Variance Components:
            estim    sqrt  nlvls  fixed  factor
sigma^2    0.0091  0.0953     18     no   IDnum
Test for Heterogeneity:
Q(df = 18) = 140.8256, p-val < .0001
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.0655   0.0245   2.6728   0.0075   0.0175   0.1135      **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RobustResults(fit1)
t test of coefficients:
        Estimate Std. Error t value Pr(>|t|)
intrcpt 0.065464   0.024489  2.6733  0.01605 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table A6.2 c: Math scores effect sizes, subgroups and other treatment arms not included in the meta-analysis

Study	Outcome	COUNTRY	N	ESa	vara	lower	upper
Sharma	Other follow-up period-grade 2 year 2	Nepal	829	-0.078304	0.002414	-0.174611	0.018004
Sharma	Other follow-up period-grade 6 year 2	Nepal	798	-0.002380	0.002506	-0.100503	0.095743
Barrera-Osorio	Grade subgroup -7	Colombia	772	0.012728	0.002591	-0.087034	0.112491
Barrera-Osorio	Grade subgroup -9	Colombia	609.5	0.224990	0.003302	0.112360	0.337620
Linden (out of school)	Grade subgroup -2	India	631	0.033871	0.003170	-0.076483	0.144225
Cristia (2012)	Grade subgroup -2	Peru	1426	-0.034169	0.001403	-0.107577	0.039238
Berlinksi	Baseline subgroup - low	Costa Rica	1663	-0.094273	0.001204	-0.162282	-0.026264
Berlinksi	Laptop Treatment Arm Baseline subgroup - low	Costa Rica	1663	-0.137180	0.001205	-0.205232	-0.069129
Berlinksi	Baseline subgroup - high	Costa Rica	1663	-0.045455	0.001203	-0.113435	0.022525
Cristia (2012)	Baseline subgroup - low	Peru	1885	-0.043863	0.001061	-0.107714	0.019988
Cristia (2012)	Baseline subgroup - high	Peru	1885	0.064917	0.001062	0.001057	0.128777

Math scores Boys

Random-Effects Model (k = 6; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0178 (SE = 0.0124)

tau (square root of estimated tau^2 value): 0.1333

I^2 (total heterogeneity / total variability): 91.85%

H^2 (total variability / sampling variability): 12.27

Test for Heterogeneity:

Q(df = 5) = 40.8743, p-val < .0001

Model Results:

estimate se zval pval ci.lb ci.ub

-0.0233 0.0572 -0.4081 0.6832 -0.1354 0.0887

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> #---

> # Sensitivity Analysis

> #---

> leave1out(fit1)

```
  estimate     se    zval   pval   ci.lb  ci.ub       Q     Qp   tau2      I2      H2
1  -0.0367 0.0692 -0.5301 0.5961 -0.1724 0.0990 37.9082 0.0000 0.0220 92.5481 13.4194
2  -0.0306 0.0706 -0.4338 0.6644 -0.1690 0.1078 40.4213 0.0000 0.0229 92.8388 13.9642
3  -0.0596 0.0515 -1.1566 0.2474 -0.1605 0.0414 27.1055 0.0000 0.0116 88.9376  9.0396
4   0.0197 0.0416  0.4731 0.6362 -0.0618 0.1011 16.8581 0.0021 0.0070 82.9316  5.8588
5  -0.0073 0.0671 -0.1091 0.9131 -0.1389 0.1242 36.4981 0.0000 0.0208 93.3984 15.1479
6  -0.0259 0.0710 -0.3648 0.7153 -0.1652 0.1133 40.8568 0.0000 0.0232 93.0623 14.4139
```

> #---

Table A6.2 d: Math scores Boys effect sizes, subgroups and other treatment arms not included in the meta-analysis

Study	Outcome	COUNTRY	N	ESa	vara	lower	upper
Berlinksi	Laptop treatment arm	Costa Rica	812	-0.235372	0.002480	-0.332981	-0.137763

Math scores Girls

Random-Effects Model (k = 6; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0166 (SE = 0.0117)

tau (square root of estimated tau^2 value): 0.1287

I^2 (total heterogeneity / total variability): 91.24%

H^2 (total variability / sampling variability): 11.41

Test for Heterogeneity:

Q(df = 5) = 40.5315, p-val < .0001

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
 -0.0572   0.0555  -1.0323   0.3020  -0.1659   0.0514
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #---------------------------------------------------
> # Sensitivity Analysis
> #---------------------------------------------------
>   leave1out(fit1)
  estimate    se    zval    pval   ci.lb  ci.ub       Q      Qp   tau2      I2      H2
1  -0.0738 0.0662 -1.1146 0.2650 -0.2035 0.0560 37.6250 0.0000 0.0198 91.7244 12.0837
2  -0.0792 0.0636 -1.2457 0.2129 -0.2039 0.0454 33.4124 0.0000 0.0182 91.0294 11.1476
3  -0.0776 0.0632 -1.2274 0.2197 -0.2014 0.0463 38.3015 0.0000 0.0183 92.6780 13.6575
4  -0.0069 0.0285 -0.2429 0.8080 -0.0627 0.0489 10.4477 0.0335 0.0025 63.3444  2.7281
5  -0.0465 0.0665 -0.6988 0.4847 -0.1769 0.0839 36.9586 0.0000 0.0203 93.1489 14.5962
6  -0.0612 0.0689 -0.8887 0.3742 -0.1963 0.0738 40.3226 0.0000 0.0217 92.4609 13.2642
> #---------------------------------------------------
```

Table A6.2 e: Math scores Girls effect sizes, subgroups and other treatment arms not included in the meta-analysis

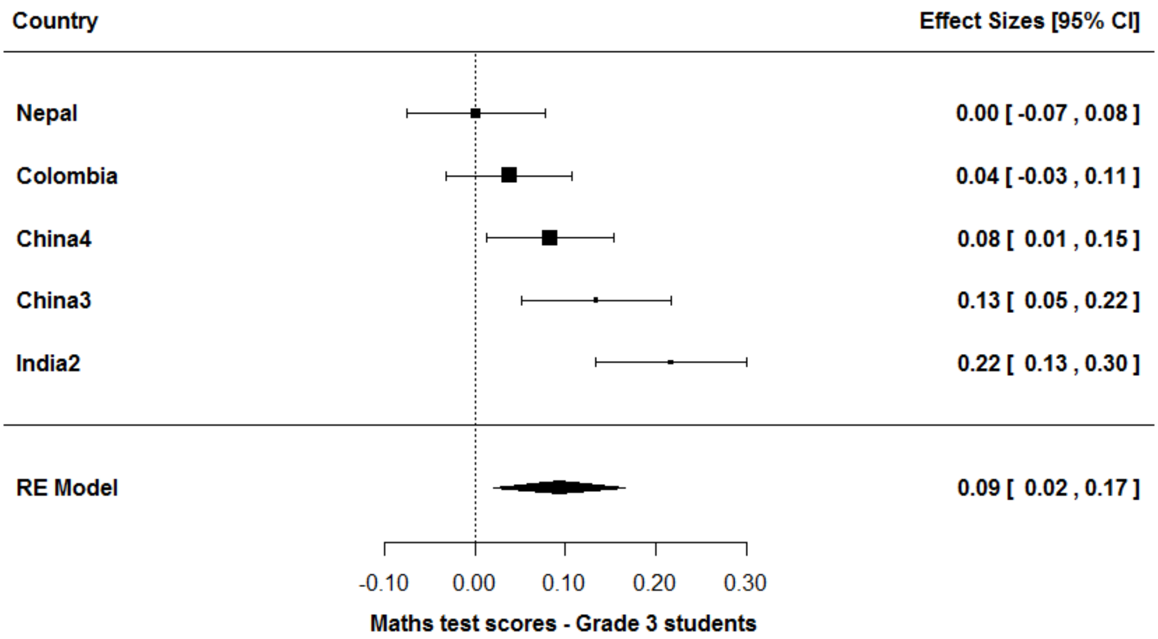
Study	Outcome	COUNTRY	N	ESa	vara	lower	upper
Berlinksi	Laptop treatment arm	Costa Rica	850	-0.259306	0.002373	-0.354779	-0.163833

Math scores Grade 3

```
Random-Effects Model (k = 5; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.0054 (SE = 0.0049)
tau (square root of estimated tau^2 value):      0.0734
I^2 (total heterogeneity / total variability):   78.21%
H^2 (total variability / sampling variability):  4.59
Test for Heterogeneity:
Q(df = 4) = 17.4677, p-val = 0.0016
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.0931   0.0372   2.5061   0.0122   0.0203   0.1659      *
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
>  leave1out(fit1)
  estimate     se    zval    pval   ci.lb  ci.ub       Q     Qp   tau2      I2     H2
1   0.0961 0.0484 1.9837 0.0473  0.0011 0.1910 17.4551 0.0006 0.0078 83.2562 5.9724
2   0.1077 0.0449 2.3992 0.0164  0.0197 0.1957 14.9744 0.0018 0.0065 80.5172 5.1327
3   0.0628 0.0273 2.2957 0.0217  0.0092 0.1163  6.2031 0.1021 0.0015 51.7144 2.0710
4   0.0836 0.0460 1.8195 0.0688 -0.0065 0.1737 15.9886 0.0011 0.0070 82.9035 5.8491
5   0.1157 0.0381 3.0411 0.0024  0.0411 0.1903 11.3181 0.0101 0.0043 73.9793 3.8431
> #----------------------------------------------------
```



Math scores Grade 4

Random-Effects Model (k = 4; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0133 (SE = 0.0136)

tau (square root of estimated tau^2 value):   0.1151

I^2 (total heterogeneity / total variability):   80.00%

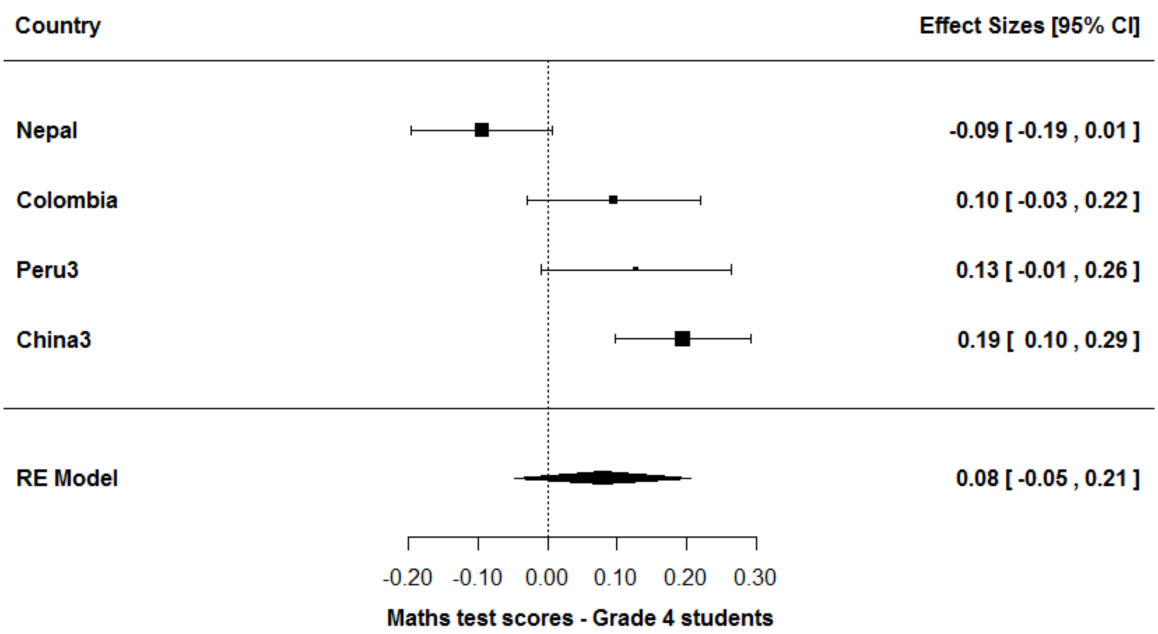
H^2 (total variability / sampling variability): 5.00

Test for Heterogeneity:

Q(df = 3) = 17.1568, p-val = 0.0007

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
  0.0794   0.0646   1.2293   0.2190  -0.0472   0.2061
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #---------------------------------------------------
> # Sensitivity Analysis
> #---------------------------------------------------
>  leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub       Q      Qp   tau2      I2     H2
1   0.0749 0.0891 0.8404 0.4007 -0.0998 0.2495 17.0443 0.0002 0.0205 86.7508 7.5476
2   0.0373 0.0713 0.5230 0.6010 -0.1024 0.1770  8.6304 0.0134 0.0114 75.3048 4.0494
3   0.0650 0.0861 0.7552 0.4501 -0.1037 0.2337 16.4960 0.0003 0.0192 86.5959 7.4604
4   0.1499 0.0340 4.4022 0.0000  0.0832 0.2166  1.6369 0.4411 0.0000  0.0000 1.0000
> #---------------------------------------------------
```



Math scores Grade 5

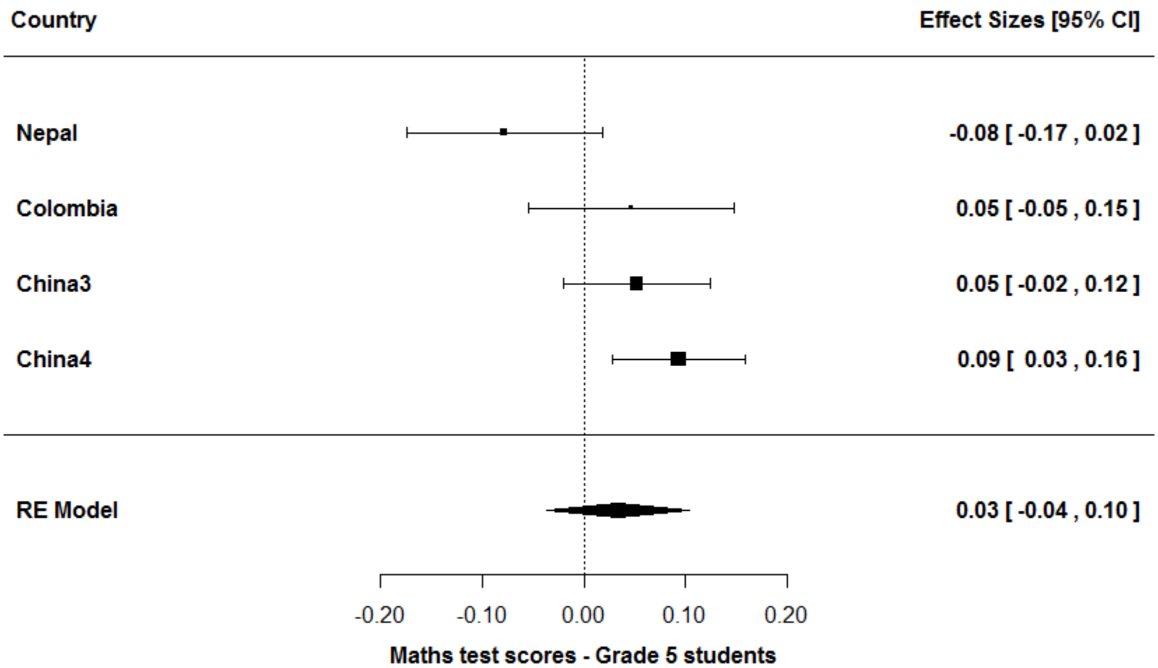
Random-Effects Model (k = 4; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0033 (SE = 0.0042)

tau (square root of estimated tau^2 value):   0.0574

I^2 (total heterogeneity / total variability):   65.71%

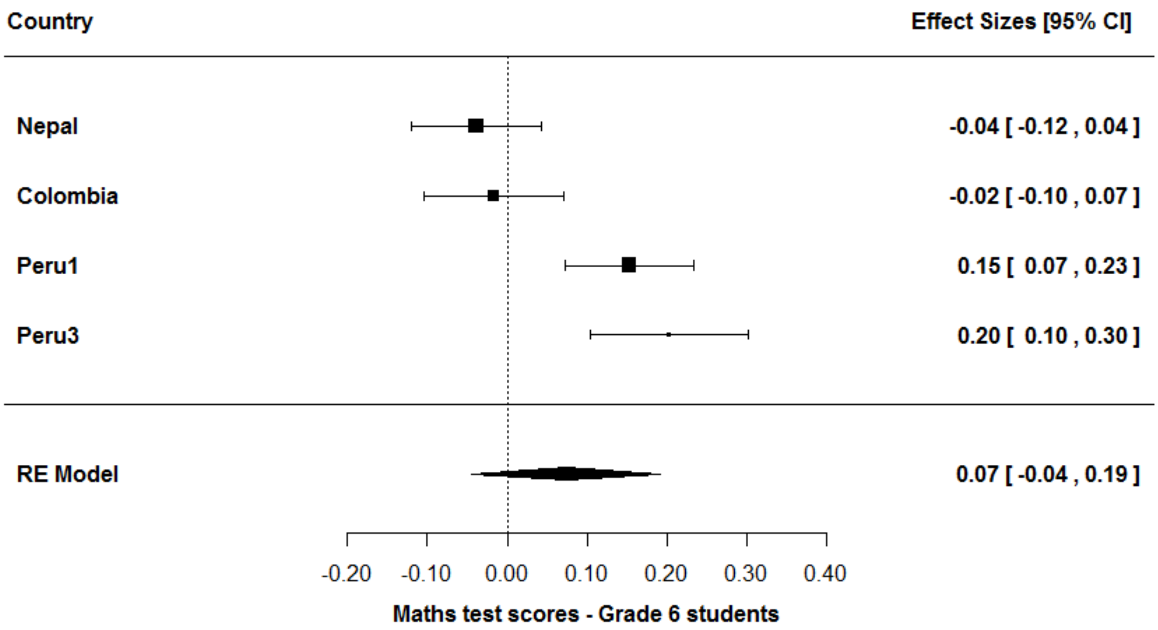
```
H^2 (total variability / sampling variability):  2.92

Test for Heterogeneity:

Q(df = 3) = 8.4006, p-val = 0.0384

Model Results:

estimate      se     zval     pval    ci.lb    ci.ub

  0.0336   0.0357   0.9404   0.3470  -0.0364   0.1035

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #---------------------------------------------------

> # Sensitivity Analysis

> #---------------------------------------------------

>   leave1out(fit1)

  estimate     se   zval   pval   ci.lb  ci.ub      Q      Qp   tau2      I2     H2

1   0.0095 0.0416 0.2289 0.8189 -0.0720 0.0911 4.9878 0.0826 0.0031 60.1957 2.5123

2   0.0276 0.0493 0.5589 0.5762 -0.0691 0.1242 8.3979 0.0150 0.0057 78.9331 4.7468

3   0.0244 0.0516 0.4724 0.6366 -0.0768 0.1256 8.3358 0.0155 0.0060 75.4125 4.0671

4   0.0692 0.0223 3.1070 0.0019  0.0255 0.1128 0.9265 0.6292 0.0000  0.0000 1.0000

> #---------------------------------------------------
```



Math scores Grade 6

Random-Effects Model (k = 4; tau^2 estimator: REML)

```
tau^2 (estimated amount of total heterogeneity): 0.0125 (SE = 0.0118)
tau (square root of estimated tau^2 value):      0.1118
I^2 (total heterogeneity / total variability):   86.55%
H^2 (total variability / sampling variability):  7.44
Test for Heterogeneity:
Q(df = 3) = 21.8138, p-val < .0001
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.0739   0.0602   1.2288   0.2191  -0.0440   0.1918
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub      Q     Qp   tau2     I2     H2
1   0.1043 0.0739 1.4117 0.1580 -0.0405 0.2490 17.1008 0.0002 0.0144 88.1963 8.4719
2   0.0471 0.0764 0.6172 0.5371 -0.1026 0.1968 15.6886 0.0004 0.0154 88.2073 8.4798
3   0.0331 0.0611 0.5413 0.5883 -0.0867 0.1529 12.9219 0.0016 0.0094 84.0172 6.2567
4   0.1124 0.0662 1.6982 0.0895 -0.0173 0.2421 12.5289 0.0019 0.0111 84.4612 6.4355
> #-----------------------------------------------------
```



Math scores Grade 8

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0029 (SE = 0.0087)

tau (square root of estimated tau^2 value): 0.0542

I^2 (total heterogeneity / total variability): 47.86%

H^2 (total variability / sampling variability): 1.92

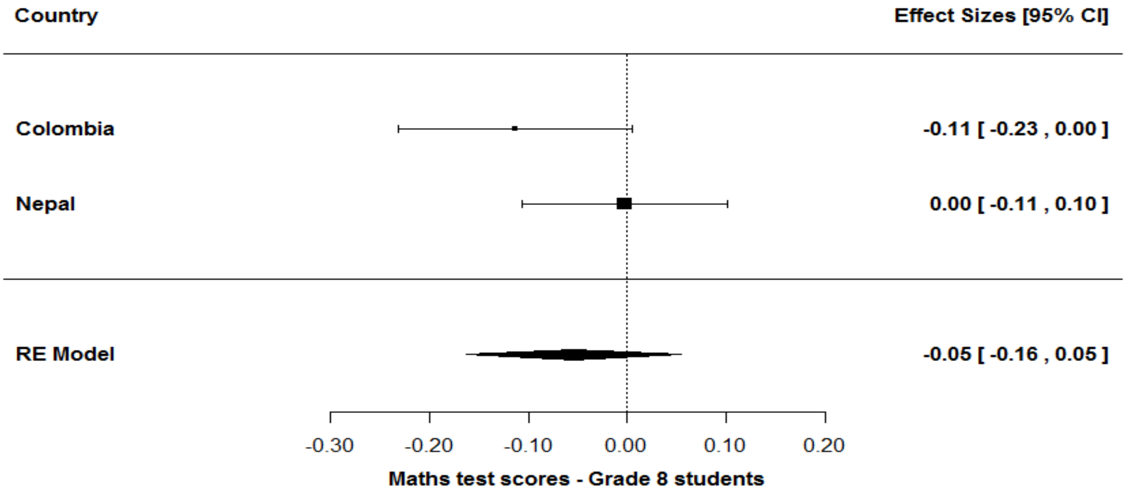
Test for Heterogeneity:

Q(df = 1) = 1.9178, p-val = 0.1661

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub
-0.0541	0.0553	-0.9783	0.3279	-0.1624	0.0542

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



Math scores Bottom Tercile

Random-Effects Model (k = 5; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0637 (SE = 0.0504)

tau (square root of estimated tau^2 value):   0.2524

I^2 (total heterogeneity / total variability):   89.58%

H^2 (total variability / sampling variability): 9.60

Test for Heterogeneity:

Q(df = 4) = 35.8518, p-val < .0001

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub
0.0631	0.1194	0.5291	0.5968	-0.1708	0.2971

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #---

> # Sensitivity Analysis

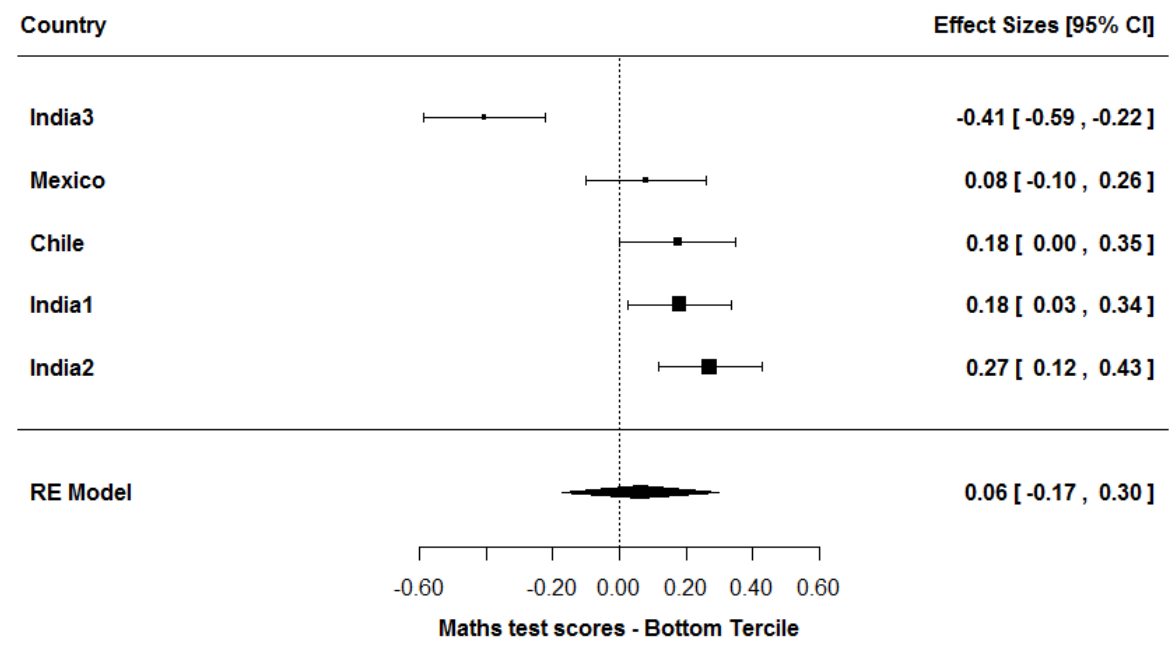
> #---

> leave1out(fit1)

	estimate	se	zval	pval	ci.lb	ci.ub	Q	Qp	tau2	I2	H2
1	0.0097	0.1389	0.0696	0.9445	-0.2626	0.2819	28.5181	0.0000	0.0694	89.9503	9.9506
2	0.1843	0.0422	4.3708	0.0000	0.1017	0.2670	2.5106	0.4734	0.0000	0.0000	1.0000
3	0.0325	0.1501	0.2168	0.8284	-0.2616	0.3267	33.8805	0.0000	0.0823	91.3815	11.6029
4	0.0582	0.1540	0.3780	0.7054	-0.2435	0.3600	35.8499	0.0000	0.0875	92.4115	13.1779

5   0.0345 0.1500 0.2301 0.8180 -0.2595 0.3286 34.5373 0.0000 0.0826 91.8708 12.3013

> #---



Math scores Middle tercile

Random-Effects Model (k = 5; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0574 (SE = 0.0462)

tau (square root of estimated tau^2 value):   0.2395

I^2 (total heterogeneity / total variability):   88.41%

H^2 (total variability / sampling variability): 8.63

Test for Heterogeneity:

Q(df = 4) = 28.3425, p-val < .0001

Model Results:

estimate   se   zval   pval   ci.lb   ci.ub

  0.0860   0.1143   0.7523   0.4518 -0.1380   0.3099

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> #---

> # Sensitivity Analysis

> #---

>   leave1out(fit1)

  estimate   se   zval   pval   ci.lb ci.ub   Q   Qp   tau2   I2   H2

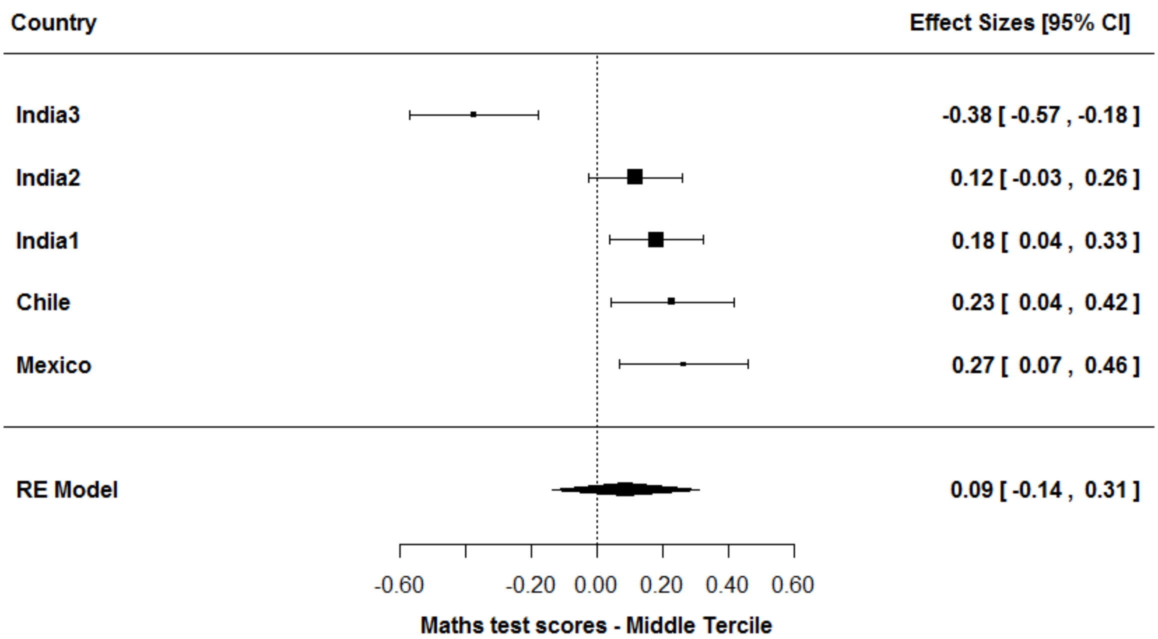
1   0.0771 0.1494 0.5159 0.6059 -0.2157 0.3699 28.2888 0.0000 0.0807 90.6423 10.6864

2 0.1845 0.0415 4.4476 0.0000 0.1032 0.2658 1.7246 0.6315 0.0000 0.0000 1.0000

3 0.0604 0.1462 0.4132 0.6794 -0.2261 0.3469 26.7504 0.0000 0.0769 90.2332 10.2388

4 0.0421 0.1363 0.3090 0.7573 -0.2251 0.3093 25.2218 0.0000 0.0669 90.4308 10.4501

5 0.0503 0.1408 0.3572 0.7209 -0.2256 0.3262 26.2600 0.0000 0.0716 90.8439 10.9217



Math scores Top tercile

Random-Effects Model (k = 5; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0250 (SE = 0.0230)

tau (square root of estimated tau^2 value): 0.1580

I^2 (total heterogeneity / total variability): 77.68%

H^2 (total variability / sampling variability): 4.48

Test for Heterogeneity:

Q(df = 4) = 15.7210, p-val = 0.0034

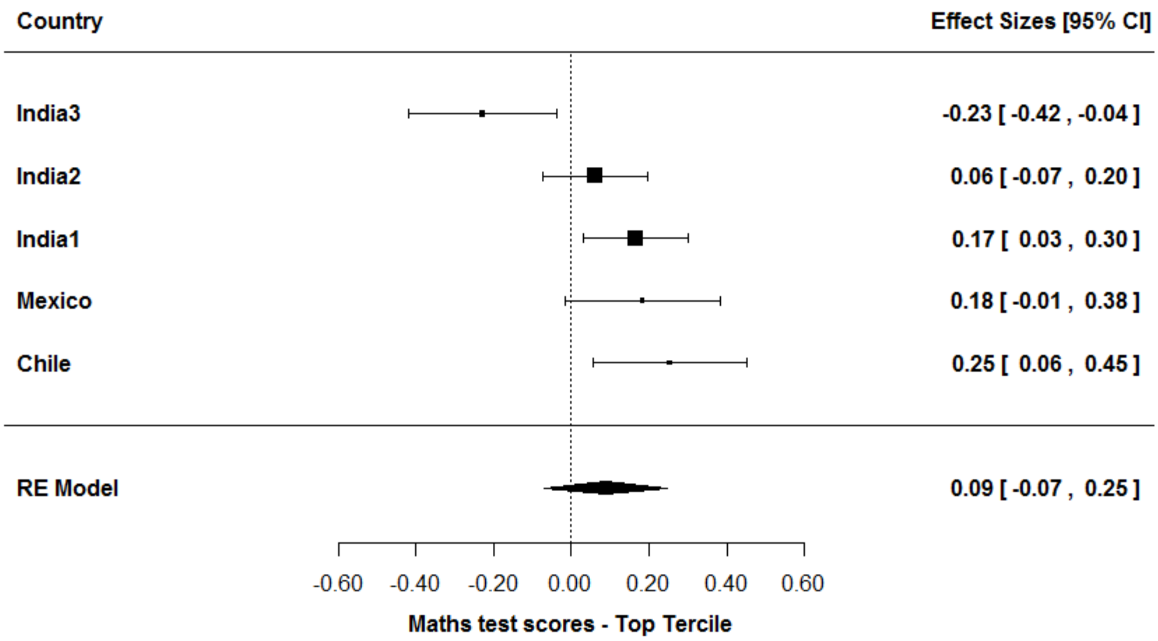
Model Results:

estimate	se	zval	pval	ci.lb	ci.ub
0.0876	0.0808	1.0848	0.2780	-0.0707	0.2459

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #---

```
> # Sensitivity Analysis
> #----------------------------------------------------
>  leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub      Q     Qp   tau2     I2     H2
1   0.0946 0.1072 0.8822 0.3777 -0.1155 0.3047 15.4598 0.0015 0.0374 82.0510 5.5713
2   0.1476 0.0412 3.5863 0.0003  0.0670 0.2283  2.8554 0.4145 0.0002  3.4449 1.0357
3   0.0660 0.1037 0.6367 0.5243 -0.1372 0.2693 14.0679 0.0028 0.0345 80.8262 5.2155
4   0.0653 0.0998 0.6537 0.5133 -0.1304 0.2609 14.7579 0.0020 0.0327 83.0807 5.9104
5   0.0497 0.0907 0.5481 0.5836 -0.1280 0.2274 12.7450 0.0052 0.0258 79.4776 4.8727
> #----------------------------------------------------
```



Remedial education

Composite

Random-Effects Model (k = 5; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.1242 (SE = 0.0882)

tau (square root of estimated tau^2 value):   0.3524

I^2 (total heterogeneity / total variability):   99.59%

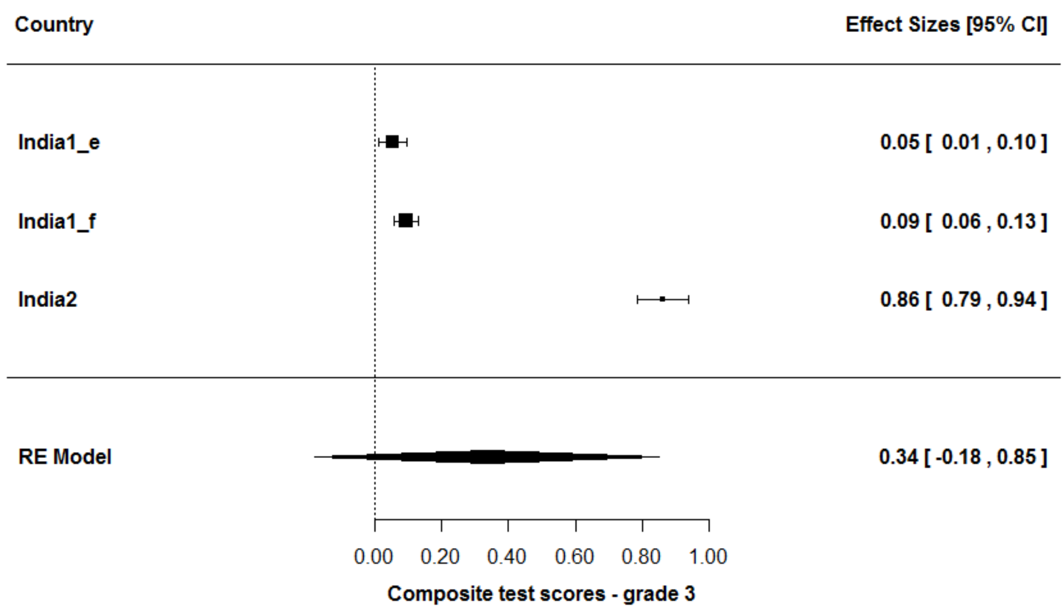
H^2 (total variability / sampling variability): 244.48

Test for Heterogeneity:

Q(df = 4) = 812.1248, p-val < .0001

```
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.2173   0.1579   1.3760   0.1688  -0.0922   0.5269
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
>  leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub       Q     Qp   tau2      I2      H2
1   0.2563 0.1976 1.2969 0.1947 -0.1311 0.6437 773.1692 0.0000 0.1557 99.6685 301.6246
2   0.2501 0.1995 1.2534 0.2101 -0.1410 0.6411 787.5196 0.0000 0.1587 99.6747 307.4228
3   0.2575 0.1973 1.3052 0.1918 -0.1292 0.6441 770.1827 0.0000 0.1551 99.6672 300.4543
4   0.2633 0.1951 1.3494 0.1772 -0.1191 0.6457 753.2157 0.0000 0.1518 99.6598 293.9594
5   0.0596 0.0110 5.4309 0.0000  0.0381 0.0812   2.9246 0.4034 0.0000  0.0000   1.0000
```

Composite - grade 3 students



Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.2058 (SE = 0.2066)

tau (square root of estimated tau^2 value):   0.4537

I^2 (total heterogeneity / total variability):   99.71%

H^2 (total variability / sampling variability): 343.11

Test for Heterogeneity:

Q(df = 2) = 358.9163, p-val < .0001

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
  0.3356   0.2624   1.2790   0.2009  -0.1787   0.8500
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Sensitivity Analysis

----

	estimate	se	zval	pval	ci.lb	ci.ub	Q	Qp	tau2	I2	H2
1	0.4766	0.3846	1.2394	0.2152	-0.2771	1.2303	316.2633	0.0000	0.2948	99.6838	316.2633
2	0.4578	0.4035	1.1347	0.2565	-0.3330	1.2486	325.6857	0.0000	0.3246	99.6930	325.6857
3	0.0756	0.0188	4.0142	0.0001	0.0387	0.1125	1.7509	0.1858	0.0003	42.8857	1.7509

> #--

Composite Grade 4

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0000 (SE = 0.0006)

tau (square root of estimated tau^2 value):   0.0067

I^2 (total heterogeneity / total variability):   10.38%

H^2 (total variability / sampling variability): 1.12

Test for Heterogeneity:

Q(df = 1) = 1.1159, p-val = 0.2908

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
  0.0862   0.0147   5.8570   <.0001   0.0574   0.1151      ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

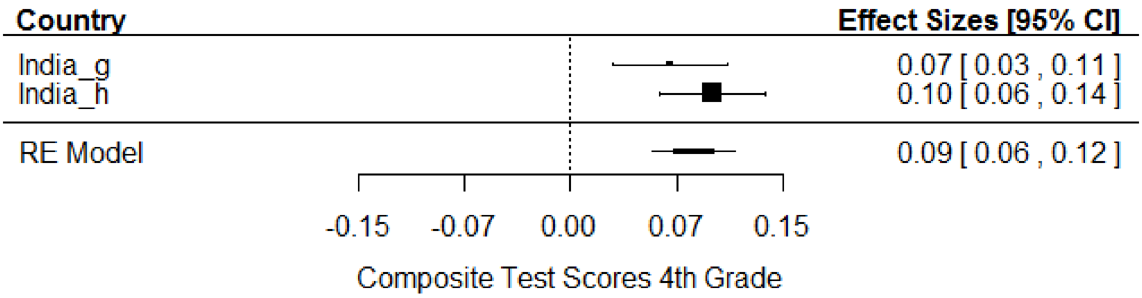
> #--

> # Sensitivity Analysis

> #--

>   leave1out(fit1)

```
  estimate     se   zval   pval  ci.lb  ci.ub     Q     Qp   tau2    I2    H2
1   0.0998 0.0188 5.2973 0.0000 0.0628 0.1367 0.0000 1.0000 0.0000   NaN   NaN
2   0.0702 0.0207 3.3952 0.0007 0.0297 0.1107 0.0000 1.0000 0.0000   NaN   NaN
> #-------------------------------------------------------
```



Language Test Scores

```
Random-Effects Model (k = 6; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.0921 (SE = 0.0585)
tau (square root of estimated tau^2 value):      0.3035
I^2 (total heterogeneity / total variability):   99.67%
H^2 (total variability / sampling variability):  305.26
Test for Heterogeneity:
Q(df = 5) = 825.2448, p-val < .0001
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.1623  0.1241   1.3074   0.1911  -0.0810   0.4056
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub        Q     Qp   tau2      I2       H2
1   0.1865 0.1492 1.2499 0.2113 -0.1060 0.4790 818.0553 0.0000 0.1110 99.7264 365.5461
2   0.1801 0.1506 1.1961 0.2317 -0.1150 0.4752 824.6302 0.0000 0.1130 99.7313 372.2188
3   0.1870 0.1491 1.2543 0.2097 -0.1052 0.4792 819.5917 0.0000 0.1108 99.7419 387.4120
4   0.1914 0.1479 1.2943 0.1955 -0.0984 0.4813 808.1398 0.0000 0.1090 99.7215 359.0647
5   0.0364 0.0102 3.5701 0.0004  0.0164 0.0564   7.3923 0.1166 0.0002 46.4724   1.8682
6   0.1907 0.1482 1.2873 0.1980 -0.0997 0.4811 757.9268 0.0000 0.1093 99.6394 277.3539
```

> #---

Language 3rd Grade

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0001 (SE = 0.0007)

tau (square root of estimated tau^2 value): 0.0101

I^2 (total heterogeneity / total variability): 20.44%

H^2 (total variability / sampling variability): 1.26

Test for Heterogeneity:

Q(df = 1) = 1.2569, p-val = 0.2622

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub	
0.0538	0.0158	3.3948	0.0007	0.0227	0.0848	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> #---

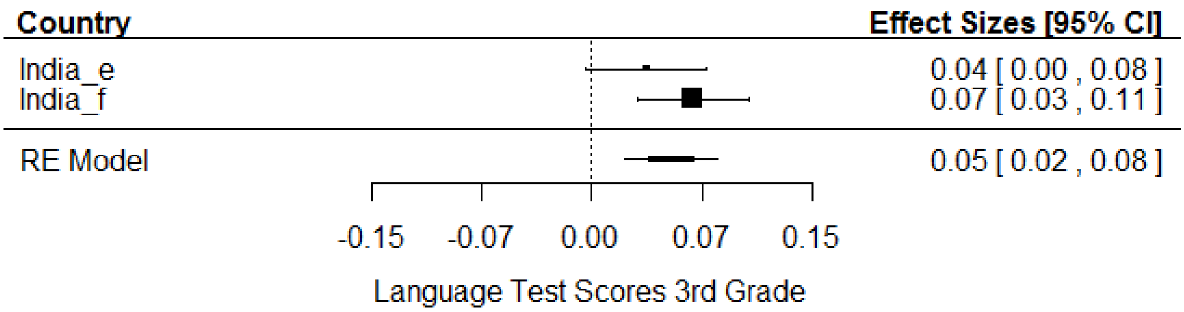
> # Sensitivity Analysis

> #---

> leave1out(fit1)

	estimate	se	zval	pval	ci.lb	ci.ub	Q	Qp	tau2	I2	H2
1	0.0686	0.0192	3.5728	0.0004	0.0310	0.1063	0.0000	1.0000	0.0000	NaN	NaN
2	0.0369	0.0208	1.7748	0.0759	-0.0039	0.0777	0.0000	1.0000	0.0000	NaN	NaN

> #---



Language 4rd Grade

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0006 (SE = 0.0014)

tau (square root of estimated tau^2 value): 0.0239

I^2 (total heterogeneity / total variability):   59.31%

H^2 (total variability / sampling variability): 2.46

Test for Heterogeneity:

Q(df = 1) = 2.4574, p-val = 0.1170

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub	
0.0691	0.0219	3.1532	0.0016	0.0261	0.1120	**

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #---

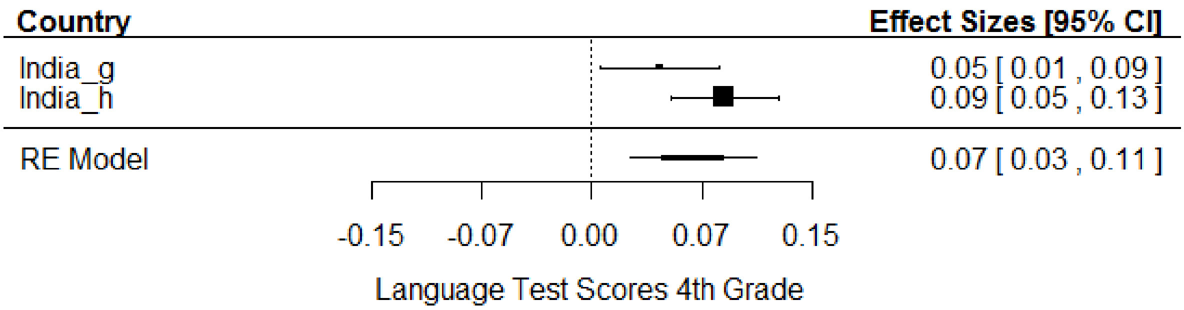
> # Sensitivity Analysis

> #---

> leave1out(fit1)

	estimate	se	zval	pval	ci.lb	ci.ub	Q	Qp	tau2	I2	H2
1	0.0902	0.0188	4.7878	0.0000	0.0532	0.1271	0.0000	1.0000	0.0000	NaN	NaN
2	0.0463	0.0207	2.2400	0.0251	0.0058	0.0868	0.0000	1.0000	0.0000	NaN	NaN

> #---



Maths Test Scores

Random-Effects Model (k = 6; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0926 (SE = 0.0588)

tau (square root of estimated tau^2 value):   0.3044

I^2 (total heterogeneity / total variability):   99.66%

H^2 (total variability / sampling variability): 292.52

Test for Heterogeneity:

Q(df = 5) = 1017.3786, p-val < .0001

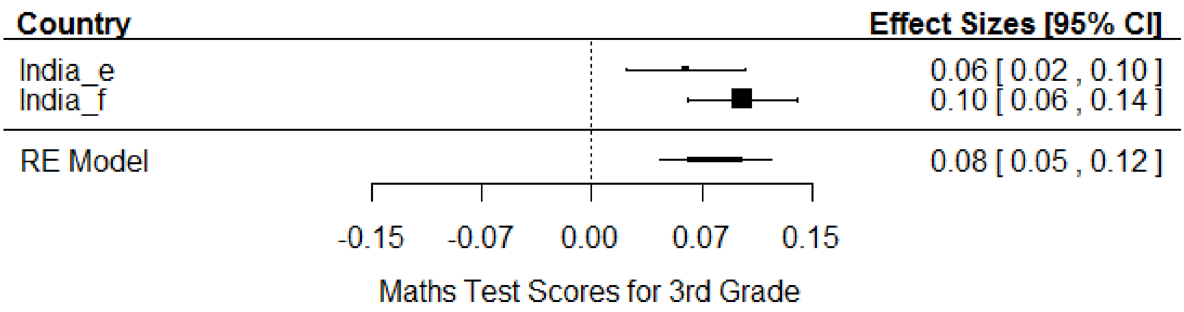
Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
  0.1920   0.1245   1.5421   0.1230 -0.0520   0.4360
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #---------------------------------------------------
> # Sensitivity Analysis
> #---------------------------------------------------
>   leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub        Q     Qp   tau2      I2       H2
1   0.2160 0.1497 1.4430 0.1490 -0.0774 0.5093  991.2972 0.0000 0.1117 99.7050
338.9600
2   0.2118 0.1506 1.4068 0.1595 -0.0833 0.5070 1003.6992 0.0000 0.1130 99.7086
343.1207
3   0.2175 0.1493 1.4571 0.1451 -0.0751 0.5100 1000.1405 0.0000 0.1111 99.7313
372.2247
4   0.2215 0.1482 1.4946 0.1350 -0.0690 0.5119  968.7028 0.0000 0.1095 99.6990
332.2407
5   0.2176 0.1493 1.4580 0.1448 -0.0749 0.5102  985.2582 0.0000 0.1110 99.7033
337.0891
6   0.0679 0.0084 8.1109 0.0000  0.0515 0.0843    4.4803 0.3449 0.0001 15.9625    1.1899
```

Maths grade 3

```
Random-Effects Model (k = 2; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.0003 (SE = 0.0010)
tau (square root of estimated tau^2 value):      0.0184
I^2 (total heterogeneity / total variability):   45.67%
H^2 (total variability / sampling variability):  1.84
Test for Heterogeneity:
Q(df = 1) = 1.8408, p-val = 0.1749
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.0840   0.0192   4.3788   <.0001   0.0464   0.1217      ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
>  leave1out(fit1)
  estimate    se   zval   pval  ci.lb  ci.ub     Q     Qp   tau2    I2    H2
1   0.1024 0.0192 5.3294 0.0000 0.0648 0.1401 0.0000 1.0000 0.0000   NaN   NaN
2   0.0640 0.0208 3.0778 0.0021 0.0232 0.1048 0.0000 1.0000 0.0000   NaN   NaN
> #-------------------------
```



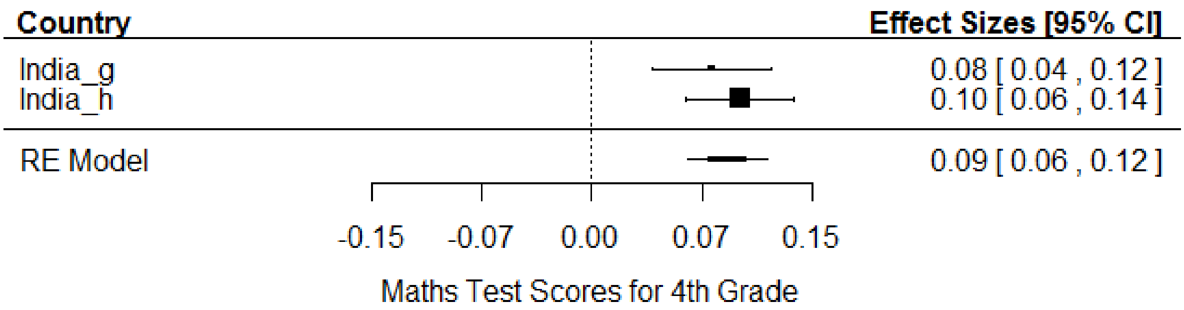
Maths grade 4

```
Random-Effects Model (k = 2; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0006)
tau (square root of estimated tau^2 value):      0
I^2 (total heterogeneity / total variability):   0.00%
H^2 (total variability / sampling variability):  1.00
Test for Heterogeneity:
Q(df = 1) = 0.4642, p-val = 0.4957
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.0919  0.0139   6.5998   <.0001   0.0646   0.1192     ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
>  leave1out(fit1)
  estimate    se   zval   pval  ci.lb  ci.ub     Q     Qp   tau2    I2    H2
```

1 0.1005 0.0188 5.3386 0.0000 0.0636 0.1374 0.0000 1.0000 0.0000 NaN NaN

2 0.0815 0.0207 3.9397 0.0001 0.0409 0.1220 0.0000 1.0000 0.0000 NaN NaN

> #--



Extra time

Language test scores

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0018)

tau (square root of estimated tau^2 value): 0

I^2 (total heterogeneity / total variability): 0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 1) = 0.0039, p-val = 0.9503

Model Results:

estimate se zval pval ci.lb ci.ub

0.1934 0.0238 8.1263 <.0001 0.1467 0.2400 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table A6.6 a: Effect sizes for studies not included in the meta-analysis for language test scores

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
Orkin, 2013	Female subgroup	897	Ethiopia	0.13694	0.00112	0.07142	0.20246
Orkin, 2013	Male subgroup	897	Ethiopia	0.17518	0.00112	0.10961	0.24074
Valenzuela, 2005	Private schools: non fee paying	751.5	Chile	0.33725	0.00226	0.24404	0.43045

Valenzuela, 2005	Private schools: fee paying	751.5	Chile	0.35920	0.00227	0.26590	0.45249

Maths test scores

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0074 (SE = 0.0130)

tau (square root of estimated tau^2 value): 0.0858

I^2 (total heterogeneity / total variability): 80.33%

H^2 (total variability / sampling variability): 5.08

Test for Heterogeneity:

Q(df = 1) = 5.0836, p-val = 0.0242

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
  0.0917   0.0677   1.3548   0.1755  -0.0410   0.2243
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table A6.6 b: Effect sizes for studies not included in the meta-analysis for maths test scores

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
Orkin, 2013	Female subgroup	897	Ethiopia	0.20033	0.00224	0.10755	0.29312
Orkin, 2013	Male subgroup	897	Ethiopia	0.21457	0.00224	0.12175	0.30738
Valenzuela, 2005	Private schools: non fee paying	751.5	Chile	0.15978	0.00267	0.05850	0.26105
Valenzuela, 2005	Private schools: fee paying	751.5	Chile	0.24039	0.00268	0.13891	0.34186

Providing materials

Language

Random-Effects Model (k = 5; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0002 (SE = 0.0003)

tau (square root of estimated tau^2 value): 0.0147

I^2 (total heterogeneity / total variability):   49.10%

H^2 (total variability / sampling variability): 1.96

Test for Heterogeneity:

Q(df = 4) = 7.7449, p-val = 0.1014

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
  0.0026   0.0094   0.2760   0.7826  -0.0159   0.0211
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
  estimate     se    zval   pval   ci.lb  ci.ub     Q     Qp   tau2     I2    H2
1   0.0006 0.0116  0.0490 0.9609 -0.0221 0.0233 7.4828 0.0580 0.0003 61.2244 2.5789
2  -0.0040 0.0082 -0.4870 0.6262 -0.0201 0.0121 3.7323 0.2919 0.0000 16.1396 1.1925
3   0.0023 0.0129  0.1802 0.8570 -0.0229 0.0275 7.7431 0.0516 0.0004 60.2900 2.5183
4   0.0098 0.0072  1.3623 0.1731 -0.0043 0.0239 2.6466 0.4494 0.0000  0.0000 1.0000
5   0.0036 0.0119  0.3037 0.7613 -0.0197 0.0269 7.6049 0.0549 0.0003 61.9562 2.6285
> #-----------------------------------------------------
```

Table A6.5 a: language scores effect sizes, subgroups and other treatment arms not included in the meta-analysis

Study	Outcome	COUNTRY	N	ESa	vara	lower	upper
Glewwe (2009)	Other follow up - year 1	Kenya	6110	0.004378	0.000327	-0.031083	0.039839
Das	Other follow up - year 1	India	13926	0.035234	0.000144	0.011743	0.058724
Glewwe (2004)	Other language - English	Kenya	15550	-0.002429	0.000129	-0.024657	0.019799

Maths

Random-Effects Model (k = 5; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0016 (SE = 0.0013)

tau (square root of estimated tau^2 value):   0.0405

I^2 (total heterogeneity / total variability): 86.78%

H^2 (total variability / sampling variability): 7.57

Test for Heterogeneity:

Q(df = 4) = 23.2424, p-val = 0.0001

Model Results:

```
estimate      se    zval     pval    ci.lb    ci.ub
 -0.0219   0.0195  -1.1232   0.2614  -0.0602   0.0163
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #------------------------------------------------------
> # Sensitivity Analysis
> #------------------------------------------------------
>   leave1out(fit1)
  estimate    se   zval   pval   ci.lb  ci.ub      Q     Qp   tau2     I2     H2
1  -0.0032 0.0080 -0.4031 0.6868 -0.0189 0.0124  3.4805 0.3233 0.0000 13.3463  1.1540
2  -0.0319 0.0223 -1.4317 0.1522 -0.0756 0.0118 17.3693 0.0006 0.0017 86.4124  7.3597
3  -0.0266 0.0253 -1.0520 0.2928 -0.0762 0.0230 22.7222 0.0000 0.0023 88.9840  9.0777
4  -0.0221 0.0256 -0.8630 0.3881 -0.0723 0.0281 22.8075 0.0000 0.0024 90.3997 10.4163
5  -0.0272 0.0248 -1.0988 0.2719 -0.0758 0.0214 22.6745 0.0000 0.0022 90.0025 10.0025
> #------------------------------------------------------
```

Table A6.5 b: math scores effect sizes, subgroups and other treatment arms not included in the meta-analysis

Study	Outcome	COUNTRY	N	ESa	vara	lower	upper
Glewwe (2009)	Other follow up - year 1	Kenya	4505	0.005464	0.000444	-0.035834	0.046761
Das	Other follow up - year 1	India	11830	0.036917	0.000169	0.011430	0.062404

Composite

Random-Effects Model (k = 5; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0002)

tau (square root of estimated tau^2 value): 0

I^2 (total heterogeneity / total variability): 0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

```
Q(df = 4) = 2.6365, p-val = 0.6204

Model Results:

estimate      se     zval     pval    ci.lb    ci.ub

  0.0082   0.0068   1.2037   0.2287  -0.0051   0.0215

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #---------------------------------------------------

> # Sensitivity Analysis

> #---------------------------------------------------

>  leave1out(fit1)

  estimate     se   zval   pval   ci.lb  ci.ub      Q     Qp   tau2     I2     H2

1   0.0107 0.0071 1.4991 0.1338 -0.0033 0.0246 1.2171 0.7489 0.0000 0.0000 1.0000

2   0.0036 0.0077 0.4636 0.6429 -0.0116 0.0188 1.0962 0.7780 0.0000 0.0000 1.0000

3   0.0097 0.0078 1.2454 0.2130 -0.0055 0.0249 2.4762 0.4796 0.0000 0.3065 1.0031

4   0.0087 0.0079 1.0988 0.2719 -0.0068 0.0241 2.6130 0.4552 0.0000 3.3375 1.0345

5   0.0078 0.0078 0.9998 0.3174 -0.0075 0.0230 2.6255 0.4530 0.0000 0.8458 1.0085
```

Table A6.5 c: composite scores effect sizes, subgroups and other treatment arms not included in the meta-analysis

Study	Outcome	COUNTRY	N	ESa	vara	lower	upper
Glewwe (2009)	Other follow up - year 1	Kenya	5661	0.009529	0.000353	-0.027311	0.046370
Glewwe (2009)	Baseline subgroup - quintile 1 (year 1)	Kenya	671	-0.059113	0.002982	-0.166143	0.047917
Glewwe (2009)	Baseline subgroup - quintile 2 (year 1)	Kenya	671	-0.023498	0.002981	-0.130509	0.083512
Glewwe (2009)	Baseline subgroup - quintile 3 (year 1)	Kenya	671	0.033845	0.002981	-0.073169	0.140859
Glewwe (2009)	Baseline subgroup - quintile 4 (year 1)	Kenya	671	0.138781	0.002988	0.031646	0.245916
Glewwe (2009)	Baseline subgroup - quintile 5 (year 1)	Kenya	671	0.175329	0.002992	0.068117	0.282541
Glewwe (2009)	Baseline subgroup - quintile 1 (year 2)	Kenya	567	-0.082955	0.003530	-0.199412	0.033502
Glewwe (2009)	Baseline subgroup - quintile 2 (year 2)	Kenya	567	-0.085779	0.003531	-0.202240	0.030681

Glewwe (2009)	Baseline subgroup - quintile 3 (year 2)	Kenya	567	-0.073391	0.003530	-0.189837	0.043055
Glewwe (2009)	Baseline subgroup - quintile 4 (year 2)	Kenya	567	0.018478	0.003527	-0.097931	0.134888
Glewwe (2009)	Baseline subgroup - quintile 5 (year 2)	Kenya	567	0.109253	0.003533	-0.007241	0.225747
Das	Other follow up - year 1	India	9891	0.038113	0.000202	0.010239	0.065986

Other outcomes

Table A6.5 d: other outcomes: effect sizes, subgroups and other treatment arms not included in the meta-analysis

Study	**Outcome**	**COUNTRY**	**N**	**ESa**	**vara**	**lower**	**upper**
Attendance							
Sabarwal	Grade - 4	Sierra Leone	299	0.017135	0.006689	-0.143168	0.177439
Sabarwal	Grade - 5	Sierra Leone	236.8	0.087727	0.008454	-0.092487	0.267941
Sabarwal	Grade 4 girls	Sierra Leone	300.5	0.038299	0.006657	-0.121616	0.198214
Sabarwal	Grade 4 boys	Sierra Leone	300.5	-0.081122	0.006661	-0.241088	0.078844
Sabarwal	Grade 5 girls	Sierra Leone	239	0.237872	0.008427	0.057943	0.417802
Sabarwal	Grade 5 boys	Sierra Leone	236.8	-0.020342	0.008446	-0.200474	0.159790
Completion							
Glewwe (2009)	Year 1	Kenya	9894	0.027645	0.000202	-0.000223	0.055513
Glewwe (2009)	Year 2	Kenya	9521	0.010350	0.000210	-0.018057	0.038758
Drop-out							
Glewwe (2009)	Year 1	Kenya	9894	-0.018592	0.000202	-0.046460	0.009275
Glewwe (2009)	Year 2	Kenya	9521	-0.026496	0.000210	-0.054905	0.001912
Enrolment							
Sabarwal	Total enrolment	Sierra Leone	325	-0.058165	0.006156	-0.211953	0.095623
Teacher Attendance							
Sabarwal	Teacher found in class	Sierra Leone	656	0.08701	0.00305	-0.02126	0.19529
Teacher performance							
Sabarwal	Teacher found teaching	Sierra Leone	656	0.189640	0.003062	0.081174	0.298105

New schools and infrastructure

Hygiene infrastructure

Enrolment

Random-Effects Model (k = 4; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0090 (SE = 0.0075)

tau (square root of estimated tau^2 value): 0.0951

I^2 (total heterogeneity / total variability): 99.14%

H^2 (total variability / sampling variability): 116.52

Test for Heterogeneity:

Q(df = 3) = 452.5603, p-val < .0001

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
  0.1065   0.0478   2.2287   0.0258   0.0128   0.2002       *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub       Q     Qp   tau2      I2      H2
1   0.1307 0.0582 2.2467 0.0247  0.0167 0.2447 344.6250 0.0000 0.0101 99.2966 142.1586
2   0.1330 0.0562 2.3682 0.0179  0.0229 0.2431 254.9282 0.0000 0.0094 99.1272 114.5753
3   0.0657 0.0355 1.8505 0.0642 -0.0039 0.1353  72.4318 0.0000 0.0037 97.3508  37.7479
4   0.0964 0.0660 1.4605 0.1442 -0.0330 0.2257 452.5426 0.0000 0.0130 99.4508 182.0741
```

Attendance

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0041 (SE = 0.0065)

tau (square root of estimated tau^2 value): 0.0639

I^2 (total heterogeneity / total variability): 88.77%

H^2 (total variability / sampling variability): 8.90

Test for Heterogeneity:

Q(df = 1) = 8.9036, p-val = 0.0028

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub	
0.1410	0.0480	2.9403	0.0033	0.0470	0.2350	**

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> leave1out(fit1)

	estimate	se	zval	pval	ci.lb	ci.ub	Q	Qp	tau2	I2	H2
1	0.1888	0.0224	8.4467	0.0000	0.1450	0.2326	0.0000	1.0000	0.0000	NaN	NaN
2	0.0929	0.0231	4.0202	0.0001	0.0476	0.1382	0.0000	1.0000	0.0000	NaN	NaN

> #---

Construction of new schools

Enrolment

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.2301 (SE = 0.3266)

tau (square root of estimated tau^2 value): 0.4797

I^2 (total heterogeneity / total variability): 99.65%

H^2 (total variability / sampling variability): 284.20

Test for Heterogeneity:

Q(df = 1) = 284.1959, p-val < .0001

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub
0.3753	0.3398	1.1043	0.2694	-0.2907	1.0413

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #---

Student Attendance

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0069 (SE = 0.0101)

tau (square root of estimated tau^2 value): 0.0832

I^2 (total heterogeneity / total variability): 97.16%

H^2 (total variability / sampling variability): 35.18

Test for Heterogeneity:

Q(df = 1) = 35.1794, p-val < .0001

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
  0.0755   0.0597   1.2639   0.2063  -0.0416   0.1925
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #----------------------------------------------------
```

Maths Test Scores

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0597 (SE = 0.0856)

tau (square root of estimated tau^2 value):   0.2443

I^2 (total heterogeneity / total variability):   98.65%

H^2 (total variability / sampling variability): 74.08

Test for Heterogeneity:

Q(df = 1) = 74.0751, p-val < .0001

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
  0.1864   0.1739   1.0719   0.2838  -0.1545   0.5273
-----------------------------------------------------------------
```

Language Test Scores

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0011)

tau (square root of estimated tau^2 value):   0

I^2 (total heterogeneity / total variability):   0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 1) = 0.0310, p-val = 0.8601

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
  0.0195   0.0148   1.3134   0.1890  -0.0096   0.0486
-----------------------------------------
```

Improvement or construction of new school infrastructure

All outcomes sub-group results, The Akshara Library program India

Outcome / sub-group	N	Country	Effect size	Variance	Lower	Upper
Language - boys	10429	India	-0.01329	0.000192	-0.04043	0.013853
Language - girls	10429	India	-0.02389	0.000192	-0.05104	0.003251
Language – grade 3	6953	India	0.004331	0.000288	-0.02891	0.037573
Language – grade 4	6953	India	-0.02634	0.000288	-0.05959	0.006899
Language – grade 5	6953	India	-0.02257	0.000288	-0.05582	0.010669
Language – quartile 1	5215	India	-0.00831	0.000384	-0.04669	0.030075
Language – quartile 2	5215	India	-0.03286	0.000384	-0.07124	0.005527
Language – quartile 3	5215	India	-0.02176	0.000384	-0.06015	0.016624
Language – quartile 4	5215	India	0.008919	0.000384	-0.02946	0.047302
Attendance - boys	10429	India	-0.00979	0.000192	-0.03693	0.01735
Attendance - girls	10429	India	0.009792	0.000192	-0.01735	0.036935
Attendance – grade 3	6953	India	0	0.000288	-0.03324	0.033242
Attendance – grade 4	6953	India	0.019188	0.000288	-0.01405	0.052431
Attendance – grade 5	6953	India	-0.00959	0.000288	-0.04284	0.023648
Attendance – quartile 1	5215	India	0.009232	0.000384	-0.02915	0.047615
Attendance – quartile 2	5215	India	0.013848	0.000384	-0.02454	0.052231
Attendance – quartile 3	5215	India	-0.00692	0.000384	-0.04531	0.03146
Attendance – quartile 4	5215	India	0.016617	0.000384	-0.02177	0.055001

Grouping students by ability

Language test scores

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0107 (SE = 0.0162)

tau (square root of estimated tau^2 value): 0.1036

I^2 (total heterogeneity / total variability): 93.55%

H^2 (total variability / sampling variability): 15.51

Test for Heterogeneity:

Q(df = 1) = 15.5095, p-val < .0001

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub
0.1200	0.0757	1.5847	0.1130	-0.0284	0.2684

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table A6.7 a: Effect sizes for studies not included in the meta-analysis for language test scores

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
Duflo et al., 2011	30 months follow-up	5007	Kenya	0.07096	0.00040	0.03177	0.11014
Duflo et al., 2015	Female subgroup	3096	India	0.26723	0.00048	0.22409	0.31037
Duflo et al., 2015	Male subgroup	3096	India	0.22192	0.00054	0.17625	0.26760
Duflo et al., 2015	CCE and LEP. Reading test.	6213	India	0.00866	0.00032	-0.02651	0.04382
Duflo et al., 2015	CCE and LEP. Written test	6213	India	0.03081	0.00032	-0.00436	0.06597

Maths test scores

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0015 (SE = 0.0027)

tau (square root of estimated tau^2 value):   0.0388

I^2 (total heterogeneity / total variability):   80.27%

H^2 (total variability / sampling variability): 5.07

Test for Heterogeneity:

Q(df = 1) = 5.0676, p-val = 0.0244

Model Results:

estimate   se   zval   pval   ci.lb   ci.ub

0.0209   0.0307   0.6831   0.4946 -0.0392   0.0810

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table A6.7 b: Effect sizes for studies not included in the meta-analysis for maths test scores

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
Duflo et al., 2011	30 months follow-up	5007	Kenya	0.06335	0.00040	0.02414	0.10256
Duflo et al., 2015	Female subgroup	3096	India	-0.01238	0.00032	-0.04760	0.02285
Duflo et al., 2015	Male subgroup	3096	India	0.03245	0.00032	-0.00278	0.06767
Duflo et al., 2015	CCE and LEP. ASER oral maths test	6213	India	0.02115	0.00032	-0.01402	0.05631
Duflo et al., 2015	CCE and LEP. Written maths test	6213	India	0.03222	0.00032	-0.00295	0.06738

7.1 Teacher incentives

Teacher attendance

Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0089 (SE = 0.0117)

tau (square root of estimated tau^2 value): 0.0943

I^2 (total heterogeneity / total variability): 75.73%

H^2 (total variability / sampling variability): 4.12

Test for Heterogeneity:

Q(df = 2) = 8.2685, p-val = 0.0160

Model Results:

estimate se zval pval ci.lb ci.ub

0.0718 0.0626 1.1480 0.2510 -0.0508 0.1944

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #---

> # Sensitivity Analysis

> #---

```
> leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub      Q     Qp   tau2      I2     H2
1   0.0093 0.0370 0.2526 0.8006 -0.0631 0.0818 0.4831 0.4870 0.0000  0.0000 1.0000
2   0.1169 0.0780 1.4973 0.1343 -0.0361 0.2698 3.9455 0.0470 0.0091 74.6547 3.9455
3   0.0892 0.1039 0.8581 0.3908 -0.1145 0.2929 7.9059 0.0049 0.0189 87.3513 7.9059
> #---------------------------------------------------
```

Table A7.1 a: Effect sizes for studies not included in the meta-analysis for teacher attendance

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
Duflo et al., 2012	Teacher absence; 30 month follow-up	3071	India	0.12632	0.00065	0.07625	0.17638
Duflo et al., 2012	Teacher absence; 30 month follow-up	882	India	0.16836	0.00228	0.07486	0.26186
Duflo et al., 2012	Teacher absence; average between 12 and 30 months	1529	India	0.14705	0.00131	0.07607	0.21803
Glewwe et al., 2010	Teacher attendance; 12 month follow-up	407	Kenya	-0.06609	0.00492	-0.20352	0.07134
Glewwe et al., 2010	Teacher attendance; 24 month follow-up	349	Kenya	0.00793	0.00573	-0.14044	0.15630
Glewwe et al., 2010	Teacher present in classroom; 24 month follow-up	481	Kenya	0.06918	0.00416	-0.05724	0.19560
Muralidharan & Sundararaman 2011	Teacher absence; 21 month follow-up	639	India	0.03146	0.00313	-0.07820	0.14112

Teacher performance: classroom management

Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0012)

tau (square root of estimated tau^2 value): 0

I^2 (total heterogeneity / total variability): 0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity: Q(df = 2) = 0.2265, p-val = 0.8929

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
 -0.0109  0.0203  -0.5358   0.5921  -0.0507   0.0289
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
>  leave1out(fit1)
  estimate    se   zval   pval   ci.lb  ci.ub     Q     Qp   tau2     I2     H2
1  -0.0025 0.0270 -0.0929 0.9260 -0.0554 0.0504 0.0050 0.9437 0.0000 0.0000 1.0000
2  -0.0127 0.0218 -0.5816 0.5608 -0.0554 0.0300 0.1750 0.6757 0.0000 0.0000 1.0000
3  -0.0165 0.0270 -0.6111 0.5411 -0.0694 0.0364 0.1268 0.7218 0.0000 0.0000 1.0000
```

Table A7.1 b: Effect sizes for studies not included in the meta-analysis for teacher performance: classroom management

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
Loyalka et al., 2015	*Incentive arm 2 – student gain*	2106	China	0.00084	0.00095	-0.05956	0.06124
Loyalka et al., 2015	Incentive arm 3 – 'pay for percentile'	2106	China	0.02853	0.00095	-0.03188	0.08893

Teacher performance: use of materials

Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0000 (SE = 0.0023)

tau (square root of estimated tau^2 value):   0.0006

I^2 (total heterogeneity / total variability):   0.02%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 2) = 1.9185, p-val = 0.3832

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
 -0.0352   0.0261  -1.3478   0.1777  -0.0864   0.0160
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
>   leave1out(fit1)
  estimate     se    zval   pval   ci.lb  ci.ub     Q     Qp   tau2     I2     H2
1  -0.0509 0.0521 -0.9766 0.3288 -0.1531 0.0513 1.6781 0.1952 0.0022 40.4088 1.6781
2  -0.0176 0.0293 -0.6010 0.5478 -0.0749 0.0398 0.1387 0.7095 0.0000  0.0000 1.0000
3  -0.0506 0.0375 -1.3480 0.1777 -0.1241 0.0230 1.4075 0.2355 0.0009 28.9533 1.4075
> #----------------------------------------------------
```

Table A7.1 c: Effect sizes for studies not included in the meta-analysis for teacher performance: use of materials

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
Glewwe et al., 2010	24 month follow-up	237	Kenya	0.15265	0.00846	-0.02766	0.33296

Teacher performance: use of assessment in instruction

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0031)

tau (square root of estimated tau^2 value):   0

I^2 (total heterogeneity / total variability):   0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 1) = 0.8707, p-val = 0.3508

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
 -0.0319   0.0294  -1.0856   0.2776  -0.0895   0.0257
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #---------------------------------------------------
> # Sensitivity Analysis
> #---------------------------------------------------
>   leave1out(fit1)
  estimate     se    zval   pval   ci.lb  ci.ub     Q     Qp   tau2    I2    H2
1   0.0125 0.0559  0.2235 0.8231 -0.0971 0.1222 0.0000 1.0000 0.0000   NaN   NaN
2  -0.0489 0.0345 -1.4140 0.1574 -0.1166 0.0189 0.0000 1.0000 0.0000   NaN   NaN
> #---------------------------------------------------
```

Table A7.1 d: Effect sizes for studies not included in the meta-analysis for teacher performance: use of assessment in instruction

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
Glewwe et al., 2010	24 month follow-up	2371	Kenya	-0.0073	0.00084		

Teacher performance: preparatory sessions

Random-Effects Model (k = 5; tau^2 estimator: REML)

tau^2 (estimate

d amount of total heterogeneity): 0 (SE = 0.0007)

tau (square root of estimated tau^2 value):   0

I^2 (total heterogeneity / total variability):   0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 4) = 1.6688, p-val = 0.7964

Model Results:

estimate   se   zval   pval   ci.lb   ci.ub

0.0692 0.0147 4.7198 <.0001 0.0405 0.0980 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #--

> # Sensitivity Analysis

> #--

> leave1out(fit1)

	estimate	se	zval	pval	ci.lb	ci.ub	Q	Qp	tau2	I2	H2
1	0.0644	0.0178	3.6112	0.0003	0.0294	0.0993	1.4374	0.6968	0.0000	0.0000	1.0000
2	0.0720	0.0152	4.7379	0.0000	0.0422	0.1018	1.1804	0.7577	0.0000	0.0000	1.0000
3	0.0634	0.0164	3.8603	0.0001	0.0312	0.0956	1.0514	0.7888	0.0000	0.0000	1.0000
4	0.0697	0.0164	4.2434	0.0000	0.0375	0.1019	1.6645	0.6448	0.0000	0.0000	1.0000
5	0.0755	0.0164	4.5919	0.0000	0.0433	0.1077	0.9643	0.8099	0.0000	0.0000	1.0000

Table A7.1 e: Effect sizes for studies not included in the meta-analysis for teacher performance: preparatory sessions

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
Glewwe et al., 2010	Follow-up after 12 months	1886	Kenya	0.04145	0.00067	-0.00916	0.09206

Teacher performance: student engagement

Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0008)

tau (square root of estimated tau^2 value): 0

I^2 (total heterogeneity / total variability): 0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 2) = 0.6019, p-val = 0.7401

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub
-0.0113	0.0164	-0.6869	0.4921	-0.0435	0.0209

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

```
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
>   leave1out(fit1)
  estimate     se    zval   pval   ci.lb  ci.ub     Q     Qp   tau2     I2     H2
1  -0.0071 0.0194 -0.3663 0.7141 -0.0452 0.0310 0.4396 0.5073 0.0000 0.0000 1.0000
2  -0.0066 0.0194 -0.3420 0.7323 -0.0447 0.0314 0.4008 0.5267 0.0000 0.0000 1.0000
3  -0.0224 0.0218 -1.0272 0.3043 -0.0651 0.0203 0.0007 0.9782 0.0000 0.0000 1.0000
> #----------------------------------------------------
> # Plot
```

Table A7.1 f: Effect sizes for studies not included in the meta-analysis for teacher performance: student engagement

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
Loyalka et al., 2015	*Teacher communication: Incentive arm 2 – student gain*	2106	China	0.03904	0.00095	-0.02136	0.09945
Loyalka et al., 2015	Teacher communication: Incentive arm 3 – ‘pay for percentile’	2106	China	0.04618	0.00095	-0.01423	0.10659

Teacher performance: teacher effort

Table A7.1 g: Effect sizes for studies not included in the meta-analysis for teacher performance: teacher effort

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
Loyalka et al., 2015	*Teacher communication: Incentive arm 2 – student gain*	2106	China	0.03904	0.00095	-0.02136	0.09945
Loyalka et al., 2015	Teacher communication: Incentive arm 3 – ‘pay for percentile’	2106	China	0.04618	0.00095	-0.01423	0.10659
Loyalka et al., 2015	*Teacher effort: Incentive arm 1 – average student exam score*	118	China	0.12982	0.01698	-0.12561	0.38526
Loyalka et al., 2015	*Teacher effort: Incentive arm 2 – student gain*	*118*	China	0.00736	0.01695	-0.24781	0.26254
Loyalka et al., 2015	Teacher effort: Incentive arm 3 – ‘pay for percentile’	118	China	-0.05814	0.01696	-0.31337	0.19708

Enrolment

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0045 (SE = 0.0089)

tau (square root of estimated tau^2 value): 0.0672

I^2 (total heterogeneity / total variability): 71.74%

H^2 (total variability / sampling variability): 3.54

Test for Heterogeneity:

Q(df = 1) = 3.5388, p-val = 0.0599

Model Results:

```
estimate     se    zval    pval   ci.lb   ci.ub
  0.0588 0.0541  1.0872  0.2769 -0.0472  0.1649
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
>  leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub      Q     Qp   tau2   I2   H2
1   0.1297 0.0586 2.2123 0.0269  0.0148 0.2446 0.0000 1.0000 0.0000  NaN  NaN
2   0.0175 0.0109 1.6098 0.1074 -0.0038 0.0389 0.0000 1.0000 0.0000  NaN  NaN
> #----------------------------------------------------
```

Table A7.1 h: Effect sizes for studies not included in the meta-analysis for enrolment

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
Barrera-Osorio, 2015	Pooled T arms 1-3 Yr1	598	Pakistan	-0.09271	0.00335	-0.20612	0.02070
Barrera-Osorio, 2015	Pooled T arms 1-3 Yr2	593	Pakistan	-0.01381	0.00337	-0.12764	0.10002
Barrera-Osorio, 2015	Pooled T arms 1-3 Yr3	583	Pakistan	0.14095	0.00344	0.02601	0.25589
Barrera-Osorio, 2015	Head teacher incentive Yr1	598	Pakistan	-0.09764	0.00335	-0.21106	0.01577
Barrera-Osorio, 2015	Head teacher incentive Yr2	593	Pakistan	-0.10096	0.00338	-0.21486	0.01293
Barrera-Osorio, 2015	Head teacher incentive Yr3	583	Pakistan	0.16181	0.00344	0.04682	0.27679
Barrera-Osorio, 2015	All teachers incentive Y1	598	Pakistan	0.05640	0.00335	-0.05698	0.16977
Barrera-Osorio, 2015	All teachers incentive Y2	593	Pakistan	0.14093	0.00338	0.02696	0.25490

Barrera-Osorio, 2015	High HT incentive, normal teacher incentive Yr 1	598	Pakistan	-0.18711	0.00336	-0.30070	-0.07351
Barrera-Osorio, 2015	High HT incentive, normal teacher incentive Yr 2	593	Pakistan	-0.06805	0.00337	-0.18191	0.04581
Barrera-Osorio, 2015	High HT incentive, normal teacher incentive Yr 3	583	Pakistan	0.05728	0.00343	-0.05754	0.17210

Attendance

Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0037)

tau (square root of estimated tau^2 value):   0

I^2 (total heterogeneity / total variability):   0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 2) = 0.0806, p-val = 0.9605

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
  0.0094   0.0241   0.3912   0.6956  -0.0379   0.0568
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub     Q     Qp  tau2     I2     H2
1   0.0104 0.0257 0.4059 0.6848 -0.0399 0.0608 0.0682 0.7940 0.0000 0.0000 1.0000
2   0.0082 0.0246 0.3312 0.7405 -0.0401 0.0564 0.0086 0.9262 0.0000 0.0000 1.0000
3   0.0119 0.0615 0.1935 0.8466 -0.1087 0.1325 0.0788 0.7790 0.0000 0.0000 1.0000
> #-----------------------------------------------------
```

Table A7.1 i: Effect sizes for studies not included in the meta-analysis for attendance

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
Duflo et al., 2012	Presence (Total days of instruction)	46184	India	0.014	0.00433	0.00115	0.02685

Duflo et al., 2012	Presence (at oral pre-test)	29113	India	0.027	0.00687	0.01072	0.04328
Duflo et al., 2012	Presence (at written pre-test)	4408	India	0.025	0.04538	-0.01676	0.06676
Glewwe et al., 2010	Present for ICS exam (24 month follow-up)	12982	Kenya	0.006	0.01541	-0.01832	0.03032
Glewwe et al., 2010	Present for ICS exam (36 month follow-up)	2277	Kenya	0.037	0.08785	-0.02111	0.09511

Completion

Random-Effects Model (k = 4; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0001 (SE = 0.0004)

tau (square root of estimated tau^2 value): 0.0117

I^2 (total heterogeneity / total variability): 22.64%

H^2 (total variability / sampling variability): 1.29

Test for Heterogeneity:

Q(df = 3) = 4.0778, p-val = 0.2532

Model Results:

```
estimate     se     zval     pval    ci.lb    ci.ub
 0.0257   0.0121   2.1283   0.0333   0.0020   0.0493      *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
>  leave1out(fit1)
  estimate     se  zval   pval   ci.lb  ci.ub     Q     Qp   tau2     I2     H2
1   0.0174 0.0103 1.6938 0.0903 -0.0027 0.0376 1.2421 0.5374 0.0000  0.0000 1.0000
2   0.0375 0.0251 1.4965 0.1345 -0.0116 0.0866 4.0748 0.1304 0.0010 52.5985 2.1096
3   0.0415 0.0193 2.1487 0.0317  0.0036 0.0793 2.2888 0.3184 0.0004 30.0046 1.4287
4   0.0239 0.0123 1.9455 0.0517 -0.0002 0.0480 3.5181 0.1722 0.0001 28.9004 1.4065
> #----------------------------------------------------
```

Table A7.1 j: Effect sizes for studies not included in the meta-analysis for completion

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
Glewwe et al., 2010	12 month Follow-up	1886	Kenya	0.02541	0.00019	-0.00159	0.05241
Behrman et al. 2012	36 month Follow-up	1886	Mexico	0.01196	0.00021	-0.01662	0.04054
Barrera-Osorio, 2015	Pooled T arms 1-3 Yr1	597	Pakistan	-0.04920	0.00335	-0.16266	0.06427
Barrera-Osorio, 2015	Pooled T arms 1-3 Yr2	591	Pakistan	0.04120	0.00338	-0.07283	0.15524
Barrera-Osorio, 2015	Pooled T arms 1-3 Yr3	591	Pakistan	0.27572	0.00351	0.15958	0.39186
Barrera-Osorio, 2015	Head teacher incentive Yr1	597	Pakistan	-0.12299	0.00336	-0.23654	-0.00944
Barrera-Osorio, 2015	Head teacher incentive Yr2	591	Pakistan	-0.06181	0.00339	-0.17585	0.05224
Barrera-Osorio, 2015	Head teacher incentive Yr3	591	Pakistan	0.28407	0.00351	0.16790	0.40025
Barrera-Osorio, 2015	All teachers incentive Y1	597	Pakistan	-0.04100	0.00335	-0.15445	0.07246
Barrera-Osorio, 2015	All teachers incentive Y3	591	Pakistan	0.11697	0.00348	0.00128	0.23266
Barrera-Osorio, 2015	High HT incentive, normal teacher incentive Yr 1	597	Pakistan	0.00000	0.00335	-0.11344	0.11344
Barrera-Osorio, 2015	High HT incentive, normal teacher incentive Yr 2	591	Pakistan	0.18954	0.00340	0.07526	0.30381
Barrera-Osorio, 2015	High HT incentive, normal teacher incentive Yr 3	591	Pakistan	0.30078	0.00352	0.18454	0.41703

Drop-out

Random-Effects Model (k = 4; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0001)

tau (square root of estimated tau^2 value):   0

I^2 (total heterogeneity / total variability):   0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 3) = 2.0754, p-val = 0.5569

Model Results:

estimate   se   zval   pval   ci.lb   ci.ub

 0.0019   0.0058   0.3256   0.7447 -0.0095   0.0133

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #---

```
> # Sensitivity Analysis
> #----------------------------------------------------
>   leave1out(fit1)
  estimate     se    zval   pval   ci.lb  ci.ub     Q     Qp   tau2     I2     H2
1   0.0025 0.0059  0.4242 0.6714 -0.0091 0.0141 1.7793 0.4108 0.0000  0.0000 1.0000
2   0.0058 0.0066  0.8773 0.3803 -0.0071 0.0187 0.5229 0.7699 0.0000  0.0000 1.0000
3  -0.0030 0.0090 -0.3308 0.7408 -0.0206 0.0147 1.5068 0.4708 0.0000  4.5765 1.0480
4  -0.0009 0.0076 -0.1243 0.9011 -0.0159 0.0140 1.6674 0.4344 0.0000 15.6882 1.1861
> #----------------------------------------------------
```

Table A7.1 k: Effect sizes for studies not included in the meta-analysis for drop-out

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
Glewwe et al., 2010	Drop-out; 24 month follow-up	12007	Kenya	-0.01327	0.00017	-0.03857	0.01202
Glewwe et al., 2010	Drop-out; 36 month follow-up	10220	Kenya	0.00456	0.00021	-0.02391	0.03304
Behrman et al. 2012	*Retention (enrolment in Spring of Yr1 given enrolment in Autumn); 6 month follow-up	12152	Mexico	-0.02030*	0.00016	-0.04544	0.00485
*N.B. Sign has been changed for estimate from Behrman et al. (2012) as it is a measure of retention, rather than of drop-out, as with the other measures included here.							

Composite test scores

```
Random-Effects Model (k = 5; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.0013 (SE = 0.0011)
tau (square root of estimated tau^2 value):      0.0359
I^2 (total heterogeneity / total variability):   95.06%
H^2 (total variability / sampling variability):  20.24
Test for Heterogeneity:
Q(df = 4) = 52.7876, p-val < .0001
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.0167   0.0173   0.9652   0.3344  -0.0173   0.0507
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #------------------------------------------------------
> # Sensitivity Analysis
> #------------------------------------------------------
>   leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub       Q     Qp   tau2     I2      H2
1   0.0049 0.0133 0.3661 0.7143 -0.0212 0.0309 44.0016 0.0000 0.0006 92.4653 13.2719
2   0.0271 0.0200 1.3553 0.1753 -0.0121 0.0664 23.6289 0.0000 0.0013 92.8131 13.9142
3   0.0111 0.0215 0.5146 0.6068 -0.0311 0.0533 18.6500 0.0003 0.0016 95.2715 21.1484
4   0.0224 0.0243 0.9246 0.3552 -0.0251 0.0700 52.6336 0.0000 0.0021 95.7958 23.7856
5   0.0249 0.0223 1.1206 0.2624 -0.0187 0.0686 52.0374 0.0000 0.0018 96.9148 32.4129
> #------------------------------------------------------
```

Table A7.1 l: Effect sizes for studies not included in the meta-analysis for composite test scores

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
Duflo et al. 2012	Female subgroup	8085	India	0.13977	0.00244	0.04291	0.23663
Duflo et al. 2012	Male subgroup	8085	India	0.10510	0.00216	0.01410	0.19611
Muralidharan & Sundararaman 2011	21 month follow-up	24665	India	0.05943	0.00008	0.04178	0.07708
Glewwe et al., 2010	District exam; 24 month follow-up	29080	Kenya	0.01419	0.00004	0.00230	0.02608
Glewwe et al., 2010	Disctrict exam; 36 month follow-up	29080	Kenya	0.01144	0.00006	-0.00371	0.02658
Glewwe et al., 2010	NGO exam; 12 month follow-up	29080	Kenya	0.01071	0.00005	-0.00316	0.02458
Glewwe et al., 2010	NGO exam; 24 month follow-up	18736	Kenya	0.00347	0.00011	-0.01678	0.02372
Barrera-Osorio, 2015	Pooled T arms 1-3 Yr1	9030	Pakistan	0.00481	0.00022	-0.02436	0.03398
Barrera-Osorio, 2015	Pooled T arms 1-3 Yr2	8085	Pakistan	0.02169	0.00025	-0.00914	0.05252
Barrera-Osorio, 2015	Pooled T arms 1-3 Yr3	8211	Pakistan	-0.00368	0.00024	-0.03427	0.02691
Barrera-Osorio, 2015	Head teacher incentive Yr1	9030	Pakistan	-0.00094	0.00022	-0.03010	0.02823
Barrera-Osorio, 2015	Head teacher incentive Yr2	8085	Pakistan	0.01046	0.00025	-0.02037	0.04128
Barrera-Osorio, 2015	Head teacher incentive Yr3	8211	Pakistan	-0.00522	0.00024	-0.03581	0.02537
Barrera-Osorio, 2015	All teachers incentive Y2	8085	Pakistan	0.00809	0.00025	-0.02274	0.03892

Barrera-Osorio, 2015	All teachers incentive Y3	8211	Pakistan	-0.01545	0.00024	-0.04604	0.01514
Barrera-Osorio, 2015	High HT incentive, normal teacher incentive Yr 1	9030	Pakistan	0.01988	0.00022	-0.00929	0.04905
Barrera-Osorio, 2015	High HT incentive, normal teacher incentive Yr 2	8085	Pakistan	0.03114	0.00025	0.00031	0.06197
Barrera-Osorio, 2015	High HT incentive, normal teacher incentive Yr 3	8211	Pakistan	0.01060	0.00024	-0.01999	0.04119

Language test scores

Random-Effects Model (k = 7; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0241 (SE = 0.0157)

tau (square root of estimated tau^2 value):   0.1553

I^2 (total heterogeneity / total variability):   98.57%

H^2 (total variability / sampling variability): 70.12

Test for Heterogeneity:

Q(df = 6) = 52.0038, p-val < .0001

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
 -0.0040   0.0624  -0.0640   0.9490  -0.1263   0.1184
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
  estimate     se    zval   pval   ci.lb  ci.ub       Q     Qp   tau2      I2       H2
1  -0.0226 0.0746 -0.3031 0.7618 -0.1689 0.1237 46.7099 0.0000 0.0298 98.9536  95.5697
2  -0.0153 0.0770 -0.1993 0.8420 -0.1662 0.1355 50.9901 0.0000 0.0319 99.0189 101.9268
```

3 -0.0166 0.0772 -0.2153 0.8296 -0.1679 0.1347 33.5781 0.0000 0.0319 96.3717 27.5608

4 -0.0220 0.0718 -0.3067 0.7591 -0.1627 0.1187 50.7338 0.0000 0.0284 98.9748 97.5446

5   0.0410 0.0187 2.1978 0.0280 0.0044 0.0777 28.6163 0.0000 0.0011 78.9301   4.7461

6 -0.0154 0.0750 -0.2053 0.8373 -0.1624 0.1316 51.6595 0.0000 0.0309 99.0515 105.4276

7 -0.0066 0.0782 -0.0849 0.9323 -0.1600 0.1467 29.1182 0.0000 0.0329 96.1873 26.2278

> #---

Table A7.1 m: Effect sizes for studies not included in the meta-analysis for language test scores

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
Duflo et al., 2012	Female subgroup	821	India	0.13185	0.00244	0.03500	0.22869
Duflo et al., 2012	Male subgroup	929	India	0.10499	0.00216	0.01398	0.19599
Contreras and Rau, 2012	24 month follow-up	1786	Chile	0.11899	0.00112	0.05334	0.18464
Muralidharan & Sundararaman 2011	24 month follow-up	12410	India	0.06855	0.00016	0.04366	0.09344
Glewwe et al., 2010	English District Y1	50842	Kenya	-0.00489	0.00004	-0.01718	0.00740
Glewwe et al., 2010	English District Y2	37620	Kenya	0.01031	0.00005	-0.00398	0.02460
Glewwe et al., 2010	English District Y3	15893	Kenya	0.00241	0.00013	-0.01958	0.02440
Glewwe et al., 2010	Swahili District Y2	37620	Kenya	0.01504	0.00005	0.00075	0.02933
Glewwe et al., 2010	Swahili District Y3	15893	Kenya	0.01719	0.00013	-0.00480	0.03917
Glewwe et al., 2010	English NGO Y1	39510	Kenya	0.00861	0.00005	-0.00534	0.02255
Glewwe et al., 2010	English NGO Y2	12996	Kenya	0.00001	0.00015	-0.02430	0.02432
Glewwe et al., 2010	English Government Y1	10430	Kenya	0.02417	0.00019	-0.00298	0.05131

Glewwe et al., 2010	English Government Y2	8427	Kenya	0.0178 1	0.0002 4	-0.01239	0.04801
Glewwe et al., 2010	English Government Y3	4053	Kenya	0.0005 0	0.0004 9	-0.04304	0.04404
Glewwe et al., 2010	Swahili Government Y1	1043 0	Kenya	0.0343 1	0.0001 9	0.00717	0.06145
Glewwe et al., 2010	Swahili Government Y2	8427	Kenya	0.0439 6	0.0002 4	0.01376	0.07416
Glewwe et al., 2010	Swahili Government Y3	4053	Kenya	- 0.0302 2	0.0004 9	-0.07376	0.01333

Maths test scores

Random-Effects Model (k = 11; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0063 (SE = 0.0035)

tau (square root of estimated tau^2 value): 0.0791

I^2 (total heterogeneity / total variability): 82.27%

H^2 (total variability / sampling variability): 5.64

Test for Heterogeneity:

Q(df = 10) = 44.2512, p-val < .0001

Model Results:

```
estimate    se     zval     pval    ci.lb   ci.ub
 0.0756   0.0266   2.8390   0.0045   0.0234   0.1277     **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
>  leave1out(fit1)
  estimate    se   zval   pval  ci.lb  ci.ub      Q     Qp   tau2     I2     H2
1    0.0758 0.0298 2.5417 0.0110 0.0173 0.1342 43.6271 0.0000 0.0073 84.1395 6.3050
2    0.0834 0.0287 2.9048 0.0037 0.0271 0.1397 41.4267 0.0000 0.0066 82.6351 5.7587
3    0.0792 0.0297 2.6668 0.0077 0.0210 0.1375 44.1879 0.0000 0.0072 83.9707 6.2386
4    0.0782 0.0298 2.6196 0.0088 0.0197 0.1367 44.2433 0.0000 0.0073 83.9512 6.2310
5    0.0623 0.0250 2.4929 0.0127 0.0133 0.1113 33.9216 0.0001 0.0048 78.9376 4.7478
```

6 0.0518 0.0178 2.9031 0.0037 0.0168 0.0868 21.0661 0.0124 0.0018 58.5022 2.4098

7 0.0761 0.0293 2.5945 0.0095 0.0186 0.1336 44.0680 0.0000 0.0071 84.7792 6.5700

8 0.0831 0.0288 2.8806 0.0040 0.0266 0.1396 41.8117 0.0000 0.0067 82.7901 5.8106

9 0.0857 0.0275 3.1115 0.0019 0.0317 0.1397 37.7944 0.0000 0.0060 81.0950 5.2896

10 0.0809 0.0295 2.7438 0.0061 0.0231 0.1386 43.6151 0.0000 0.0071 83.5379 6.0746

11 0.0744 0.0292 2.5472 0.0109 0.0172 0.1317 43.6119 0.0000 0.0071 84.6594 6.5186

Table A7.1 n: Effect sizes for studies not included in the meta-analysis for maths test scores

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
Glewwe et al., 2010	District exam; 24 month follow-up	1983	Kenya	0.01215	0.00101	-0.05009	0.07440
Glewwe et al., 2010	District exam; 36 month follow-up	1983	Kenya	-0.01373	0.00013	-0.03571	0.00826
Glewwe et al., 2010	NGO exam; 12 month follow-up	1983	Kenya	0.00721	0.00005	-0.00674	0.02115
Glewwe et al., 2010	NGO exam; 24 month follow-up	1983	Kenya	0.01636	0.00101	-0.04589	0.07861
Glewwe et al., 2010	Government exam; 12 month follow-up	1983	Kenya	0.03187	0.00019	0.00473	0.05901
Glewwe et al., 2010	Government exam; 24 month follow-up	1983	Kenya	0.02641	0.00101	-0.03584	0.08866
Glewwe et al., 2010	Government exam; 36 month follow-up	1983	Kenya	0.01115	0.00049	-0.03239	0.05469
Contreras and Rau, 2012	24 month follow-up	1786	Chile	0.13521	0.00112	0.06955	0.20088
Muralidharan & Sundararaman 2011	21 month follow-up	12255	India	0.09033	0.00101	0.02805	0.15261
Behrman et al. 2012	Baseline achievement subgroup: basic	768	Mexico	0.05340	0.00730	-0.11408	0.22088

Behrman et al. 2012	Baseline achievement subgroup: pre-basic	768	Mexico	-0.03011	0.00837	-0.20942	0.14920
Behrman et al. 2012	Baseline achievement subgroup: proficient or advanced	768	Mexico	0.01662	0.02941	-0.31952	0.35276
Behrman et al. 2012	Grade 10; 24 month follow-up	1897	Mexico	0.00248	0.00105	-0.06116	0.06612
Behrman et al. 2012	Grade 11; 24 month follow-up	1897	Mexico	0.04100	0.00105	-0.02265	0.10465
Behrman et al. 2012	Grade 12; 24 month follow-up	1897	Mexico	-0.01319	0.00105	-0.07683	0.05045
Behrman et al. 2012	Grade 10; 36 month follow-up	1886	Mexico	0.11887	0.00106	0.05499	0.18275
Behrman et al. 2012	Grade 11; 36 month follow-up	1886	Mexico	-0.00705	0.00106	-0.07087	0.05678
Behrman et al. 2012	Grade 12; 36 month follow-up	1886	Mexico	0.04236	0.00106	-0.02148	0.10619
Loyalka, 2015	Levels small	768	China	0.08617	0.00261	-0.01390	0.18624
Loyalka, 2015	Gains small	768	China	0.04238	0.00260	-0.05765	0.14242
Loyalka, 2015	Gains large	771	China	-0.03897	0.00259	-0.13880	0.06087
Loyalka, 2015	Pay-for-percentile small	768	China	0.09454	0.00261	-0.00554	0.19462
Loyalka, 2015	Pay-for-percentile large	771	China	0.19808	0.00261	0.09801	0.29815
Loyalka, 2015	Levels bottom third	1228	China	0.03805	0.00163	-0.04106	0.11715
Loyalka, 2015	Levels middle third	1228	China	0.14907	0.00163	0.06986	0.22828

Loyalka, 2015	Levels top third	1228	China	0.04941	0.00163	-0.02970	0.12852
Loyalka, 2015	Gains bottom third	1228	China	0.00865	0.00163	-0.07045	0.08775
Loyalka, 2015	Gains middle third	1228	China	0.01657	0.00163	-0.06253	0.09567
Loyalka, 2015	Gains top third	1228	China	-0.02300	0.00163	-0.10210	0.05610
Loyalka, 2015	Pay-for-percentile bottom third	1228	China	0.06464	0.00163	-0.01448	0.14376
Loyalka, 2015	Pay-for-percentile middle third	1228	China	0.18589	0.00164	0.10663	0.26516
Loyalka, 2015	Pay-for-percentile top third	1228	China	0.11262	0.00163	0.03346	0.19179

7.2 Teacher hiring

Completion

Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0007 (SE = 0.0009)

tau (square root of estimated tau^2 value): 0.0272

I^2 (total heterogeneity / total variability): 89.08%

H^2 (total variability / sampling variability): 9.15

Test for Heterogeneity:

Q(df = 2) = 8.1503, p-val = 0.0170

Model Results:

estimate se zval pval ci.lb ci.ub

0.0439 0.0174 2.5183 0.0118 0.0097 0.0780 *

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #---

> # Sensitivity Analysis

> #---

```
>  leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub      Q     Qp   tau2      I2     H2
1   0.0582 0.0066 8.8258 0.0000  0.0453 0.0711 1.7039 0.1918 0.0000 41.3107 1.7039
2   0.0285 0.0275 1.0378 0.2994 -0.0253 0.0824 4.9803 0.0256 0.0012 79.9208 4.9803
3   0.0342 0.0343 0.9979 0.3183 -0.0330 0.1014 7.5843 0.0059 0.0021 86.8149 7.5843
> #---------------------------------------------------
```

Composite test scores

```
Random-Effects Model (k = 3; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.0028 (SE = 0.0033)
tau (square root of estimated tau^2 value):      0.0531
I^2 (total heterogeneity / total variability):   92.79%
H^2 (total variability / sampling variability):  13.87
Test for Heterogeneity:
Q(df = 2) = 11.5772, p-val = 0.0031
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.0578   0.0329   1.7582   0.0787  -0.0066   0.1222        .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #---------------------------------------------------
> # Sensitivity Analysis
> #---------------------------------------------------
>  leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub       Q     Qp   tau2      I2      H2
1   0.0718 0.0604 1.1901 0.2340 -0.0465 0.1901 10.6816 0.0011 0.0066 90.6381 10.6816
2   0.0293 0.0131 2.2295 0.0258  0.0035 0.0551  2.7582 0.0968 0.0002 63.7447  2.7582
3   0.0836 0.0470 1.7787 0.0753 -0.0085 0.1758  6.6173 0.0101 0.0038 84.8882  6.6173
> #---------------------------------------------------
```

Language test scores

Sensitivity analysis:

```
Random-Effects Model (k = 2; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.0001 (SE = 0.0012)
```

tau (square root of estimated tau^2 value): 0.0119

I^2 (total heterogeneity / total variability): 16.47%

H^2 (total variability / sampling variability): 1.20

Test for Heterogeneity:

Q(df = 1) = 1.1972, p-val = 0.2739

Model Results:

estimate se zval pval ci.lb ci.ub

0.0640 0.0161 3.9829 <.0001 0.0325 0.0956 ***

Table A7.1 o: Effect sizes for studies not included in the meta-analysis for language test scores

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
Muralidharan & Sundararaman, 2013	Additional contract teacher: 12 month follow-up	12230	India	0.03879	0.00016	0.01372	0.06386
Duflo, Dupas and Kremer, 2012; Duflo, Dupas and Kremer, 2007	Additional civil-service teacher	1623	Kenya	0.04412	0.00123	-0.02468	0.11294
Estrada, 2013	Competitive teacher recruitment	1324	Mexico	-0.03051	0.00151	-0.10669	0.04566
Bau & Das, 2014	Civil-service teachers versus contract teachers: Urdu	1324	Pakistan	0.69083	0.00160	0.61241	0.76925
Bau & Das, 2014	Civil-service teachers versus contract teachers: English	1324	Pakistan	0.84882	0.00164	0.76929	0.92835
Ome, 2012	Competitive teacher recruitment: Primary school	12230	Colombia	0.02022	0.00016	-0.00484	0.04528
Ome, 2012	Competitive teacher recruitment: Secondary school	5610	Colombia	0.09875	0.00035	0.06172	0.13578
Ome, 2012	Competitive teacher recruitment: High school	12230	Colombia	-0.00478	0.00016	-0.02984	0.02028

Maths test scores

Sensitivity analysis:

Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0043 (SE = 0.0071)

tau (square root of estimated tau^2 value): 0.0657

I^2 (total heterogeneity / total variability): 85.90%

H^2 (total variability / sampling variability): 7.09

Test for Heterogeneity:

Q(df = 1) = 7.0922, p-val = 0.0077

Model Results:

estimate se zval pval ci.lb ci.ub

0.1025 0.0499 2.0550 0.0399 0.0047 0.2002 *

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table A7.1 p: Effect sizes for studies not included in the meta-analysis for maths test scores

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
Muralidharan & Sundararaman, 2013	Additional contract teacher: 12 month follow-up	12399	India	0.05158	0.0002	0.026683	0.076478
Duflo, Dupas and Kremer, 2012; Duflo, Dupas and Kremer, 2007	Additional civil-service teacher	1623	Kenya	0.06383	0.0012	-0.004992	0.132649
Estrada, 2013	Competitive teacher recruitment	1617	Mexico	-0.0141	0.0012	-0.083014	0.054849
Ome, 2012	Competitive teacher recruitment: Primary school	15550	Colombia	0.03289	0.0001	0.010664	0.055124
Ome, 2012	Competitive teacher recruitment: Secondary school	5609	Colombia	0.09781	0.0004	0.060772	0.134838

Ome, 2012	Competitive teacher recruitment: High school	15550	Colombia	-0.0021	0.0001	-0.02436	0.020091
Vegas and de Laat, 2003	Civil-service teachers versus contract teachers	837	Togo	0.26751	0.0051	0.127592	0.407426
Bau & Das, 2014	Civil-service teachers versus contract teachers	947	Pakistan	0.62453	0.0022	0.532282	0.716768

7.3 Diagnostic Feedback

Language test scores

Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0000 (SE = 0.0004)

tau (square root of estimated tau^2 value):   0.0004

I^2 (total heterogeneity / total variability):   0.04%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 2) = 2.3321, p-val = 0.3116

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
  0.0065   0.0107   0.6025   0.5468  -0.0145   0.0274
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
  estimate     se    zval   pval   ci.lb  ci.ub     Q     Qp   tau2      I2     H2
1   0.0136 0.0118  1.1501 0.2501 -0.0096 0.0367 0.2817 0.5956 0.0000  0.0000 1.0000
2   0.0006 0.0224  0.0287 0.9771 -0.0432 0.0445 2.3321 0.1267 0.0006 57.1195 2.3321
3  -0.0052 0.0157 -0.3280 0.7429 -0.0360 0.0257 1.1216 0.2896 0.0001 10.8398 1.1216
> #-----------------------------------------------------
```

Table A7.2 a: Effect sizes for studies not included in the meta-analysis for language test scores

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
Duflo et al., 2015	Reading test. Female students	3100	India	0.02611	0.00065	-0.02368	0.07589
Duflo et al., 2015	Reading test. Male students	3100	India	-0.01905	0.00065	-0.06883	0.03073
Duflo et al., 2015	Written test. Female students	3100	India	0.06643	0.00065	0.01663	0.11622
Duflo et al., 2015	Written test. Male students	3100	India	0.01071	0.00065	-0.03907	0.06049
Duflo et al., 2015	CCE and LEP. Reading test.	6213	India	0.00866	0.00032	-0.02651	0.04382
Duflo et al., 2015	CCE and LEP. Written test	6213	India	0.03081	0.00032	-0.00436	0.06597

Maths test scores

Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0000 (SE = 0.0003)

tau (square root of estimated tau^2 value): 0.0005

I^2 (total heterogeneity / total variability): 0.06%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 2) = 1.8054, p-val = 0.4055

Model Results:

```
estimate      se    zval    pval    ci.lb   ci.ub
  0.0112  0.0096  1.1642  0.2443  -0.0076  0.0300
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
  estimate    se   zval   pval   ci.lb  ci.ub    Q    Qp   tau2    I2    H2
1   0.0070 0.0104 0.6714 0.5020 -0.0134 0.0273 0.6448 0.4220 0.0000  0.0000 1.0000
2   0.0176 0.0114 1.5479 0.1216 -0.0047 0.0398 0.6943 0.4047 0.0000  0.0000 1.0000
```

3 0.0119 0.0203 0.5888 0.5560 -0.0278 0.0517 1.7649 0.1840 0.0004 43.3403 1.7649

> #---

Table A7.3 a: Effect sizes for studies not included in the meta-analysis for maths test scores

Study	Description	N	COUNTRY	ES	Variance	Lower CI	Upper CI
Duflo et al., 2015	ASER oral maths test. Female students	3100	India	0.02245	0.00064	-0.02733	0.07223
Duflo et al., 2015	ASER oral maths test. Male students	3100	India	0.00164	0.00064	-0.04814	0.05142
Duflo et al., 2015	Written maths test. Female students	3100	India	0.05255	0.00064	0.00275	0.10233
Duflo et al., 2015	Written maths test. Male students	3100	India	-0.00714	0.00064	-0.05693	0.04264
Duflo et al., 2015	CCE and LEP. ASER oral maths test	6213	India	0.02115	0.00032	-0.01402	0.05631
Duflo et al., 2015	CCE and LEP. Written maths test	6213	India	0.03222	0.00032	-0.00295	0.06738

8.1 School based management interventions

Teacher Attendance

Random-Effects Model (k = 4; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0472 (SE = 0.0561)

tau (square root of estimated tau^2 value): 0.2172

I^2 (total heterogeneity / total variability): 68.65%

H^2 (total variability / sampling variability): 3.19

Test for Heterogeneity:

Q(df = 3) = 9.5090, p-val = 0.0232

Model Results:

estimate se zval pval ci.lb ci.ub

-0.0077 0.1311 -0.0586 0.9533 -0.2645 0.2492

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #---

```
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
  estimate     se    zval   pval   ci.lb  ci.ub     Q     Qp   tau2     I2     H2
1   0.0469 0.1697  0.2763 0.7823 -0.2857 0.3795 7.9315 0.0190 0.0647 74.8263 3.9724
2  -0.1336 0.0842 -1.5873 0.1125 -0.2986 0.0314 0.8067 0.6681 0.0000  0.0000 1.0000
3  -0.0008 0.1852 -0.0044 0.9965 -0.3639 0.3622 9.4971 0.0087 0.0815 79.2166 4.8115
4   0.0579 0.1616  0.3585 0.7200 -0.2588 0.3747 7.1928 0.0274 0.0566 72.2580 3.6046
> #-----------------------------------------------------
```

Table A8.1 a: teacher attendance effect sizes, subgroups and other treatment arms

Study	Outcome	COUNTRY	N	ESa	vara	lower	upper
Blimpo	Grant Only treatment arm	Gambia	3274	0.015103	0.007299	-0.15235	0.18256

Enrolment

```
Random-Effects Model (k = 3; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0020)
tau (square root of estimated tau^2 value):      0
I^2 (total heterogeneity / total variability):   0.00%
H^2 (total variability / sampling variability):  1.00
Test for Heterogeneity:
Q(df = 2) = 0.5898, p-val = 0.7446
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.0149   0.0260   0.5731   0.5666  -0.0360   0.0658
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub     Q     Qp   tau2     I2     H2
1   0.0046 0.0318 0.1432 0.8861 -0.0578 0.0669 0.2736 0.6010 0.0000 0.0000 1.0000
```

2 0.0284 0.0318 0.8918 0.3725 -0.0340 0.0907 0.0509 0.8216 0.0000 0.0000 1.0000

3 0.0117 0.0318 0.3687 0.7123 -0.0506 0.0741 0.5603 0.4541 0.0000 0.0000 1.0000

Table A8.1 b: enrolment effect sizes, subgroups and other treatment arms

Study	Outcome	N	Country	ESa	vara	lower	upper
Blimpo	Grant only treatment arm	274	Gambia	0.036247	0.002025	-0.05194	0.124439
Beasley	Gender Subgroup- Girls	988	Niger	0.025648	0.002024	-0.06254	0.113837
Beasley	Gender Subgroup-Boys	988	Niger	0.033266	0.002025	-0.05492	0.121457
Beasley	Grade Subgroup-1	988	Niger	-0.02559	0.002024	-0.11378	0.062601
Beasley	Grade Subgroup-2	988	Niger	0.150563	0.00203	0.062253	0.238872
Beasley	Grade Subgroup-3	988	Niger	-0.02553	0.002024	-0.11372	0.062661
Beasley	Grade Subgroup-4	988	Niger	-0.02893	0.002025	-0.11712	0.059262
Beasley	Grade Subgroup-5	988	Niger	0.022854	0.002024	-0.06533	0.111041
Beasley	Grade Subgroup-6	988	Niger	-0.04226	0.002025	-0.13046	0.04593

Dropout

Random-Effects Model (k = 7; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0000 (SE = 0.0010)

tau (square root of estimated tau^2 value): 0.0029

I^2 (total heterogeneity / total variability): 0.44%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 6) = 6.2609, p-val = 0.3946

Model Results:

estimate se zval pval ci.lb ci.ub

-0.0230 0.0161 -1.4324 0.1520 -0.0545 0.0085

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

```
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
>  leave1out(fit1)
  estimate    se   zval   pval   ci.lb  ci.ub     Q     Qp   tau2     I2     H2
1  -0.0165 0.0164 -1.0020 0.3163 -0.0487 0.0157 3.1724 0.6734 0.0000 0.0000 1.0000
2  -0.0171 0.0168 -1.0125 0.3113 -0.0501 0.0160 4.9804 0.4183 0.0000 0.0340 1.0003
3  -0.0323 0.0179 -1.8010 0.0717 -0.0674 0.0028 4.9156 0.4263 0.0000 0.0126 1.0001
4  -0.0181 0.0179 -1.0113 0.3119 -0.0533 0.0170 5.8969 0.3164 0.0000 0.1095 1.0011
5  -0.0222 0.0165 -1.3463 0.1782 -0.0544 0.0101 6.2119 0.2861 0.0000 0.2661 1.0027
6  -0.0286 0.0189 -1.5138 0.1301 -0.0656 0.0084 5.9505 0.3111 0.0002 7.8160 1.0848
7  -0.0295 0.0180 -1.6361 0.1018 -0.0649 0.0058 5.6230 0.3446 0.0000 1.0216 1.0103
> #----------------------------------------------------
```

Table A8.1 c: Dropout effect sizes, subgroups and other treatment arms

Study	Outcome	N	Country	ESa	vara	lower	upper
Murnane	PEC1 treatment arm	1566	Mexico1	-0.01654	0.001277	-0.086583	0.053509
Carnoy	Treatment 2	394.2	Brazil	-0.31764	0.005138	-0.458127	-0.177154
Carnoy	Treatment 3	394.2	Brazil	-0.21782	0.005104	-0.357844	-0.077800
Beasley	Subgroup-Grade 1	531	Niger	-0.15572	0.003778	-0.276193	-0.035252
Beasley	Subgroup-Grade 2	434	Niger	-0.05796	0.004610	-0.191042	0.075121
Beasley	Subgroup-Grade 3	525	Niger	-0.11863	0.003816	-0.239713	0.002448
Beasley	Subgroup-Grade 4	454	Niger	-0.07303	0.004408	-0.203160	0.057106
Beasley	Subgroup-Grade 5	394.2	Niger	0.03186	0.005074	-0.107756	0.171479
Beasley	Subgroup-Grade 6	466	Niger	0.01305	0.004292	-0.115358	0.141453
Beasley	Subgroup- Girls	753	Niger	-0.05618	0.002657	-0.157211	0.044853
Beasley	Subgroup- Boys	754	Niger	-0.07091	0.002654	-0.171890	0.030064

Completion

Random-Effects Model (k = 8; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0024 (SE = 0.0022)

tau (square root of estimated tau^2 value): 0.0486

I^2 (total heterogeneity / total variability): 77.18%

H^2 (total variability / sampling variability): 4.38

Test for Heterogeneity:

Q(df = 7) = 22.6108, p-val = 0.0020

Model Results:

```
estimate      se    zval    pval   ci.lb   ci.ub
  0.0453  0.0236  1.9166  0.0553 -0.0010  0.0917    .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>  leave1out(fit1)
  estimate    se   zval   pval  ci.lb  ci.ub     Q    Qp  tau2    I2    H2
1  0.0552 0.0217 2.5498 0.0108  0.0128 0.0977 18.6755 0.0047 0.0016 72.2238 3.6002
2  0.0590 0.0290 2.0333 0.0420  0.0021 0.1158 14.9065 0.0210 0.0030 66.5027 2.9853
3  0.0531 0.0345 1.5398 0.1236 -0.0145 0.1206 21.1809 0.0017 0.0050 76.6943 4.2908
4  0.0465 0.0268 1.7335 0.0830 -0.0061 0.0990 22.5161 0.0010 0.0031 83.4614 6.0465
5  0.0338 0.0193 1.7465 0.0807 -0.0041 0.0716 15.2859 0.0181 0.0013 67.4378 3.0710
6  0.0445 0.0256 1.7342 0.0829 -0.0058 0.0948 22.3101 0.0011 0.0028 82.0707 5.5775
7  0.0322 0.0219 1.4743 0.1404 -0.0106 0.0750 17.6920 0.0070 0.0015 68.2108 3.1457
8  0.0464 0.0319 1.4550 0.1457 -0.0161 0.1089 21.3644 0.0016 0.0042 86.2178 7.2557
> #-----------------------------------------------------
```

Table A8.1 d: Completion effect sizes, subgroups and other treatment arms

Study	Outcome	N	Country	ESa	vara	lower	upper
Murnane	PEC1 treatment arm	13490	Mexico1	0.004329	0.000148	-0.019536	0.028194
Carnoy	Treatment 2	171.5	Brazil	0.360570	0.011851	0.147197	0.573943
Carnoy	Treatment 3	171.5	Brazil	0.271640	0.011769	0.059006	0.484274

Composite test scores

Random-Effects Model (k = 9; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0148 (SE = 0.0095)

tau (square root of estimated tau^2 value): 0.1216

I^2 (total heterogeneity / total variability): 85.65%

H^2 (total variability / sampling variability): 6.97

Test for Heterogeneity:

Q(df = 8) = 45.2978, p-val < .0001

Model Results:

```
estimate      se    zval    pval   ci.lb   ci.ub
 -0.0097  0.0462 -0.2108  0.8330 -0.1003  0.0808
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
  estimate     se    zval   pval   ci.lb  ci.ub       Q     Qp   tau2      I2     H2
1  -0.0324 0.0469 -0.6903 0.4900 -0.1242 0.0595 33.1271 0.0000 0.0127 81.4220 5.3827
2  -0.0150 0.0534 -0.2803 0.7793 -0.1196 0.0897 44.9104 0.0000 0.0181 88.2091 8.4811
3  -0.0320 0.0474 -0.6738 0.5004 -0.1249 0.0610 34.2300 0.0000 0.0131 81.8886 5.5214
4  -0.0142 0.0542 -0.2617 0.7936 -0.1203 0.0920 43.5217 0.0000 0.0184 86.7539 7.5494
5  -0.0069 0.0535 -0.1294 0.8971 -0.1117 0.0979 38.1585 0.0000 0.0179 86.5021 7.4086
6   0.0004 0.0493  0.0086 0.9931 -0.0961 0.0970 42.4319 0.0000 0.0155 87.2941 7.8704
7   0.0287 0.0355  0.8079 0.4191 -0.0409 0.0983 29.6450 0.0001 0.0066 74.6640 3.9470
8  -0.0045 0.0506 -0.0884 0.9296 -0.1036 0.0947 43.7228 0.0000 0.0165 87.9881 8.3251
9  -0.0172 0.0515 -0.3343 0.7381 -0.1180 0.0836 45.2972 0.0000 0.0172 88.4092 8.6275
```

Table A8.1 e: Composite effect sizes, subgroups and other treatment arms

Study	Outcome	N	Country	ESa	vara	lower	upper
Khattri	2003 Treatment arm	1736	Philippines1	0.210623	0.001158	NA	NA
Blimpo	Grant- Grade 3	1512	Gambia	0.006429	0.001323	NA	NA
Blimpo	Grant- Grade 5	1451	Gambia	0.017501	0.001378	NA	NA
Carneiro	Grade 3 Male First Follow Up	338	Senegal	0.041			
Carneiro	Grade 3 Female First Follow Up	338	Senegal	0.217			

Carneiro	Grade 5 Male First Follow Up	332	Senegal	0.043			
Carneiro	Grade 5 Female First Follow Up	332	Senegal	0.009			

Language arts

Random-Effects Model (k = 20; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0153 (SE = 0.0061)

tau (square root of estimated tau^2 value):   0.1238

I^2 (total heterogeneity / total variability):   84.87%

H^2 (total variability / sampling variability): 6.61

Test for Heterogeneity:

Q(df = 19) = 114.9422, p-val < .0001

Model Results:

estimate   se   zval   pval   ci.lb   ci.ub

 -0.0130   0.0306 -0.4240   0.6715 -0.0730   0.0470

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> #--

> # Sensitivity Analysis

> #--

>   leave1out(fit1)

  estimate   se   zval   pval   ci.lb ci.ub   Q   Qp   tau2   I2   H2

1   -0.0229 0.0307 -0.7438 0.4570 -0.0831 0.0374 99.0814 0.0000 0.0144 83.4547 6.0440

2   -0.0143 0.0325 -0.4416 0.6588 -0.0780 0.0493 114.9335 0.0000 0.0165 85.2452 6.7775

3   -0.0107 0.0320 -0.3331 0.7391 -0.0734 0.0521 113.9002 0.0000 0.0161 85.8565 7.0704

4   -0.0109 0.0323 -0.3374 0.7358 -0.0742 0.0524 113.2806 0.0000 0.0163 85.6200 6.9541

5   -0.0226 0.0304 -0.7441 0.4568 -0.0823 0.0370 108.3748 0.0000 0.0143 84.2834 6.3627

6   -0.0228 0.0308 -0.7418 0.4582 -0.0831 0.0375 99.2723 0.0000 0.0144 83.4768 6.0521

7   -0.0120 0.0324 -0.3687 0.7123 -0.0756 0.0516 113.7409 0.0000 0.0165 85.3975 6.8481

8   -0.0123 0.0325 -0.3787 0.7049 -0.0760 0.0514 113.8631 0.0000 0.0165 85.2408 6.7754

9   -0.0068 0.0315 -0.2174 0.8279 -0.0686 0.0549 110.0928 0.0000 0.0155 85.2543 6.7817

10 -0.0029 0.0304 -0.0956 0.9239 -0.0624 0.0566 103.9326 0.0000 0.0141 84.0980 6.2885

11	-0.0142	0.0322	-0.4398	0.6601	-0.0772	0.0489	114.9407	0.0000	0.0163	85.9428	7.1138
12	-0.0109	0.0321	-0.3387	0.7349	-0.0738	0.0520	113.9351	0.0000	0.0162	85.8539	7.0691
13	-0.0165	0.0320	-0.5143	0.6071	-0.0793	0.0463	114.4980	0.0000	0.0161	85.7759	7.0303
14	0.0085	0.0244	0.3486	0.7274	-0.0394	0.0564	75.0742	0.0000	0.0080	75.0742	4.0119
15	-0.0141	0.0322	-0.4380	0.6614	-0.0773	0.0490	114.9386	0.0000	0.0163	85.9285	7.1065
16	-0.0116	0.0322	-0.3595	0.7192	-0.0746	0.0515	114.2627	0.0000	0.0163	85.8858	7.0851
17	-0.0159	0.0321	-0.4948	0.6207	-0.0788	0.0470	114.7087	0.0000	0.0162	85.8324	7.0584
18	-0.0149	0.0322	-0.4644	0.6424	-0.0780	0.0481	114.9049	0.0000	0.0163	85.8969	7.0906
19	-0.0227	0.0308	-0.7372	0.4610	-0.0831	0.0377	99.6931	0.0000	0.0145	83.5253	6.0699
20	-0.0095	0.0323	-0.2955	0.7676	-0.0727	0.0537	109.6851	0.0000	0.0162	85.0167	6.6741

> #---

Table A8.1 f: Language effect sizes, subgroups and other treatment arms

Study	Outcome	N	Country	ESa	vara	lower	upper
Khattri	2003 Treatment	10334	Philippines1	0.147753	0.000194	0.120449	0.175057
Blimpo	Grant Treatment	4817	Gambia	-0.03293	0.000415	-0.07287	0.007007
Yamauchi	English	2406	Philippines2	0.127622	0.000833	0.071055	0.18419
Beasley	Subgroup-Grade 1	179	Niger	0.165105	0.011211	-0.04243	0.372636
Beasley	Subgroup-Grade 2	316	Niger	0.059248	0.006332	-0.09672	0.215211
Beasley	Subgroup-Grade 3	162	Niger	-0.1262	0.01237	-0.3442	0.091792
Beasley	Subgroup-Grade 4	402	Niger	-0.06973	0.004978	-0.20802	0.068561
Beasley	Subgroup-Grade 5	109	Niger	-0.11246	0.018378	-0.37817	0.153244
Beasley	Subgroup-Grade 6	179	Niger	-0.07385	0.011181	-0.2811	0.133402
Pradhan	Subgroup-Boys	45	Indonesia	0.241353	0.044768	-0.17335	0.656059
Pradhan	Subgroup-Girls	45	Indonesia	0.315426	0.044997	-0.10034	0.731192
Carneiro	Grade 3 Male First Follow Up	680	Senegal	-0.01183	0.002941	-0.11813	0.094464
Carneiro	Grade 3 Female First Follow Up	680	Senegal	-0.17657	0.002953	-0.28308	-0.07007
Carneiro	Grade 5 Male First Follow Up	662	Senegal	-0.09695	0.003025	-0.20475	0.010841
Carneiro	Grade 5 Female First Follow Up	662	Senegal	-0.05835	0.003022	-0.1661	0.049405

Maths

```
Random-Effects Model (k = 21; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.0033 (SE = 0.0019)
tau (square root of estimated tau^2 value):      0.0572
I^2 (total heterogeneity / total variability):   56.40%
H^2 (total variability / sampling variability):  2.29
Test for Heterogeneity:
Q(df = 20) = 45.9563, p-val = 0.0008
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.0143   0.0172   0.8308   0.4061  -0.0195   0.0481
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
   estimate     se   zval    pval    ci.lb  ci.ub       Q      Qp   tau2      I2     H2
1    0.0072 0.0165 0.4355 0.6632 -0.0252 0.0396 36.1473 0.0101 0.0025 48.3514 1.9362
2    0.0078 0.0171 0.4575 0.6473 -0.0257 0.0413 39.1177 0.0043 0.0028 51.6326 2.0675
3    0.0175 0.0179 0.9790 0.3276 -0.0176 0.0527 42.7403 0.0014 0.0034 55.9713 2.2712
4    0.0185 0.0173 1.0704 0.2844 -0.0154 0.0525 42.9872 0.0013 0.0031 55.8053 2.2627
5    0.0183 0.0176 1.0436 0.2967 -0.0161 0.0528 42.9204 0.0013 0.0032 55.8979 2.2675
6    0.0157 0.0178 0.8833 0.3771 -0.0191 0.0505 45.2828 0.0006 0.0034 57.9579 2.3786
7    0.0109 0.0181 0.6028 0.5467 -0.0246 0.0464 45.2398 0.0006 0.0035 57.4868 2.3522
8    0.0163 0.0180 0.9024 0.3668 -0.0191 0.0516 44.3749 0.0008 0.0035 57.2356 2.3384
9    0.0088 0.0176 0.4971 0.6191 -0.0258 0.0433 41.9530 0.0018 0.0031 54.3561 2.1909
10   0.0185 0.0173 1.0649 0.2869 -0.0155 0.0525 43.0684 0.0013 0.0031 55.8834 2.2667
11   0.0194 0.0172 1.1277 0.2594 -0.0143 0.0530 42.1084 0.0017 0.0030 54.9513 2.2198
12   0.0112 0.0177 0.6301 0.5287 -0.0236 0.0459 45.1327 0.0007 0.0034 57.8182 2.3707
13   0.0136 0.0179 0.7599 0.4473 -0.0215 0.0487 45.9497 0.0005 0.0035 58.5521 2.4127
14   0.0145 0.0179 0.8128 0.4163 -0.0205 0.0495 45.7828 0.0005 0.0035 58.4069 2.4042
```

15 0.0155 0.0178 0.8690 0.3848 -0.0194 0.0503 45.4055 0.0006 0.0034 58.0691 2.3849

16 0.0139 0.0179 0.7766 0.4374 -0.0212 0.0489 45.9201 0.0005 0.0035 58.5271 2.4112

17 0.0195 0.0171 1.1369 0.2556 -0.0141 0.0531 41.9622 0.0018 0.0030 54.8077 2.2128

18 0.0150 0.0178 0.8431 0.3992 -0.0199 0.0500 45.6015 0.0006 0.0034 58.2455 2.3949

19 0.0145 0.0179 0.8128 0.4163 -0.0205 0.0495 45.7828 0.0005 0.0035 58.4069 2.4042

20 0.0092 0.0178 0.5179 0.6045 -0.0256 0.0441 42.9512 0.0013 0.0033 55.2442 2.2343

21 0.0141 0.0183 0.7679 0.4425 -0.0218 0.0500 45.6576 0.0006 0.0036 57.8517 2.3726

Table A8.1 g: Maths effect sizes, subgroups and other treatment arms

Study	Outcome	N	Country	ESa	vara	lower	upper
Khattri	2003 treatment group	1449	Phillipines 1	0.176679	0.001386	NA	NA
Blimpo	Grant only treatment arm	1449	Gambia	-0.058242	0.001381	NA	NA
Carnoy	Treatment Group 2	1138	Brazil	0.035237	0.001758	NA	NA
Carnoy	Treatment Group 3	1138	Brazil	-0.045274	0.001758	NA	NA
Carnoy	Treatment Group 4	1138	Brazil	0.075329	0.001759	NA	NA
Carneiro	Subgroup - Grade 3 Male at first follow-up	680	Senegal	-0.133622	0.002948	NA	NA
Carneiro	Subgroup - Grade 3 Female at first follow-up	680	Senegal	-0.175852	0.002953	NA	NA
Carneiro	Subgroup - Grade 5 Male at first follow-up	661	Senegal	-0.044355	0.003026	NA	NA
Carneiro	Subgroup - Grade 5 Female at first follow-up	661	Senegal	-0.068349	0.003027	NA	NA

8.2 Community based monitoring and accountability interventions

Enrolment

```
Random-Effects Model (k = 12; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.0174 (SE = 0.0105)
tau (square root of estimated tau^2 value):      0.1320
I^2 (total heterogeneity / total variability):   72.52%
H^2 (total variability / sampling variability):  3.64
Test for Heterogeneity:
Q(df = 11) = 37.6374, p-val < .0001
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.1652   0.0454   3.6415   0.0003   0.0763   0.2542     ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
>   leave1out(fit1)
   estimate     se   zval   pval  ci.lb  ci.ub      Q     Qp   tau2     I2     H2
1    0.1820 0.0465 3.9179 0.0001 0.0910 0.2731 32.2059 0.0004 0.0163 70.7707 3.4212
2    0.1883 0.0414 4.5439 0.0000 0.1071 0.2695 26.3553 0.0033 0.0116 63.2127 2.7183
3    0.1382 0.0383 3.6136 0.0003 0.0633 0.2132 25.2215 0.0049 0.0095 60.2029 2.5127
4    0.1751 0.0493 3.5549 0.0004 0.0786 0.2717 36.0270 0.0001 0.0192 74.0462 3.8530
5    0.1671 0.0504 3.3136 0.0009 0.0683 0.2659 37.6321 0.0000 0.0205 75.2464 4.0398
6    0.1653 0.0504 3.2776 0.0010 0.0664 0.2641 37.6015 0.0000 0.0205 75.2485 4.0402
7    0.1614 0.0501 3.2198 0.0013 0.0632 0.2597 37.0883 0.0001 0.0201 74.9556 3.9929
8    0.1577 0.0495 3.1876 0.0014 0.0607 0.2547 36.0592 0.0001 0.0194 74.2833 3.8885
9    0.1557 0.0490 3.1798 0.0015 0.0597 0.2517 35.3127 0.0001 0.0189 73.7527 3.8099
10   0.1514 0.0475 3.1863 0.0014 0.0583 0.2445 33.2459 0.0002 0.0174 72.1052 3.5849
11   0.1707 0.0487 3.5029 0.0005 0.0752 0.2663 37.3308 0.0000 0.0193 75.4832 4.0788
12   0.1738 0.0482 3.6084 0.0003 0.0794 0.2682 36.7486 0.0001 0.0187 74.8884 3.9822
> #----------------------------------------------------
```

Table A8.2 a: Enrolment effect sizes, sub-groups and other treatment arms

Study	Outcome	COUNTRY	N	ESa	vara	lower	upper
Mizala	4 year follow up	Chile	364	-0.00814	0.005495	-0.15342	0.137149
Zeitlin	Participatory scorecard treatment arm	Uganda2	364	0.047363	0.005496	-0.09794	0.192668
Banerjee	Treatment 1 - boys	India1	138	0.145953	0.014531	-0.09032	0.382223
Banerjee	Treatment 3 - boys	India1	138	-0.49528	0.014937	-0.73483	-0.25573
Banerjee	Treatment 1 - girls	India1	138	0.002211	0.014493	-0.23375	0.238167
Banerjee	Treatment 3 - girls	India1	138	0.02951	0.014494	-0.20646	0.265479

Attendance

Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0131 (SE = 0.0148)

tau (square root of estimated tau^2 value): 0.1143

I^2 (total heterogeneity / total variability): 88.69%

H^2 (total variability / sampling variability): 8.84

Test for Heterogeneity:

Q(df = 2) = 18.2666, p-val = 0.0001

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
  0.0445   0.0702   0.6346   0.5257  -0.0930   0.1821
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
-----------------------------------------------------
Sensitivity Analysis
>-----------------------------------------------------
  estimate     se    zval   pval   ci.lb  ci.ub       Q     Qp   tau2      I2      H2
1   0.1022 0.0731  1.3988 0.1619 -0.0410 0.2454  5.4147 0.0200 0.0087 81.5318  5.4147
2  -0.0220 0.0468 -0.4702 0.6382 -0.1137 0.0697  2.8019 0.0942 0.0028 64.3093  2.8019
3   0.0536 0.1200  0.4467 0.6551 -0.1816 0.2889 18.2603 0.0000 0.0272 94.5236 18.2603
```

Table A8.2 b: Attendance effect sizes, sub-groups and other treatment arms

Study	Outcome	Country	N	ESa	vara	lower	upper
Banerjee	Treatment 3	India1	1695	-0.03212	0.00118	-0.09945	0.035209
Banerjee	Treatment 1	India1	1695	-0.02983	0.00118	-0.09716	0.037503
Banerjee	Treatment 3 - boys	India1	1334	-0.03414	0.001499	-0.11004	0.041757
Banerjee	Treatment 2 - boys	India1	1334	-0.07703	0.0015	-0.15295	-0.00111
Banerjee	Treatment 1 - boys	India1	1334	-0.03632	0.001499	-0.11222	0.039575
Banerjee	Treatment 3 - girls	India1	1153	-0.02255	0.001735	-0.10418	0.059083
Banerjee	Treatment 2 - girls	India1	1153	-0.09822	0.001737	-0.1799	-0.01654
Banerjee	Treatment 1 - girls	India1	1153	-0.03312	0.001735	-0.11476	0.048515
Zeitlin	Participatory scorecard treatment arm	Uganda2	1060	0.200927	0.001978	0.113747	0.288108

Dropout

Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0129 (SE = 0.0163)

tau (square root of estimated tau^2 value):   0.1134

I^2 (total heterogeneity / total variability):   79.84%

H^2 (total variability / sampling variability): 4.96

Test for Heterogeneity:

Q(df = 2) = 8.0685, p-val = 0.0177

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
  0.0546   0.0737   0.7411   0.4586  -0.0899   0.1991
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
  estimate     se    zval   pval   ci.lb  ci.ub      Q     Qp   tau2      I2     H2
1  -0.0109 0.0344 -0.3175 0.7508 -0.0784 0.0565 0.1804 0.6710 0.0000  0.0000 1.0000
2   0.1017 0.1099  0.9254 0.3547 -0.1137 0.3171 6.5251 0.0106 0.0206 84.6745 6.5251
```

3 0.0904 0.1254 0.7211 0.4709 -0.1553 0.3362 7.0713 0.0078 0.0270 85.8583 7.0713

Completion

Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0028)

tau (square root of estimated tau^2 value): 0

I^2 (total heterogeneity / total variability): 0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 2) = 1.3181, p-val = 0.5174

Model Results:

```
estimate      se    zval    pval   ci.lb   ci.ub
  0.0650  0.0299  2.1720  0.0299  0.0063  0.1236      *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #---------------------------------------------------
> # Sensitivity Analysis
> #---------------------------------------------------
>  leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub     Q     Qp   tau2      I2     H2
1   0.0663 0.0403 1.6448 0.1000 -0.0127 0.1452 1.2654 0.2606 0.0007 20.9717 1.2654
2   0.0816 0.0344 2.3701 0.0178  0.0141 0.1490 0.3668 0.5448 0.0000  0.0000 1.0000
3   0.0347 0.0417 0.8337 0.4044 -0.0469 0.1164 0.2299 0.6316 0.0000  0.0000 1.0000
> #---------------------------------------------------
```

Composite test scores

Random-Effects Model (k = 7; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0204 (SE = 0.0128)

tau (square root of estimated tau^2 value): 0.1427

I^2 (total heterogeneity / total variability): 93.13%

H^2 (total variability / sampling variability): 14.56

Test for Heterogeneity:

Q(df = 6) = 38.8351, p-val < .0001

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
  0.0969   0.0563   1.7209   0.0853  -0.0135   0.2073        .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
>   leave1out(fit1)
  estimate    se   zval   pval   ci.lb  ci.ub      Q     Qp   tau2     I2      H2
1   0.1146 0.0654 1.7510 0.0799 -0.0137 0.2428 35.3837 0.0000 0.0237 93.8609 16.2891
2   0.1075 0.0678 1.5861 0.1127 -0.0253 0.2403 38.5942 0.0000 0.0256 94.2805 17.4842
3   0.1075 0.0677 1.5876 0.1124 -0.0252 0.2403 38.5852 0.0000 0.0256 94.2956 17.5302
4   0.1021 0.0685 1.4912 0.1359 -0.0321 0.2363 38.5885 0.0000 0.0262 94.4028 17.8659
5   0.1154 0.0649 1.7791 0.0752 -0.0117 0.2426 35.3574 0.0000 0.0234 93.9522 16.5350
6   0.0448 0.0166 2.6938 0.0071  0.0122 0.0774  6.3526 0.2734 0.0003 20.5902  1.2593
7   0.0985 0.0685 1.4372 0.1507 -0.0358 0.2328 37.5216 0.0000 0.0262 94.4133 17.8995
> #----------------------------------------------------
```

Table A8.2 c: Composite test score effect sizes, sub-groups and other treatment arms

Study	Outcome	N	COUNTRY	ESa	vara	lower	upper
Andrabi	Village level scores	371	Pakistan	0.478755	0.005545	0.3328	0.62471
Zeitlin	Participatory scorecard treatment arm	3076	Uganda2	0.101996	0.001302	0.03127	0.172721
Bjorkman	Rural area	371	Uganda1	0.173205	0.005411	0.029028	0.317382
Reinikka	Full sample	374	Uganda1	0.129272	0.005359	-0.01421	0.272751
Reinikka	Boys	371	Uganda1	0.077876	0.005395	-0.06609	0.221838
Reinikka	Girls	373	Uganda1	0.181223	0.005384	0.037407	0.325039

Language test scores

Random-Effects Model (k = 9; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0183 (SE = 0.0120)

tau (square root of estimated tau^2 value):   0.1353

I^2 (total heterogeneity / total variability):   95.11%

H^2 (total variability / sampling variability): 20.45

Test for Heterogeneity:

Q(df = 8) = 39.7007, p-val < .0001

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
  0.1150   0.0529   2.1730   0.0298   0.0113   0.2188       *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
----------------------------------------------------
Sensitivity Analysis
----------------------------------------------------
  estimate     se   zval   pval   ci.lb  ci.ub      Q     Qp   tau2     I2     H2
1   0.0600 0.0337 1.7816 0.0748 -0.0060 0.1260 19.9553 0.0057 0.0053 86.2742  7.2856
2   0.1358 0.0609 2.2295 0.0258  0.0164 0.2552 39.6814 0.0000 0.0212 90.5246 10.5537
3   0.1271 0.0629 2.0201 0.0434  0.0038 0.2505 37.1632 0.0000 0.0231 96.0356 25.2244
4   0.1371 0.0603 2.2753 0.0229  0.0190 0.2552 35.0411 0.0000 0.0206 89.0431  9.1267
5   0.1397 0.0564 2.4784 0.0132  0.0292 0.2502 38.1295 0.0000 0.0179 95.2819 21.1950
6   0.1104 0.0536 2.0586 0.0395  0.0053 0.2155 38.7857 0.0000 0.0185 95.7225 23.3782
7   0.1137 0.0572 1.9898 0.0466  0.0017 0.2258 38.8778 0.0000 0.0203 96.0769 25.4903
8   0.1009 0.0574 1.7567 0.0790 -0.0117 0.2134 30.8712 0.0001 0.0191 95.7192 23.3604
9   0.1127 0.0609 1.8517 0.0641 -0.0066 0.2320 35.5365 0.0000 0.0219 96.2520 26.6808
> #----------------------------------------------------
```

Table A8.2 d: Language test score effect sizes, other outcomes and other treatment arms

Study	Outcome	COUNTRY	N	ESa	vara	lower	upper
Andrabi	English	Pakistan	198.8	0.393713	0.010255	0.195227	0.592199
Lassibille	French	Madagascar	7804	0.003295	0.000256	-0.02808	0.034672

Maths test scores

Random-Effects Model (k = 9; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0224 (SE = 0.0135)

tau (square root of estimated tau^2 value):   0.1496

I^2 (total heterogeneity / total variability):   96.51%

H^2 (total variability / sampling variability): 28.66

Test for Heterogeneity:

Q(df = 8) = 66.3819, p-val < .0001

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub
0.0881	0.0559	1.5755	0.1152	-0.0215	0.1976

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Sensitivity Analysis

----

	estimate	se	zval	pval	ci.lb	ci.ub	Q	Qp	tau2	I2	H2
1	0.1013	0.0648	1.5614	0.1184	-0.0258	0.2283	63.8728	0.0000	0.0266	94.7646	19.1006
2	0.0498	0.0409	1.2189	0.2229	-0.0303	0.1299	47.5675	0.0000	0.0098	93.1763	14.6547
3	0.0815	0.0657	1.2400	0.2150	-0.0473	0.2103	57.2145	0.0000	0.0276	97.1857	35.5330
4	0.0993	0.0655	1.5166	0.1294	-0.0290	0.2277	63.3925	0.0000	0.0271	93.9879	16.6331
5	0.0704	0.0619	1.1373	0.2554	-0.0509	0.1916	40.4269	0.0000	0.0240	96.7751	31.0085
6	0.0785	0.0616	1.2734	0.2029	-0.0423	0.1993	64.4914	0.0000	0.0251	97.2281	36.0761
7	0.0986	0.0563	1.7498	0.0802	-0.0118	0.2090	64.9334	0.0000	0.0222	96.9100	32.3628
8	0.1144	0.0558	2.0482	0.0405	0.0049	0.2238	59.4798	0.0000	0.0192	96.2255	26.4935
9	0.1031	0.0635	1.6233	0.1045	-0.0214	0.2277	65.4230	0.0000	0.0256	97.0832	34.2838

> #---

8.3 Public Private Partnership

Enrolment

Random-Effects Model (k = 6; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0135)

tau (square root of estimated tau^2 value): 0

I^2 (total heterogeneity / total variability): 0.00%

H^2 (total variability / sampling variability): 1.00

Test for Heterogeneity:

Q(df = 5) = 4.5283, p-val = 0.4761

Model Results:

estimate se zval pval ci.lb ci.ub

```
  0.1190   0.0597   1.9948   0.0461   0.0021   0.2359       *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #---------------------------------------------------
> # Sensitivity Analysis
> #---------------------------------------------------
>   leave1out(fit1)
  estimate    se   zval   pval   ci.lb  ci.ub     Q     Qp   tau2     I2     H2
1   0.0943 0.0653 1.4433 0.1489 -0.0338 0.2223 3.6663 0.4530 0.0000  0.0000 1.0000
2   0.1442 0.0654 2.2054 0.0274  0.0160 0.2723 3.6420 0.4566 0.0000  0.0000 1.0000
3   0.1471 0.0654 2.2502 0.0244  0.0190 0.2753 3.4238 0.4896 0.0000  0.0000 1.0000
4   0.1143 0.0692 1.6509 0.0987 -0.0214 0.2500 4.4959 0.3430 0.0026 10.8477 1.1217
5   0.1348 0.0668 2.0173 0.0437  0.0038 0.2657 4.1817 0.3820 0.0009  4.2449 1.0443
6   0.0797 0.0653 1.2198 0.2225 -0.0483 0.2076 2.3286 0.6756 0.0000  0.0000 1.0000
> #---------------------------------------------------
```

Table A8.3 a: enrolment effect sizes, subgroups and other treatment arms not included in the meta-analysis

Study	Outcome	COUNTRY	N	ESa	vara	lower	upper
Barrera-Osorio	Total enrolment check 1 (4 month follow-up)	Uganda	94	0.248041	0.021440	-0.038952	0.535034
Barrera-Osorio	Grade subgroup- 1 (check 1)	Uganda	94	0.473067	0.021872	0.183200	0.762933
Barrera-Osorio	Grade subgroup- 2 (check 1)	Uganda	94	0.051295	0.021284	-0.234647	0.337238
Barrera-Osorio	Grade subgroup- 3 (check 1)	Uganda	94	0.174045	0.021357	-0.112391	0.460481
Barrera-Osorio	Grade subgroup- 4 (check 1)	Uganda	94	0.215290	0.021400	-0.071433	0.502012
Barrera-Osorio	Grade subgroup- 5 (check 1)	Uganda	94	0.141441	0.021330	-0.144812	0.427693
Barrera-Osorio	Grade subgroup- 6 (check 1)	Uganda	94	0.070350	0.021290	-0.215634	0.356333
Barrera-Osorio	Total enrolment check 2 (6 month follow-up)	Uganda	93.75	0.322052	0.021726	0.033154	0.610950
Barrera-Osorio	Grade subgroup- 1 (check 2)	Uganda	93.75	0.460910	0.022017	0.170081	0.751739
Barrera-Osorio	Grade subgroup- 2 (check 2)	Uganda	93.75	0.171716	0.021527	-0.115855	0.459287

Barrera-Osorio	Grade subgroup- 3 (check 2)	Uganda	93.75	0.289259	0.021672	0.000719	0.577799
Barrera-Osorio	Grade subgroup- 4 (check 2)	Uganda	93.75	0.275743	0.021652	-0.012661	0.564146
Barrera-Osorio	Grade subgroup- 5 (check 2)	Uganda	93.75	0.228826	0.021588	-0.059154	0.516807
Barrera-Osorio	Grade subgroup- 6 (check 2)	Uganda	93.75	0.072097	0.021462	-0.215039	0.359233
Barrera-Osorio	Grade subgroup- 1 (check 3) (11 month follow-up)	Uganda	94	0.381160	0.021663	0.092680	0.669640
Barrera-Osorio	Grade subgroup- 2 (check 3)	Uganda	94	0.472282	0.021870	0.182428	0.762135
Barrera-Osorio	Grade subgroup- 3 (check 3)	Uganda	94	0.139180	0.021328	-0.147061	0.425422
Barrera-Osorio	Grade subgroup- 4 (check 3)	Uganda	94	0.147885	0.021335	-0.138400	0.434171
Barrera-Osorio	Grade subgroup- 5 (check 3)	Uganda	94	0.218708	0.021404	-0.068041	0.505457
Barrera-Osorio	Grade subgroup- 6 (check 4)	Uganda	94	0.174686	0.021358	-0.111754	0.461126

Girls' Enrolment

```
Random-Effects Model (k = 7; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0051 (SE = 0.0153)

tau (square root of estimated tau^2 value):      0.0711

I^2 (total heterogeneity / total variability):   19.07%

H^2 (total variability / sampling variability):  1.24

Test for Heterogeneity:

Q(df = 6) = 7.4517, p-val = 0.2811

Model Results:

estimate      se     zval     pval    ci.lb    ci.ub

  0.2185   0.0616   3.5499   0.0004   0.0979   0.3392      ***

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> #---------------------------------------------------

> # Sensitivity Analysis

> #---------------------------------------------------

>   leave1out(fit1)
```

```
  estimate     se    zval    pval   ci.lb   ci.ub      Q      Qp    tau2      I2     H2
1   0.2555 0.0599 4.2684 0.0000 0.1382 0.3728 4.7641 0.4453 0.0000  0.0000 1.0000
2   0.1974 0.0683 2.8912 0.0038 0.0636 0.3311 6.5626 0.2553 0.0065 23.2985 1.3038
3   0.1721 0.0597 2.8826 0.0039 0.0551 0.2891 3.1769 0.6727 0.0000  0.0000 1.0000
4   0.2309 0.0715 3.2269 0.0013 0.0906 0.3711 7.1647 0.2087 0.0092 30.0375 1.4293
5   0.2287 0.0720 3.1752 0.0015 0.0875 0.3698 7.2599 0.2020 0.0096 30.9587 1.4484
6   0.2144 0.0727 2.9478 0.0032 0.0718 0.3569 7.4140 0.1916 0.0103 32.3523 1.4782
7   0.2314 0.0714 3.2407 0.0012 0.0915 0.3714 7.1377 0.2106 0.0091 29.7724 1.4239
> #---------------------------------------------------
```

Table A8.3 b: girls' enrolment effect sizes, subgroups and other treatment arms not included in the meta-analysis

Study	Outcome	COUNTRY	N	ESa	vara	lower	upper
Barrera-Osorio	Grade subgroup - 1 (check 1)	Uganda	94	0.514278	0.021980	0.223695	0.804861
Barrera-Osorio	Grade subgroup - 2 (check 1)	Uganda	94	0.109632	0.021309	-0.176478	0.395742
Barrera-Osorio	Grade subgroup - 3 (check 1)	Uganda	94	0.306668	0.021527	0.019097	0.594239
Barrera-Osorio	Grade subgroup - 4 (check 1)	Uganda	93	Inf	Inf	NA	Inf
Barrera-Osorio	Grade subgroup - 5 (check 1)	Uganda	94	0.146793	0.021334	-0.139487	0.433073
Barrera-Osorio	Grade subgroup - 6 (check 1)	Uganda	94	0.057482	0.021285	-0.228472	0.343437
Barrera-Osorio	Grade subgroup - 1 (check 2)	Uganda	93	0.486954	0.022143	0.195297	0.778611
Barrera-Osorio	Grade subgroup - 2 (check 2)	Uganda	93	0.246959	0.021669	-0.041563	0.535481
Barrera-Osorio	Grade subgroup - 3 (check 2)	Uganda	93	0.373794	0.021881	0.083866	0.663721
Barrera-Osorio	Grade subgroup - 4 (check 2)	Uganda	93	0.345920	0.021827	0.056350	0.635490
Barrera-Osorio	Grade subgroup - 5 (check 2)	Uganda	93	0.252666	0.021677	-0.035907	0.541239
Barrera-Osorio	Grade subgroup - 6 (check 2)	Uganda	93	0.071501	0.021519	-0.216019	0.359022

Completion

```
Random-Effects Model (k = 2; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0047 (SE = 0.0079)
tau (square root of estimated tau^2 value):      0.0683
I^2 (total heterogeneity / total variability):   83.12%
H^2 (total variability / sampling variability):  5.93

Test for Heterogeneity:
Q(df = 1) = 5.9258, p-val = 0.0149

Model Results:

estimate      se    zval    pval   ci.lb   ci.ub
  0.0757  0.0530  1.4294  0.1529 -0.0281  0.1795

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

No sensitivity analysis as only 2 studies

Composite test scores

```
Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0133 (SE = 0.0136)
tau (square root of estimated tau^2 value):      0.1155
I^2 (total heterogeneity / total variability):   97.85%
H^2 (total variability / sampling variability):  46.52

Test for Heterogeneity:
Q(df = 2) = 85.5202, p-val < .0001

Model Results:

estimate      se    zval    pval   ci.lb   ci.ub
  0.0651  0.0674  0.9659  0.3341 -0.0670  0.1973

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
  estimate    se   zval   pval   ci.lb  ci.ub      Q     Qp   tau2     I2     H2
```

```
1   0.1033 0.0970  1.0649 0.2869 -0.0868 0.2934 49.8961 0.0000 0.0184 97.9958 49.8961
2  -0.0049 0.0108 -0.4570 0.6477 -0.0262 0.0163  0.4885 0.4846 0.0000  0.0000  1.0000
3   0.0946 0.1052  0.8998 0.3682 -0.1115 0.3008 80.6508 0.0000 0.0218 98.7601 80.6508
> #-----------------------------------------------------
```

Table A8.3 c: composite scores effect sizes, subgroups and other treatment arms not included in the meta-analysis

Study	Outcome	COUNTRY	N	ESa	vara	lower	upper
Barrera-Osorio & Raju	QAT score	Pakistan	11730	0.00331	0.00017	-0.02228	0.02891
Muralidharan	4 year follow-up	India	5316	0.076925	0.000377	0.038894	0.114956

Language Arts

```
Random-Effects Model (k = 7; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.0031 (SE = 0.0020)
tau (square root of estimated tau^2 value):      0.0560
I^2 (total heterogeneity / total variability):   94.32%
H^2 (total variability / sampling variability):  17.59
Test for Heterogeneity:
Q(df = 6) = 36.1790, p-val < .0001
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.0440   0.0224   1.9652   0.0494   0.0001   0.0879       *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>  leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub      Q     Qp   tau2      I2      H2
1   0.0485 0.0272 1.7846 0.0743 -0.0048 0.1018 35.6749 0.0000 0.0040 94.8820 19.5390
2   0.0462 0.0275 1.6821 0.0926 -0.0076 0.1000 35.6860 0.0000 0.0041 94.9937 19.9749
3   0.0560 0.0203 2.7606 0.0058  0.0162 0.0958 22.0780 0.0005 0.0021 92.3815 13.1259
```

4 0.0461 0.0275 1.6795 0.0930 -0.0077 0.1000 35.6297 0.0000 0.0041 94.9949 19.9794

5 0.0484 0.0272 1.7769 0.0756 -0.0050 0.1017 35.8029 0.0000 0.0040 94.8943 19.5859

6 0.0400 0.0256 1.5608 0.1186 -0.0102 0.0902 34.0762 0.0000 0.0036 95.7155 23.3400

7 0.0251 0.0118 2.1225 0.0338 0.0019 0.0484 16.2594 0.0061 0.0006 79.0418 4.7714

> #--

Table A8.3 d: Language Arts effect sizes, subgroups and other treatment arms not included in the meta-analysis

Study	Outcome	COUNTRY	N	ESa	vara	lower	upper
Muralidharan	Telugu 4 year follow up	India	5316	-0.008968	0.000376	-0.046985	0.029050
Muralidharan	English 2 year follow up	India	5316	0.062153	0.000376	0.024127	0.100179
Muralidharan	English 4 year follow up	India	5316	0.043432	0.000376	0.005410	0.081454
Muralidharan	Hindi 4 year follow up	India	5316	0.212185	0.000378	0.174061	0.250309
Saavedra	Subgroup 14_16y	Chile	17650	0.041997	0.000113	0.021131	0.062864
Saavedra	Subgroup 5_9y	Chile	7606	0.044476	0.000263	0.012690	0.076263
Angrist	Subgroup girls	Colombia	1968	0.124550	0.001018	0.062007	0.187093
Angrist	Subgroup boys	Colombia	1968	0.238161	0.001023	0.175457	0.300865
Barrera-Osorio	2011 English scores	Uganda	2085	0.024820	0.000959	-0.035886	0.085527

Maths

Random-Effects Model (k = 7; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0018 (SE = 0.0012)

tau (square root of estimated tau^2 value): 0.0420

I^2 (total heterogeneity / total variability): 90.67%

H^2 (total variability / sampling variability): 10.72

Test for Heterogeneity:

Q(df = 6) = 39.8840, p-val < .0001

Model Results:

estimate se zval pval ci.lb ci.ub

0.0464 0.0172 2.7054 0.0068 0.0128 0.0800 **

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
>   leave1out(fit1)
  estimate     se   zval    pval  ci.lb  ci.ub       Q     Qp   tau2      I2      H2
1   0.0470 0.0205 2.2941 0.0218 0.0068 0.0871 39.8603 0.0000 0.0022 93.3218 14.9740
2   0.0401 0.0191 2.1045 0.0353 0.0028 0.0774 23.4259 0.0003 0.0018 90.0384 10.0385
3   0.0473 0.0211 2.2372 0.0253 0.0059 0.0888 39.8219 0.0000 0.0023 91.9609 12.4392
4   0.0505 0.0207 2.4363 0.0148 0.0099 0.0911 37.1476 0.0000 0.0022 91.6049 11.9117
5   0.0558 0.0150 3.7255 0.0002 0.0265 0.0852 27.8222 0.0000 0.0011 86.6565  7.4943
6   0.0508 0.0206 2.4681 0.0136 0.0105 0.0912 36.3244 0.0000 0.0022 91.5061 11.7731
7   0.0352 0.0134 2.6227 0.0087 0.0089 0.0615 28.9342 0.0000 0.0009 84.7992  6.5786
> #----------------------------------------------------
```

Table A8.3 e: Maths effect sizes, subgroups and other treatment arms not included in the meta-analysis

Study	Outcome	COUNTRY	N	ESa	vara	lower	upper
Saavedra	Age subgroup - 5_9 yrs	Chile	7606	0.079439	0.000263	0.047644	0.111234
Saavedra	Age subgroup - 10_13 yrs	Chile	4950	0.074708	0.000404	0.035296	0.114119
Saavedra	Age subgroup - 14_16 yrs	Chile	17900	0.076157	0.000112	0.055432	0.096883
Muralidharan	4 year follow up	India	5316	-0.016044	0.000376	-0.054062	0.021973
Angrist	Gender subgroup - girls	Colombia	2056	0.406173	0.000993	0.344415	0.467930
Angrist	Gender subgroup- boys	Colombia	2056	0.003696	0.000973	-0.057435	0.064826
Barrera-Osorio	Math scores 2011	Uganda	2085	0.047085	0.000959	-0.013627	0.107798
Barrera-Osorio	Math scores 2011 (independent)	Uganda	2085	0.107676	0.000961	0.046928	0.168424
Barrera-Osorio	Math scores 2012 (independent)	Uganda	2056	0.102790	0.000974	0.041619	0.163961

Table A8.3 f: Miscellaneous Access effect sizes, subgroups and other treatment arms not included in the meta-analysis

Study	Outcome	N	COUNTRY	ESa	vara	lower	upper
Barrera-Osorio	Drop-out	2145	Colombia	-0.183637	0.000936	-0.243612	-0.123662

9. Multilevel interventions

Attendance

Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0591 (SE = 0.0605)

tau (square root of estimated tau^2 value):   0.2431

I^2 (total heterogeneity / total variability):   99.89%

H^2 (total variability / sampling variability): 893.95

Test for Heterogeneity:

Q(df = 2) = 49.7403, p-val < .0001

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
  0.1570   0.1419   1.1058   0.2688  -0.1212   0.4351
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #----------------------------------------------------
> # Sensitivity Analysis
> #----------------------------------------------------
>   leave1out(fit1)
  estimate     se   zval   pval   ci.lb  ci.ub       Q     Qp   tau2      I2      H2
1   0.2352 0.2136 1.1008 0.2710 -0.1836 0.6539 44.1893 0.0000 0.0893 97.7370 44.1893
2   0.2273 0.2217 1.0252 0.3053 -0.2072 0.6618 47.5122 0.0000 0.0963 97.8953 47.5122
3   0.0184 0.0081 2.2803 0.0226  0.0026 0.0342  3.8785 0.0489 0.0001 74.2166  3.8785
> #----------------------------------------------------
```

Table A9 a: Attendance samples not included in the meta-analysis

Study	Outcome	COUNTRY	N	ESa	vara	lower	upper
Rosati	8-11 year old children - Marginal Effects	Mexico	13	NA	NA	NA	NA
Rosati	12-16 year old children - Marginal Effects	Mexico	13	NA	NA	NA	NA
De Hoop	Attendance - Child was in school the last day school was in session	Burkina Faso	29	0.595238	0.020886	0.311981	0.878496

Dropout

```
Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0202 (SE = 0.0238)

tau (square root of estimated tau^2 value):      0.1420

I^2 (total heterogeneity / total variability):   98.79%

H^2 (total variability / sampling variability):  82.83

Test for Heterogeneity:

Q(df = 2) = 36.3297, p-val < .0001

Model Results:

estimate      se     zval     pval    ci.lb    ci.ub

 -0.1553   0.0886  -1.7515   0.0799  -0.3290   0.0185        .

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> #----------------------------------------------------

> # Sensitivity Analysis

> #----------------------------------------------------

>   leave1out(fit1)

  estimate     se     zval   pval   ci.lb   ci.ub      Q     Qp   tau2     I2      H2

1  -0.2002 0.1782 -1.1234 0.2613 -0.5494  0.1491  8.6300 0.0033 0.0569 88.4125  8.6300

2  -0.2367 0.1337 -1.7707 0.0766 -0.4987  0.0253  4.9902 0.0255 0.0297 79.9607  4.9902

3  -0.0843 0.0430 -1.9601 0.0500 -0.1685 -0.0000 29.6497 0.0000 0.0036 96.6273 29.6497
```

Completion

Random-Effects Model (k = 4; tau^2 estimator: REML)

```
tau^2 (estimated amount of total heterogeneity): 0.0070 (SE = 0.0059)
tau (square root of estimated tau^2 value):      0.0839
I^2 (total heterogeneity / total variability):   97.77%
H^2 (total variability / sampling variability):  44.84
Test for Heterogeneity:
Q(df = 3) = 77.1847, p-val < .0001
Model Results:
estimate      se     zval     pval    ci.lb    ci.ub
  0.1279   0.0427   2.9959   0.0027   0.0442   0.2115       **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #-----------------------------------------------------
> # Sensitivity Analysis
> #-----------------------------------------------------
>   leave1out(fit1)
  estimate     se   zval   pval  ci.lb  ci.ub       Q     Qp   tau2      I2      H2
1   0.1250 0.0608 2.0558 0.0398 0.0058 0.2441 60.7188 0.0000 0.0108 98.2819 58.2041
2   0.1582 0.0424 3.7307 0.0002 0.0751 0.2412 30.5637 0.0000 0.0051 96.4928 28.5129
3   0.0915 0.0296 3.0926 0.0020 0.0335 0.1495 42.2180 0.0000 0.0025 95.4281 21.8728
4   0.1386 0.0595 2.3300 0.0198 0.0220 0.2551 77.0663 0.0000 0.0103 98.1375 53.6905
> #-----------------------------------------------------
```

Table A9 b: Child Scholarship program (CSP) Kenya: completion and subgroup effect sizes

Study	Outcome/subgroup	COUNTRY	N	ESa	vara	lower	upper
Kremer	Grades advanced/Grade 1 prior to prog.	Kenya	558	0.167556	0.003597	0.050008	0.285104
Kremer	Grades advanced/Grade 2 prior to prog.	Kenya	579	0.256143	0.003483	0.140478	0.371809
Kremer	Grades advanced/Grade 3 prior to prog.	Kenya	510	0.11633	0.003928	-0.00651	0.239174
Kremer	Grades advanced/Grade 4 prior to prog.	Kenya	575	0.047881	0.003479	-0.06773	0.163492
Kremer	Grades advanced/Grade 5 prior to prog.	Kenya	461	0.222026	0.005043	0.082833	0.36122

Kremer	Grades advanced/Grade 6 prior to prog.	Kenya	461	0.269346	0.006097	0.116302	0.42239
Kremer	Grades advanced/Grade 7 prior to prog.	Kenya	461	-0.04043	0.007144	-0.2061	0.125234
Kremer	Years enrolled/Grade 1 prior to prog.	Kenya	461	0.365017	0.004411	0.234848	0.495186
Kremer	Years enrolled/Grade 2 prior to prog.	Kenya	461	0.25873	0.004375	0.129093	0.388367
Kremer	Years enrolled/Grade 3 prior to prog.	Kenya	461	0.348113	0.004404	0.21804	0.478185
Kremer	Years enrolled/Grade 4 prior to prog.	Kenya	461	0.12058	0.004346	-0.00864	0.249795
Kremer	Years enrolled/Grade 5 prior to prog.	Kenya	461	0.370885	0.004413	0.240682	0.501088
Kremer	Years enrolled/Grade 6 prior to prog.	Kenya	461	0.300144	0.004387	0.170321	0.429968
Kremer	Years enrolled/Grade 7 prior to prog.	Kenya	461	0.03721	0.004339	-0.0919	0.16632
Kremer	Years enrolled	Kenya	1906	0.365851	0.000629	0.316688	0.415014

Enrolment

Table A9 c: BRIGHT program Burkina Faso: enrolment and subgroup effect sizes

Study	Outcome/subgroup	COUNTRY	N	ESa	vara	lower	upper
Kazianga	Enrolment (full sample)	Burkina Faso	9158	0.110405	0.000111	0.089712	0.131098
Kazianga	Enrolment (age 6)	Burkina Faso	2567	0.194826	0.000783	0.139987	0.249664
Kazianga	Enrolment (age 7)	Burkina Faso	2567	0.190633	0.000783	0.1358	0.245466
Kazianga	Enrolment (age 8)	Burkina Faso	2567	0.222304	0.000784	0.167426	0.277182
Kazianga	Enrolment (age 9)	Burkina Faso	2567	0.263915	0.000786	0.208969	0.318862
Kazianga	Enrolment (age 10)	Burkina Faso	2567	0.216013	0.000784	0.161145	0.270882
Kazianga	Enrolment (age 11)	Burkina Faso	2567	0.135091	0.000781	0.080319	0.189862

Kazianga	Enrolment (age 12)	Burkina Faso	2567	0.151531	0.000781	0.096744	0.206319
De Hoop	Enrolment (boys w/o female sibling)	Burkina Faso	1334	0.488	0.020595	0.206719	0.769281
De Hoop	Enrolment (boys w/ female sibling)	Burkina Faso	1334	0.736364	0.021356	0.449938	1.022789
De Hoop	Enrolment (girls)	Burkina Faso	1334	0.644	0.021037	0.35972	0.92828

> #---

Composite scores

Random-Effects Model (k = 3; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0077 (SE = 0.0079)

tau (square root of estimated tau^2 value): 0.0877

I^2 (total heterogeneity / total variability): 98.03%

H^2 (total variability / sampling variability): 50.70

Test for Heterogeneity:

Q(df = 2) = 107.2985, p-val < .0001

Model Results:

```
estimate      se     zval     pval    ci.lb    ci.ub
  0.0222   0.0514   0.4325   0.6653  -0.0785   0.1229
```

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #---

> # Sensitivity Analysis

> #---

> leave1out(fit1)

```
  estimate     se    zval   pval   ci.lb  ci.ub       Q      Qp   tau2       I2      H2
1  -0.0232 0.0155 -1.4993 0.1338 -0.0536 0.0071  1.9340 0.1643 0.0003 48.2940  1.9340
2   0.0392 0.0839  0.4678 0.6399 -0.1252 0.2037 47.4279 0.0000 0.0138 97.8915 47.4279
3   0.0546 0.0673  0.8118 0.4169 -0.0773 0.1866 91.2977 0.0000 0.0090 98.9047 91.2977
```

Table A9 d: BRIGHT School test score effect sizes & subgroups not included in meta-analysis

Study	Outcome/subgroup	COUNTRY	N	ESa	vara	lower	upper
Kazianga	Total Test score (age 6)	2567	Burkina Faso	0.202707	0.000783	0.147858	0.257556
Kazianga	Total Test score (age 7)	2567	Burkina Faso	0.236221	0.000785	0.181321	0.29112
Kazianga	Total Test score (age 8)	2567	Burkina Faso	0.232746	0.000784	0.177852	0.28764
Kazianga	Total Test score (age 9)	2567	Burkina Faso	0.288687	0.000787	0.233694	0.34368
Kazianga	Total Test score (age 10)	2567	Burkina Faso	0.223250	0.000784	0.168371	0.27813
Kazianga	Total Test score (age 11)	2567	Burkina Faso	0.128753	0.000781	0.073988	0.183519
Kazianga	Total Test score (age 12)	2567	Burkina Faso	0.119785	0.000781	0.065027	0.174543

Math scores

Random-Effects Model (k = 10; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.2766 (SE = 0.1306)

tau (square root of estimated tau^2 value):   0.5260

I^2 (total heterogeneity / total variability):   99.89%

H^2 (total variability / sampling variability): 921.47

Test for Heterogeneity:

Q(df = 9) = 7511.6006, p-val < .0001

Model Results:

estimate   se   zval   pval   ci.lb   ci.ub

  0.1567   0.1665   0.9414   0.3465 -0.1695   0.4830

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #---

> # Sensitivity Analysis

> #---

>   leave1out(fit1)

estimate se zval pval ci.lb ci.ub Q Qp tau2 I2 H2

1 0.1609 0.1861 0.8650 0.3870 -0.2037 0.5256 7383.5251 0.0000 0.3111 99.8843 863.9596

2 0.1589 0.1861 0.8541 0.3930 -0.2058 0.5236 7491.8546 0.0000 0.3112 99.9099 1109.4915

3 0.1619 0.1860 0.8702 0.3842 -0.2027 0.5265 7490.9716 0.0000 0.3110 99.9068 1073.1658

4 0.2665 0.1398 1.9064 0.0566 -0.0075 0.5406 2593.5026 0.0000 0.1754 99.8125 533.2038

5 0.1730 0.1852 0.9339 0.3503 -0.1901 0.5360 7511.4133 0.0000 0.3083 99.8924 929.0465

6 0.0641 0.1546 0.4144 0.6786 -0.2390 0.3672 5777.4140 0.0000 0.2148 99.8673 753.3726

7 0.0628 0.1536 0.4085 0.6829 -0.2384 0.3639 5888.6368 0.0000 0.2120 99.8663 747.7292

8 0.1759 0.1849 0.9517 0.3412 -0.1864 0.5382 7510.0186 0.0000 0.3071 99.9079 1086.3019

9 0.1816 0.1840 0.9871 0.3236 -0.1790 0.5422 7502.3207 0.0000 0.3043 99.9078 1084.9341

10 0.1614 0.1860 0.8677 0.3855 -0.2032 0.5261 7453.7228 0.0000 0.3110 99.8964 965.7142

> #---

Table A9 e: Math scores effect sizes and subgroups

Study	Outcome	COUNTRY	N	ESa	vara	lower	upper
Chay	Math Scores (year 1988 - 90)	2644	Chile	0.018347	0.000756	-0.03556	0.072254
Bellei	Maths Scores (1 year follow-up)	10360	Chile	0.037288	0.00019	0.010292	0.064284
De Hoop	Maths scores (girls)	2587	Burkina Faso	0.558954	0.008312	0.380256	0.737652
De Hoop	Maths scores (boys with female sibling)	2587	Burkina Faso	0.568237	0.008323	0.389426	0.747048
De Hoop	Maths scores (boys without female sibling)	2587	Burkina Faso	0.557099	0.00831	0.378423	0.735775

Language scores

Random-Effects Model (k = 14; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.1629 (SE = 0.0645)

tau (square root of estimated tau^2 value): 0.4037

I^2 (total heterogeneity / total variability): 99.56%

H^2 (total variability / sampling variability): 226.14

Test for Heterogeneity:

Q(df = 13) = 3660.4901, p-val < .0001

Model Results:

```
estimate      se    zval    pval    ci.lb    ci.ub
  0.0430  0.1084  0.3969  0.6914  -0.1694   0.2554
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #--------------------------------------------------
> # Sensitivity Analysis
> #--------------------------------------------------
>   leave1out(fit1)
   estimate     se    zval   pval   ci.lb  ci.ub        Q     Qp   tau2      I2      H2
1    0.0372 0.1169  0.3179 0.7505 -0.1919 0.2663 3648.4906 0.0000 0.1761 99.5733
234.3810
2    0.0401 0.1170  0.3430 0.7316 -0.1892 0.2695 3657.9313 0.0000 0.1765 99.5742
234.8694
3    0.0341 0.1166  0.2920 0.7703 -0.1946 0.2627 3643.2449 0.0000 0.1754 99.5904
244.1699
4    0.0481 0.1165  0.4133 0.6794 -0.1802 0.2764 3659.9191 0.0000 0.1755 99.6184
262.0671
5    0.1372 0.0613  2.2375 0.0253  0.0170 0.2574  802.8955 0.0000 0.0474 98.4582
64.8599
6    0.0438 0.1171  0.3739 0.7085 -0.1857 0.2732 3660.0386 0.0000 0.1766 99.5745
235.0009
7    0.0193 0.1142  0.1688 0.8659 -0.2046 0.2432 3465.8634 0.0000 0.1681 99.5577
226.0796
8   -0.0101 0.1022 -0.0986 0.9215 -0.2104 0.1903 2806.2112 0.0000 0.1343 99.4541
183.1774
9    0.0404 0.1169  0.3456 0.7297 -0.1888 0.2696 3660.0822 0.0000 0.1763 99.6123
257.9404
10   0.0620 0.1152  0.5385 0.5902 -0.1638 0.2878 3612.2011 0.0000 0.1711 99.5935
246.0077
```

11 0.0317 0.1163 0.2723 0.7854 -0.1964 0.2597 3649.6928 0.0000 0.1745 99.6069 254.3848

12 0.0370 0.1169 0.3170 0.7513 -0.1920 0.2661 3647.9774 0.0000 0.1760 99.5733 234.3548

13 0.0373 0.1169 0.3191 0.7497 -0.1918 0.2664 3652.5818 0.0000 0.1761 99.5873 242.2928

14 0.0456 0.1170 0.3894 0.6970 -0.1838 0.2749 3658.5842 0.0000 0.1765 99.5930 245.6888

> #---

Table A9 f: language scores effect sizes, subgroups and other treatment arms not included in the meta-analysis

Study	Outcome/subgroup	COUNTRY	N	ESa	vara	lower	upper
Chay	Language scores (1988-90)	2644	Chile	0.042864	0.000757	-0.01105	0.096777
Lockheed	School mean grade 4 literacy, % near mastery	206.8	Jamaica	-0.002113	0.015873	-0.249050	0.244824
Lockheed	School mean grade 4 literacy, % non-mastery	206.8	Jamaica	0.055323	0.015879	-0.191661	0.302308
Paqueo	Indigenous-Rural sample	449	Mexico	0.344508	0.004520	0.212729	0.476287
De Hoop	Maths scores (boys without female sibling)	960.5	Burkina Faso	0.180165	0.020081	-0.09758	0.457913
De Hoop	Maths scores (girls)	960.5	Burkina Faso	0.570455	0.020814	0.287687	0.853222
De Hoop	Maths scores (boys with female sibling)	960.5	Burkina Faso	0.711111	0.021264	0.425299	0.996923

Appendix I: Full qualitative synthesis

Child Level

School Based Health Interventions

In the following section, we report the results of the synthesis of qualitative findings from our included school- based health programmes, presented using the hypothesised programme theory as an overall framework for structuring the synthesis. This first section looks at intervention and implementation features that may be associated with relative success and failure, followed by a summary of the contextual barriers and facilitators to success.

We identified only three additional documents (three project documents) related to two programmes covered by the included impact evaluations. For most programmes, there was limited additional information identified for the included programmes outside of the impact evaluations themselves and therefore following synthesis is mainly based on information provided in the impact evaluations. The lack of qualitative information for this intervention type may be due to the fact that the majority of included studies are RCTs often evaluating a one- off trial to inform future roll out of a government-run programme.

Process and Implementation

Delivery of treatment

In most of the programmes teachers were responsible for delivering the specific health intervention to the students in the school (n= 9). In general, school staff had been trained previously by the research team, implementers or public health officials on how to deliver the treatment and were monitored through regular spot checks. An exception is the Primary School Deworming Project in Kenya in which public health nurses and officers delivered the deworming treatment.

In a context where school resources are limited, an additional emphasis on implementing school based health interventions could draw attention (resources, time etc.) away from education due to a multitasking effort by school staff. Similarly, weak incentives for teachers to improve health may be keeping SBH programmes from reaching their full potential due to reduced compliance or diversion of resources to traditional responsibilities. For some health treatments, the community might not consider it acceptable to use teachers to deliver them.

Several studies report on the use of school staff as the main vehicle to deliver the SBH intervention:

Teachers in Kenya considered delivering the malaria control programme discruptive and beyond the scope of their work (Brooker et al. 2015; IE)

In Kenya, the use of teachers to deliver the testing part of the intervention was not considered acceptable because the intervention included taking blood samples. This was thought to be beyond the teachers' scope, overburdening them and undermining their ability to carry out their regular duties. The use of teachers to deliver other parts of the treatment was considered acceptable. The taking of blood samples caused the most concern in the population, resulting in rumours that the testing was covert HIV testing. Such rumours regarding blood are common with health interventions are common in Sub-Saharan Africa (Brooker et al. 2015).

The use of health workers to implement the SBH intervention may be critical in terms of training and supervising teachers and handling referral cases (Brooker et al. 2015 IE)

Based on focus group discussions and in depth interviews, Brooker et al. find that the use of health workers to implement the SBH intervention was critical to successful implementation in schools. Their involvement is particularly necessary in terms of training and supervising the delivery agents implementing the intervention in the school, in facilitating safe waste disposal, and in handling referral cases arising from schools (Brooker et al. 2015 IE).

Compliance with treatment

A number of studies also report on compliance rates with the treatment regime as well as to what extent treatment was delivered as intended.

Treatment may not have been delivered as intended, with divergence from treatment regime and target population (Miguel et al. 2004 (IE); Clarke et al. (IE), Mahawithanage et al. 2007, Fernando et al. 2006 IE))

Clarke et al. report that most children had received treatment on all three occasions- more than 7 days of treatment were completed by 1946 [75%] and 1990 [86%] in each group (p. 132) in each group) although only 1070 (41%) of children in the intervention group and 1394 (60%) in the placebo group had received treatment on all 9 days and were included in the per-protocol analysis of health outcomes. Those children who completed treatment on all 9 days were evaluated in the per-protocol analysis (ATET) (Clarke et al. 2008).

Mahawithanage et al. also report that not all children received a Vitamin A capsule every 4 months as intended: 25 children (8.4%) received one dose, 75 (25.3%) received two doses and 197 (66.3%) children received all three doses (Mahawithanage et al. 2007)

Several studies elucidated reasons for lack of compliance with the treatment regime, for example absence from school on the day of drug administration, not having breakfast, having a headache, or being on medication for some other condition can cause of drug noncompliance.

In the case of the Primary School deworming Project in Kenya, 78% of those students assigned to receive treatment received at least some medical treatment through the program in 1998, 72% in 1999. 19% of girls 13 or older also received medical treatment due to confusion in the field about students' age and confusion amongst the Ministry of Health nurses administering the drugs. The authors note that one possible explanation for the smaller impact of the program on school participation in 1999 is the lower proportion of pupils taking deworming drugs compared to 1998, which should reduce both treatment effects on the treated and externality effects. Furthermore, since approximately 80 percent of the students enrolled prior to the start of the program were present in school on a typical day in 1998, absence from school on the day of drug administration was a major cause of drug noncompliance (Miguel et al. 2004).

Similarly, in Sri Lanka, children received varying amount of chloroquine and placebo tablets during the intervention period as a result of various reasons such as being absent, not having breakfast, having a headache, or being on medication for some other condition such as a common cold (Fernando et al. 2006).

There was a decline in full supervision (a proxy for compliance) with time due to logistical difficulties in providing the complex treatment regime on more than one visit (Brooker et al 2015).

The Intermittent Screening and Treatment for Malaria trial in Kenya observed an apparent decline in full supervision (a proxy for compliance) with time, falling from 96.9% at the first round to 81.7% at the fifth round. The authors note that it became logistically difficult for children who were absent on screening day and subsequently treated on a repeat visit to be followed up on treatment day two and three by the nurse (Halliday et al. 2014 IE)

High compliance rates for delivery of treatment (Kleiman- Weiner et al. 2015 IE; Luo et al. 2012 IE; Wong et al. 2014)

In contrast, Kleiman- Weiner et al. report high compliance level with treatment. About once per month, the research team sent out inspectors to undertake unannounced compliance checks in the chewable vitamin schools. According to these observations and interviews with the teachers who dispersed the vitamins, the compliance level was almost 100%. (Kleiman-Weiner et al. 2013). Similarly, based on interviews during compliance checks, Luo et al. find that compliance of dispensing iron supplements to be almost 100% (Luo et al. 2012). In Wong et al.'s study inspectors interviewed the sample children, their parents, teachers and teacher who were not part of the intervention during surprise inspection visits. According to the interviews, the level of compliance was nearly perfect.

Project management

In China, highly decentralised implementation of the programme and poor project management might have resulted in the distribution of eyeglasses in ways inconsistent with project criteria (Gleewe et al. 2014 IE)

In China, Glewwe et al. find that the effect of providing eyeglasses differed between counties with a much larger impact observed in County 1 than in County 2. Economic, social and environmental conditions do not differ dramatically in the two counties. The authors therefore suggest that explanations related to the implementation of the project or data collection are likely explanations for the large difference in results. For example, in County 1 all data was collected by well- trained county CDC staff who participated in trainings whereas County 2 adopted a decentralised approach that was loosely monitored and teachers were busy and had little incentive to conduct data collection carefully. In addition, aspects of project management such as monitoring whether children continued to wear eyeglasses or replacing lost or damaged glasses was superior in County 1 (Glewwe et al. 2014)

Funding Limitations

In Zambia, funding limitations resulted in the scale down of the intervention in the third year. (Grigorenko et al. 2007; IE)

Parental/ Community support

In Zambia, community sensitisation activities played a key role in gaining parents support (CAI 2007 (Project Document)

In the case of the School Health and Nutrition Study in Zambia, parents were suspicious of the fact that teachers would administer the drugs. These concerns were overcome with improvements in children's health. Community sensitisation activities (including media campaigns) were necessary activities in gaining parents support (CAI 2007).

Parents adequate knowledge of the disease may be important for compliance with the treatment (Brooker et al 2014 IE; Gleewe et al. 2014 (IE)).

Similarly, in Kenya, based on focus group discussion and in depth interviews, Brooker et al. find that adequate knowledge of Malaria and its consequences amongst school stakeholders were significant for the positive attitudes towards the programme. For example, some parents considered the intervention to be treatment of the clinical disease rather than the reduction of asymptomatic parasitemia and in a few cases parents encouraged children not to take the medication and used the drugs to treat other sick siblings. In other cases children threw away their medication because they did not perceive themselves to be ill (Brooker et al. 2015.).

In the case of the Gansu Vision Intervention Project in China, Glewwe et al. find that a lack of parental awareness of vision problems significantly affected whether children wear eyeglasses. Of the 1978 students in the treatment township, 1384 accepted, while the other 594 declined eyeglasses. The main reasons given for declining the offer were objection of the household head (187) and refusal by the child (89). Similar patterns hold for the 25 townships in the compliant sample; in particular, only 69.8% of the students offered eyeglasses accepted them (Glewwe et al. 2014).

Clarke et al. comment that studies undertaken after completion of the trial showed the IPT strategy to be well accepted by teachers, parents and schoolchildren" – (data not shown) (Clarke et al. 2014)

Contextual factors influencing SBH effectiveness

A range of contextual factors external to the programme can act to facilitate or hinder the effectiveness of school health programmes. Several of the included studies contained findings related to the community context and epidemiological profile of the targeted population within which SBH was implemented.

Baseline health and nutrition status

Baseline nutritional status and prevalence and intensity of infection may be important in mediating the effect of SBH program (Simeons et al 1995 (IE); Sylvia et al 2014 (IE) Luo et al. Wong et al. 2014 (IE)).

Student’s biological state is one of many factors affecting school achievement. The prevalence and the intensity of infection and the underlying nutritional status of schoolchildren at baseline may be critical in mediating the benefits of SBH treatment. Several studies report on subgroup analysis based on baseline nutritional status or infection intensity.

For example, in Jamaica, Simeons et al. (1995) find no improvement in growth or learning and attendance after the treatment of T. Trichiura with albendazole. However, attendance improved for stunted children who were treated compared to those receiving the placebo. The authors also report a significant treatment- by intensity- interaction for spelling outcomes. Children with heavy infection at baseline improved in spelling compared to those that received a placebo (Simeons et al. 1995).

Sylvia et al.’s study finds that neither type of anaemia reduction program (subsidy or health incentive intervention) led to significant changes in student performance on standardized exams. However, both treatment arms significantly improved exam scores of students who were initially anaemic at the start of the trial but reduced exam scores for initially healthy students. Possible interpretation of the findings is that an added focus on improving student health led to a redistribution of school resources either by a reallocation of inputs to students who were initially anaemic, or by an overall reallocation of school resources from academic inputs to anaemia reduction inputs (Sylvia et al. 2014).

In China, Luo et al. find a larger impact of the supplement treatment on Hb levels and learning if the students were anaemic at baseline (Hb levels below 120g/L) compared to the average student (Luo et al. 2012).

In contrast, Wong et al. 2014 find that the effects of the supplementation intervention on children's Hb levels and math test scores, though positive, are weaker among anaemic children than among non- anaemic children. The authors offer two possible explanations for the surprising results. 1) Anaemia may co- exist with other health conditions and learning barriers; 2) Reducing anaemia to the more anaemic children may lead to positive spillovers.

Infection intensity at baseline did not moderate the treatment effect (Watkins et al. 1996 IE; Brooker et al. 2015 IE)

On the contrary, Watkins et al. (1996) did not find that baseline infection intensity moderated the treatment effect on attendance. Brooker et al.'s subgroup analysis of the impact of IST intervention on anaemia according to Plasmodium prevalence at baseline demonstrated no different impact by prevalence category at either follow- up (Brooker et al. 2015).

Rapid re- infection between screening/ treatment rounds

Halliday et al. find no effect of school- based intermittent screening and treatment (IST) for Malaria with AL on health and education of school children. One possible explanation put forward by the authors is that in a low- to- moderate malaria transmission setting most of the children did not require treatment and those who did live in focal high transmission regions where rapid re- infection occurred between screening rounds. Rates of re- infection and acquisition of new infections between screening rounds may allow no time for haematological recovery, indicated by similar percentage of children RDT positive at each screening round (Halliday et al. 2014 IE).

Significant external event

External events including natural disasters and disease outbreaks during the intervention period may have been a barrier to the success of SBH programmes. (Glewwe et al. 2014; IE, Miguel et al. 2004; IE, Fernando et al. 2006; IE; Jukes et al. 2014).

Several studies reported on significant external events that occurred at the time of intervention, which may have had an influence on the effectiveness of the SBH trial/ programme:

Fernando et al. (2006) report that an epidemic of malaria in the placebo group and periods of severe drought during the period of the intervention led to an increase in school absenteeism during the intervention period compared to the pre- intervention period.

Miguel et al. (2004) point out that the larger participation differences between treatment and comparison schools in 1998 may also have been due to the widespread El Niño flooding in this region in early 1998, which substantially increased worm loads between early 1998 and early 1999.

In the case of the Gansu Vision Intervention Project, the SARS epidemic of 2003 resulted in a delay in the implementation of the project (Glasses were provided in 2004 instead of 2003).

Jukes et al. 2014 report on a number of external factors that might have hindered participation in the study: A massive oil spill in the Guimaras caused many of the students to experience dizziness, headache, and respiratory problems. The disruption to fishing also increased hunger levels. These problems led to a high rate of absenteeism. Sample size was further reduced by heavy rains, making many schools inaccessible, and by scheduling difficulties with staff conducting biomedical assessments.

School-feeding programs

In the following section, we report the results of the thematic synthesis of qualitative findings from our included programmes, presented using the hypothesised programme theory as an overall framework for structuring the synthesis. This first section looks at intervention and implementation features, including targeting, that may be associated with relative success and failure, followed by a review of the contextual barriers and facilitators to success.

We identified sixteen additional qualitative, mixed methods studies, process evaluations and project documents related to the eleven programmes covered by the included impact evaluations. For most programmes, there was limited additional information identified for the included programmes outside of the impact evaluations themselves and therefore much of the following synthesis is based on information provided in the impact evaluations. This is with the exception of the Mid-day Meal Scheme in India; for this programme we identified six qualitative, mixed methods studies and project documents assessing the implementation of the programme in one or more states.

Process and implementation

Uptake and adherence to the intervention

We only identified information from three of the included school feeding programs on the take up of the intervention or adherence to consumption of the provided daily meals or snacks (Diagne et al. 2014 – Senegal; Kleiman-Weiner, 2013 – China; Buttenheim et al. 2013 – Laos).

Almost complete child compliance with the snack program (Kleiman-Weiner al., 2013)

Kleiman-Weiner al. (2013) report that the compliance level with the egg a day program was at almost 100%, according to observations and interview with the teachers who dispersed the eggs. This information was recorded once per month, when the research team sent out inspectors to undertake unannounced compliance checks in the schools provided with eggs.

Schools unable to set up school feeding program in time (Diagne et al. 2014)

Diagne et al. (2014) report that some treatment schools were late in establishing a parent association to receive the food from the WFP for the school canteen program in Senegal. In other cases the program implementers were unable to make contact with the school director before implementation of the program. Authors explain the lack of positive outcomes treatment schools as potentially being a result of this lack of adherence in eight of the schools allocated to set up a school canteen for the feeding program.

Lack of take-up due to high cost of the feeding program for target villages (Buttenheim et al., 2011)

One study, Buttenheim et al. (2011), reports that there was numerous villages that turned down the offer of a school feeding program and discusses this in relation to the high costs associated with delivering the school feeding program in the northern regions of Laos, which is geographically isolated and not well served by transport infrastructure. Villages that decided not to take up the feeding program were asked on their reasons for not participating: "…the most common response was that the WFP food delivery point was too far away; another frequent response was lack of access to a road...Problems were also cited with the necessity to build the food storage warehouse and to recruit sufficient village volunteers to run the program." (Buttenheim et al., 2011:26). The authors suggest that this indicates that the villages that had the most to gain from participating in school feeding were missing out on the intervention due to lack of social capital and isolation.

Implementation fidelity and service quality delivery

We identified information relating to implementation fidelity and service quality delivery for six of the included feeding programs. There was fairly detailed information regarding implementation of the government run Mid-Day Meal scheme in India, although not for states covered by the evaluation. For all other programs, there was either no information or very limited, context specific information from which we cannot draw many generalisable conclusions.

Poor basic infrastructure a barrier to successful implementation of feeding programs (Deze and Goyal, 2003; PEO, 2010, He, 2010)

Deze and Goyal (2003) report the results of a mixed-methods survey on the Mid- Day Meal scheme in India, assessing implementation of the feeding program in the states of Karnataka, Rajasthan and Chhattisgarh. This study was conducted in 2003 - the middle of the impact evaluation study period – and found that over half of the schools they surveyed in Karnataka, Rajasthan and Chhattisgarh had no kitchen facilities, and shortages in fuel supplies and utensils. Almost half of the head teachers in the surveyed schools felt that drinking water arrangements were of inadequate standard. This often meant no drinking water source on premise. This finding is repeated in a process evaluation by the Programme Evaluation Organisation of the Government of India (2010), that found that in the states of Andhra Pradesh, Arunachal Pradesh, Jammu & Kashmir, Jharkhand, Maharashtra, Meghalaya and West Bengal, less than 75 percent of the sample schools had access to drinking water. Deze and Goyal (2003) attribute this lack of basic infrastructure to serious lack of financial resources to support the programs from the state governments in Rajasthan and Chhattisgarh. For example, in the official guidelines for the running of the program in Rajasthan, there was no mention of any expenditure on infrastructure and equipment and it seems that this matter is left to local initiative.

He (2010) reports that in the standard feeding program in Sri Lanka, WFP monitoring reports described issues with implementing the program in vulnerable parts of the country, particularly in areas that suffer from poor road transport infrastructure, which prevented delivery of food. This led to days in schools without provision of any school meal or delays in the provision.

Schools meals not provided on a daily basis as intended (Buttenheim et al., 2013)

Buttenheim et al. (2013) found large regional variation in implementation of the WFP school feeding program in Laos. In the region of Phongsaly, 97% of schools that took up the program reported providing meals in school every day during the assessed term, however in Khua, only 27 out of 47 schools (57%) reported providing a meal every day during the term. These figures suggest serious implementation issues, however these are not reported by the impact evaluation.

Poor basic infrastructure and logistical arrangement for preparing meals obstructing teacher activities in India (Deze and Goyal, 2003; Khera, 2006; PEO, 2010)

Deze and Goyal's (2003) mixed methods study of the Mid-Day Meal scheme in India found that "in a majority of schools, there is no proper kitchen facility…Food is often cooked in the open, in a makeshift shed, or in a classroom." (ibid, 2003: 4677). This is not only potentially unhygienic, but also distracts the attention of the students. They report that they did not find any cases of teachers working as chefs but many cases where the teachers spent time organising the mid-day meal. PEO's (2010) evaluation of the implementation of the MDMS found that in their sample of schools in most states, teachers spent one to two hours daily on activities related to the MDMS, thus teaching time. This finding is repeated by Khera (2006), who reports results from various field studies and government documents on the MDMS that found issues with infrastructure, such as water supply, kitchen sheds and storage facilities, and that often the preparation of food disturbed lessons. The CUTS International (2006) study of the implementation of the MDMS in Rajasthan found that 68 per cent of teachers spend more than an hour in coordinating the implementation of the feeding program. Interviews with teachers found that many complained that the MDMS had increased their workload and they felt their ability to teach effectively was hampered.

Food provided to non-enrolled students (Buttenheim et al., 2011)

An important assumption of the theory of change is that transfers of food to children are conditional on their enrolment or attendance in school. Buttenheim et al.'s (2011) evaluation of the WFP school feeding program in Laos found evidence that the snack was provided to non-enrolled children. A survey undertaken as part of the evaluation found that in the region of Khua, the per cent of schools that reported providing a snack to non-enrolled children was at 47%. At the child level, 11 per cent of non-enrolled school-aged children and preschool children in Phongsaly and 19 per cent of the same population of children in Khua were reported by a parent to have consumed a WFP school snack in the past 24 hours

Food provided to non-targeted children (He, 2010)

He's (2010) evaluation of the two school feeding programs in Sri Lanka reported that the WFP found that in 44% of schools, meals were provided to children above the grades targeted by the program. It is not clear whether this meant that the small amount of food that was delivered by the WFP was split among more children or if additional food was supplied.

School meal schemes are popular with children and families (Altman, 2013; 2003; PEO, 2010; CUTS, 2006)

We identified evidence from two of the included programs that the school feeding schemes were popular with children and families. Epstein et al.'s (2004) study on the PAE feeding program in Chile found that on an acceptability ranking of between one and seven, a sample of children rated the service between 5.1 and 6.2 in 1997, between 5.5 and 6.7 in 2000, and between 6.0 and 6.7 in 2001. Anecdotally, McEwan (2011) reports that the WFP selected the Chilean program as one of the top five school feeding programmes in the world.

PEO (2010) found of their process evaluation of the Mid-Day Meal Scheme (MDMS) found that a large proportion of children in sample schools in the states of Andhra Pradesh, Arunachal Pradesh, Himachal Pradesh, Jammu & Kashmir, Maharashtra and Tamil Nadu were of the opinion that the meals provided were of good quality. This finding is repeated by the CUTS (2006) mixed method study of the MDMS in Rajasthan, which adopted a citizen report card approach to assess satisfaction with the program, and found that most of the interviewed parents and students interviewed reported that the quality and taste of the meals was good and that they felt the nutrition and health status of their children had improved as a result.

Programs targeting school meals at the neediest schools (McEwan, 2013; Adrogue et al., 2011)

McEwan (2011) reports that the evidence suggests that the PAE feeding program in Chile is succeeding in targeting school meals at the most vulnerable schools as per its objectives. He assesses this by looking at the food calories of school feeding rations around the vulnerability index cut-off point that is used to target school meals. He finds that the food calories sharply increase around this cut off point, suggesting that the meals were targeted to needy schools as they were supposed to be. He does state that this assumes that meal rations contained the actual number of food calories recorded by the implementing agency JUNAEB, but that the incentives were not strong for private sector companies delivering the program to not deliver as they were supposed to.

Adrogue et al. (2011)'s evaluation of the school feeding programme in Argentina also attempted to evaluate the success of targeting. By constructing an index of socioeconomic status (SES) for students covered by the program and matching to schools, they report that the program is effectively located in the most disadvantaged schools, that is, in those with students with lower SES and in those with the lower school performance.

Contextual barriers and facilitators

There are a range of theorised contextual factors external to school feeding programs that can potentially act to facilitate or hinder the effectiveness (see section 4.2.1). However, we identified only very limited information regarding the influence of context in our included studies. We describe these factors below.

Existing educational and nutritional status/food security (McEwan, 2013; Altman, 2013; Omwami et al., 2011; He, 2010)

One of the key assumptions of the theory of change is that the program responds to community needs, that is, the target group has a clear need for additional food intake (inadequate energy intake/ nutritional deficiency) or that that the food ration represents an economic / nutritional benefit to the family. McEwan (2013) and Altman (2013) both report that Chile has now eliminated most extreme childhood malnutrition and now faces issues with childhood obesity in primary schools, with similar rates found to those found in the US. Primary school enrolment is also high in the country. Therefore the mechanisms for learning increases over early primary grades through two of the three channels in the school feeding theory of change, that is, improved nutrition and improving enrolment and attendance, is limited in this context. He (2010) suggests that in contexts of already high rates of primary school enrolment, provision of school meals may not be enough to bring those hard to reach children into schools.

Conversely, many of the other included studies came from contexts of poor baseline nutritional status, food insecurity and low rates of primary school participation, as reported in the impact evaluations (Tan et al. 1999 – Philippines; Kazianga et al., 2012 – Burkina Faso; Buttenheim et al., 2011 – Laos; Omwami et al., 2011 – Kenya; Jayaraman & Simroth, 2015 – India; Diagne et al., 2014 – Senegal; Cheung et al., 2013). During the study in Kenya, the area covered by the feeding program was hit by drought which caused food shortages at the household level (Omwami et al., ibid). The authors found that attendance rates declined for all students in both the treatment and comparison schools over the period that the study took place with the exception of the last study survey period when the drought was over. They suggest that it was likely that children were involved in seeking other sources of food for their households rather than attending school. This suggests that the success of feeding programs can be affected by the local food security status, presumably dependent on whether the food ration received represents an economic / nutritional benefit to the family over keeping a child at home.

Presence of school user fees (Omwami et al., 2011)

We identified evidence from one study that the presence of school fees was having a negative impact of children's participation in school during a feeding program. Omwami et al.'s (2011) evaluation of the Child Nutrition Project (CNP) in Kenya found that absenteeism and dropout from school due to the enforcement of school user fee was reported as an issue in the study, as children whose parents have not paid were excluded from school. This is in a context where fifty per cent of households live below the poverty level (ibid). In this case, the benefits of food provided at school were not able to exceed the direct financial costs of attending school in Kenya.

Capacity of local education organisations or the community to manage the feeding program (Buttenheim et al., 2011; Diagne et al. 2014; CUTS International, 2006)

We identified some limited evidence from the included programs to suggest that the existing level of community social capital and interest in participating may be a barrier to the effective participation in, and management of, school feeding programs. This is a relevant factor when there is a demand on the community to be involved in the management of the program.

As previously reported, Buttenheim et al. (2011) found that a number of villages in northern Laos decided not to participate in the WFP feeding program due to the requirement of having to build the food storage facilities, to travel to pick up WFP food allocations and to recruit sufficient village volunteers to run the program. There is also some indication that there was implementation issues in one district that was implementing the program (schools meals only provided 57% of intended days during the term). In the school feeding program in Senegal, several treatment schools were late in establishing a parent association to receive the food from the WFP (Diagne et al., 2014), which the authors use to help describe lack of positive results of the program.

In the Mid-Day Meal Scheme (MDMS) in India, the state government was expected to assign responsibility for the management of the meal program to a local level organisation or body such as a village education committee, gram panchayats or municipality, and a parent-teacher committee to monitor the management. The CUTS International (2006) survey of the program implementation in Rajasthan found that 84 per cent of the parents and 85 percent of the gram panchayats were not assisting in any part of the MDMS in the schools. Only 22 percent of the parents in surveyed villages reported that a parent monitoring committee existed in the villages. The reasons for the lack of participation in the management of the program is not clear from this survey, however it is likely that the burden of the management of the program then is shifted to other actors (for example teachers, as suggested by our analysis of process and implementation factors).

Merit Based Scholarships

Frequency

For most interventions in this category incentives were paid once annually, after examinations. The GSP paid one annual sum to schools (scholarship for tuition fees for high achieving girls) and one annual sum to families for school supplies (Kremer et al., 2008). In Gurgaon, India, the participants received one payment paid either to children or their parents, depending on which treatment arm they were in (Berry, 2013). One intervention in Nepal paid incentives to students after each semester, rather than annually like the others in this category (Sharma et al., 2011). In this programme payments were made to students following the end of semester exams in Semesters 1 and 2, and another payment made following the end of year District exams (ibid).

Targeting

Some Merit-based Scholarship programmes were targeted at underachieving students or poorly-performing school districts (Li et al., 2014; Blimpo et al., 2010).

For instance, The Fall Challenge Programme was targeted at migrant schools in Beijing where Chinese government policy prevents rural, migrant children from attending local public schools. As a result, most are enrolled in under-resourced migrant schools. Authors obtained a full list of migrant schools in Beijing and randomised from this list. Then, using students combined maths and Chinese scores, students were grouped in to quartiles according to their performance. In the Individual Incentive programme arm, targeted low-achieving students. The peer-incentive arm of the Fall Challenge Programme also targeted low-achieving students to receive the intervention but did so by pairing them with a peer from the high-achieving group, and offering incentives to both on the condition that the lower-achieving student obtained one of the largest increases in test score in his class. While, in this intervention arm the high-achieving students are not the target of the intervention as such, the authors do note that the high-achieving peers do not see a negative effect on their own results as a result of taking part in the programme (Li et al., 2010; Li et al., 2014).

Similarly in Benin, Blimpo et al. (2010) targeted low-achievers. Though in this case the intervention targeted schools where the percentage of students who passed the secondary school certification exam (BEPC) was under 65%, rather than poorly performing individuals.

Some programmes target students high-achievers and incentivised them to continue in school (Kremer et al., 2008; Barerra-Osorio & Filmer, 2012)

Meanwhile, in Kenya, the GSP was targeted at girls, a typically disadvantaged group in the region, who already demonstrated potential to perform well. Scholarships were offered to the top 15% in the year and payments were only continued if they maintained or increased their score in the following year (Kremer et al., 2008). Initial entry onto the program was decided by the girls' scores on district exams covering 5 subjects (ibid).

Some programmes were targeted at disadvantaged communities or groups such as poor or rural areas, ethnic minorities, or girls (Li et al., 2014; Blimpo et al., 2010; Kremer et al., 2008; Yi et al 2015; Liu et al., 2013).

Regional poverty levels were used for socio-economic targeting in all programmes. Poor counties were targeting for ECFA programmes and it is not clear how this was implemented (Yi et al 2015; Liu et al. 2013), in addition a student survey was used to identify the poorest students in the grade (Yi et al. 2012; Liu et al. 2013).

There is limited information on the success of targeting methods, though Sharma et al (2011) comment that one of the major drawbacks of Merit-based scholarship programmes, especially where awards are made to already high-achieving students is that they may inadvertently target the least poor in a community (Sharma et al., 2011).

<u>Implementation</u>

Implementation fidelity and service quality

Four of the included studies commented, albeit briefly, on implementation fidelity and service quality.

While some checks were put in place to reduce the incidence of cheating in exams on which payment of incentives was based, some cheating did still occur (Berhman et al., 2012)

Cheating seemed to be a worry in these programmes. For example, Li et al carried out robustness checks to rule out any cheating which would have affected the results. Similarly, a cheating analysis was conducted for the ALI. Berhman et al (2012) discussed the study's analysis of the occurrence of cheating during mathematics tests. To minimise cheating, external examiners administered the tests (on which bonus payments were based). However, to determine to what extent cheating had occurred an independent statistical analysis was done. It showed higher levels of cheating in treatment arms with direct incentives (i.e. individual student incentives) as opposed to treatment arms where only teachers were incentivised (not included in our analysis). The authors present their treatment effects taking this cheating analysis into account (ibid).

Cancellation of district exams caused disruption to implementation of the GSP, leaving the implementing NGO to run the exams in one district instead (Kremer et al., 2008)

In the GSP, where initial exam performance determined whether girls would be allowed into the scholarship programme, one of the school districts cancelled their exams leaving the implementing NGO to run them instead (Kremer et al., 2008).

There is a small, but fairly unlikely, risk that teachers may inflate childrens' grades so that they receive the incentives (Sharma et al., 2011)

On a related note, the Sharma (2011) comments that checks were made to ensure that teachers did not "strike a deal" with students to boost their score and share a proportion of the child's incentives. However, the author concluded that this is unlikely given that the payments were made on the basis of aggregate scores and as such would require a risky degree of collusion between students and other subject teachers (ibid).

Other cash-transfer and scholarship interventions present in treatment schools may account for lack of impact on drop-out rates (Berhman et al., 2012)

Almost 40 percent of the students in the ALI programme received a Cash-Transfer as part of the Oportunidades. In addition several students from low-income families were receiving scholarship payments for successful progression from one grade to another. The authors say that this explains the lack of effect of their ALI programme on drop-out (Behrman et al, 2012).

Disbursement of one scholarship programme was based on adherence to certain conditionalities. However, these were only moderately enforced. (Barrera-Osorio & Filmer, 2012).

In Cambodia, the scholarship programme required recipients to adhere to certain conditionalities such as regular attendance and passing examinations. However, these were not always enforced and, in many cases, students still received their scholarship even though they had had some absences. Payments were stopped where absence was excessive (Barrera-Osoria & Filmer, 2012).

The same study reported delays and other problems in programme start-up which meant that students were unable to apply to be eligible for the scholarship until they had already begun the school year. This also meant that scholarships were paid in one lump-sum as opposed to two tranches (ibid).

Take-up/Adherence

Parental commitment to reward their child for good work may improve attendance but does not seem to affect attainment (Berry, 2013)

Berry (2013) tested whether parents intend to reward their children for achievement in education but have difficulty committing to do so. In the parent treatment of his study (where parents received the incentive rather than the child), participants were asked whther they would like to commit to giving their child the monetary reward rather than keeping it for the household. Half were asked ex ante, before the programme began, and half were ask ex post, after their child had already achieved their goal in the reading test. More parents who chose ex-post than ex-ante rewarded their child with a toy. Of those who did give the money directly to their child, this was mostly for the purpose of buying food for themselves. Moreover, the outcomes of the study should some, albeit weak, evidence for an effect of parental commitment ex ante on attendance but not on attainment (Berry, 2013).

<u>Outcomes along the causal chain: process outcomes</u>

We now report descriptive findings for 'intermediate' outcomes, that is, those outcomes which may explain whether or not changes in learning outcomes are seen. Four of our included studies report on the intermediate outcomes so we have limited data to present here. As per our suggested programme theory Merit-based scholarships are thought to work as a fee subsidy for academic performance which should facilitate increased involvement through attendance, and child effort in school and in doing homework. The programme theory also suggests that teacher effort and attendance may increase as a result of parental involvement or pressure to ensure their child does well and, as such, will receive the reward.

Student effort and motivation tends to increase (Berhman et al., 2012; Blimpo et al., 2010; Kremer et al., 2008), and doesn't necessarily detract from other key activities (Behrman et al., 2012; Kremer et al., 2008).

Student effort seems to have been measured using indicators such as study time or reported enjoyment of studying. For instance, the Cambodia programme found an increase in the amount of time students spent studying outside of school (Barrera-Osorio & Filmer, 2012). Similarly, in Mexico, Behrman et al (2012) reports of the Aligning Learning Incentives (ALI) programme that students spent more time studying mathematics and were significantly less likely to text or watch TV while doing homework. They were also more likely to help their classmates with work. Additionally, the authors note that while the incentive was related to mathematics scores, the results showed no difference in reported study time for other subjects which indicates that the programme did not shift student effort away from other skills (ibid).

Blimpo et al., 2010 reported that students reported putting more study time and effort into subjects requiring Higher Order Thinking Skills (HOTS) such as mathematics, despite incentives being awarded on the basis of performance across all subjects. The results showed a statistically significant improvement in HOTS subjects which, the authors suggest, may means that incentives can be tied to performance across all subjects but still achieve greater results on subjects such as maths and science.

Further, the data suggests that a team (or group) tournament-style programme saw a higher increase in reported study time than a team target-style programme. It was suggested that while the target group reported higher levels of individual effort, there was more cooperation in the tournament group (Blimpo et al., 2010).

Kremer et al preferred measures of Student Motivation for assessing the impact on intermediate outcomes in the GSP. Using a survey of eight questions which asked students whether they preferred certain school activities over non-school activities such as doing chores or playing sport. Results showed that overall students seem to prefer school activity in 72% of the questions presented. Though there was no evidence to suggest either a positive or negative effect of the incentives on reported extra tutoring or homework completion. Further, the indicated preference for school activity does not seem to have affected time spent on chores at home. An important observation given girls' expected role in the household in these districts of Kenya (Friedman et al., 2001; Kremer et al., 2008).

The cost of education can act as a barrier for low income households. The effectiveness of methods used to target such low income households may impact on the programmes educational outcomes (Yi et al., 2015; Liu et al., 2013)

In countries where students are likely get a paid job outside of school there is a high opportunity cost for education. This opportunity cost is reported to be a significant factor for both of the ECFA programmes (grade 7 and 9). In the 7th grade programme Yi et al. (2012 pPN) measure student opinion on this opportunity cost, a key factor in the potential success of the programme. Students on average anticipate having to spend 3.5 times the value of their household assets a year to attend an academic high school, they also anticipated earning an average of 1,248 yuan a year outside of school. This opportunity cost accumulates to 44,928 yuan a year in the mind of the average student. In addition to this, 90% of students were unaware of policies providing financial assistance to vocational high school and 89% were unaware of such policies for academic high schools (ibid).

The ECFA appears to have affected children's intentions to progress to high school (Yi et al, 2015; Liu et al., 2013)

The proportion of students in the ECFA programme that planned to attend academic high school increased by 15 percentage points in the treatment group, the number of students that planned to attend vocational high school increased by 2% and the number of undecided students dropped by 17 percentage points. High opportunity costs of high school education, the ease of getting a low skilled job outside of school and the competitive nature of the education system in the context of ECFA means that student plans for attending high school is a significant step in the causal chain. Students must choose to stay in education and must then study to pass the high school entrance exam. Liu et al. (2013) report on ECFA's (grade 9) impact of self-esteem, as increasing self-esteem may impact a students plans for further education. The authors used the Rosenberg Self-esteem scales (SES) I and II to measure the programme impacts on self-esteem and found no statistically significant impacts.

Teacher Effort/Attendance

Merit-based scholarship programmes may have a positive effect on teacher attendance (Kremer et al., 2008)

There are mixed results on the ALI's effect on Teacher Effort (Berhman et al., 2012)

Results suggest that for the GSP teacher attendance increased significantly across the two included districts which could, according to the authors, correspond to an increase in test scores for students (Kremer et al., 2008).

The authors seem to equate increased attendance with increased effort here and suggest that possible mechanisms behind the positive effect could include social prestige and ego rents. Presumably, this is because the GSP involves presenting winners with a certificate at a public ceremony as well as cash incentives. Thus winners and their teachers will be known within the community. Authors also suppose that teachers could be motivated by the expectation or possibility of receiving gifts from winners' parent (Kremer et al., 2008).

Effects on Teacher Effort are reported for the ALI. Authors state that a higher proportion of teachers in the treatment group reported preparing students for their exams and helping students outside of class time. Tenth grade teachers were also less likely to use only multiple-choice question in their testing; this is notable because long-answer question take much more time and effort to grade. However, the authors did not find any increase in the amount of time teachers spent preparing for class meaning that overall results on teacher effort for this study are mixed (Berhman et al., 2012).

With reference to the finding that teachers reported helping students preparing from exams and aiding students outside class time, the authors reported a positive affect for the treatment arm involving only student incentives. While we do not discuss the other treatments here, it is useful to point out that the student and teacher incentive treatment arm produced a much larger effect. In their discussion of the mechanisms behind this effect, the authors conclude that while motivation to help students win incentives may increase teacher effort, the effect is significantly augmented when they too are rewarded for student performance (Berhman et al., 2012).

There do not seem to be any peer-effects of merit-based scholarships (Barrera-Osorio & Filmer, 2012).

The theory of change for the Cambodia programme assumed that there may be some positive spillover effects for those who did not receive the scholarship. However, the results showed no such affects for either attendance or learning outcomes (ibid).

<u>Contextual Barriers and Facilitators</u>

There is limited information on contextual barriers and facilitators within the included studies though we can at least comment briefly on a few points. The program theory for this intervention type mentions four major assumptions; incentive based learning, teaching and school environment, non-financial barriers to education, and family of which we can comment on four related contextual points.

Students' existing academic ability may be a predictor of how large an effect the incentive will have (Blimpo et al., 2010).

The mechanisms by which incentives may have an effect and suggests that students' existing ability is a factor in whether or not these interventions improve learning outcomes. For instance, a very low-ability student may have to exert a high amount of effort to achieve a grade which qualifies for incentive-payment. Depending on whether the student thinks this is achievable, the time and effort costs may be much higher than the gains. Meanwhile, high-ability students may not have to exert much effort at all to gain the grades required to receive an award and so do not represent much of an effect. The results from this evaluation show that while there was an overall positive effect on learning, most of this can be attributed to the students with a medium level of ability (ibid).

An increase in student effort does not appear to have had a negative effect on girls' expected household duties (Kremer et al., 2008).

In the case of the GSP programme in Kenya, Kremer et al. report in relation to student effort that while there was an increase in time spent studying outside school that this did not appear to have a negative effect on doing household chores. Since, the participants are in a community where women are traditionally confined to the household and are expected to undertake household chores this observation is promising. It suggests that the scholarship programme can still have an effect despite existing pressures on girls out-of-school time (Friedmen et al., 2011; Kremer et al 2008).

Social prestige and social pressure may be important factors in teacher effort and parental support (Kremer et al., 2008)

As mentioned above, social prestige may be an explanatory factor behind increases in teacher attendance and effort (Kremer et al., 2008). The public certificate-giving ceremony in the GSP means that the community are aware of the programme and as such teachers may be motivated by the prestige of one, or more, of their students winning a scholarship (ibid).

In addition, Social pressure may play an important part in this particular intervention. The authors note that the materials grant although not conditional was, by in large, spent on school supplies. The authors attributed this to social pressure from the community, again due to knowledge of the intervention and the public certificate-giving ceremony (Kremer et al., 2008).

High poverty rates and the costs of education are a barrier for the three of the Chinese scholarship programmes (Chen et al, 2013; Liu et al, 2013).

All three programmes implemented in China are implemented in rural areas. The primary barrier to education for students in rural areas is poverty. Fewer than half of the junior high graduates in poor rural areas attend senior high school (Liu et al, 2013). Tuition fees are some of the highest in the developing world and are a primary reason for low senior high enrolment (Chen 2013). In the Ningshan region the cost of attending senior high school surpasses annual household income (Liu et al, 2013).

Providing information

In the following section, we report the results of the synthesis of qualitative findings from included Providing Information to Children programs. Given that there were only three included studies in this intervention category, and no additional qualitative or process documents, findings on this are scarce. The first section looks at intervention and implementation features that may be associated with relative success and failure, followed by a brief summary of the contextual factors that potentially affect the success of these interventions.

Intervention Design

Information on the returns to education

In all three programs included in this intervention category, participants were given information on the potential returns to education, in terms of wage earnings, after leaving school. For two studies this involved providing children with national data, for instance average wages by education level.

In a different approach, the Abre la Caja described the education experiences and current earnings of 13 so-called role models rather than using national data or average earnings (Dinkelman & Martinez, 2011).

Additional information categories

Two of the studies provided additional relevant information; one program in xx provided participants with additional information about available academic scholarships and student loans for further study while, in China, the program included information delivered as part a series of counselling sessions. These counselling sessions not only included the information described above but also helped students identify career interests and highlighted the importance of acquiring skills (Loyalka et al., 2013).

Delivery methods

Delivery methods differed in terms of how the information was presented and who presented them. In most cases, information about returns to education was delivered verbally. However, Jensen (no date) also gave participants a written statement of the statistics presented to take home with them while in the Chinese counselling intervention, children were given workbooks as part of the 4-session information and career planning course. Meanwhile, in Chile, the Abre la Caja program delivered information verbally but via a DVD which, depending on the treatment arm, was either watched at school or at home (with parents) (Dinkleman & Martinez, 2011).

All three of our studies varied in terms of who delivered the information. For example, in China teachers were trained by professional counsellors to deliver either the one-off information session or the series of four career counselling sessions by a professional counsellor. Teachers then delivered the intervention to their own students in class (Loyalka et al., 2013). Jensen's experiment (Jensen n.d.) used "trained enumerators" to deliver the intervention to children. We can only assume that these were external to the school though not much detail is given.

Targeting

School Grade

Two studies targeted their interventions by school grade, keeping critical stages in schooling in mind. For instance, Abre la Caja, was targeted at Grade 4 students pointing out that, in Chile, children must choose in Grade 8 whether or not they will carry on into secondary school and also choose a specific study track. As such, the authors chose to implement the intervention before this critical stage (Dinkleman & Martinez, 2011).

Similarly, Grade 7 students were targeted for the Chinese information and counselling program. Authors cited literature which suggests a positive result from targeting wage information at primary and secondary school students. The authors also found that Grade 7 and the subsequent grades are critical points where students often drop out.

Socioeconomic Status and Place of Residence

Loyalka et al (2013) targeted the intervention at poor rural areas, largely because it was thought that people in rural areas are more likely to have inaccurate information on returns to schooling.

Conversely, the Dominican Republic intervention was targeted only at non-rural areas due to a lack of information on average earnings and the difficulty in estimating such earnings for agricultural households. This was particularly important for the program given that a household survey was first administered to estimate the returns to schooling in the area and then to provide that information to participants in the interventions (Jensen, n.d.).

Gender

Further, the programme in the Dominican Republic was targeted at boys due to low female labour force participation and the fact that most of the girls in the authors' focus groups did not expect to ever work (Jensen, n.d.)

Process and implementation

Students provided with a DVD to take home did not necessarily watch it (Dinkleman & Martinez, 2011).

In the Abre la Caja program, authors noted that one school did not participate in the follow-up and further that, for those who were given the DVD to watch at home with their family, only 60% reported actually watching it. Given that 89% of families in the sample had a working DVD player in the home the figure cannot be explained by lack of access. The majority of students in the programme had the opportunity to watch the DVD (Dinkleman & Martinez, 2011).

Impacts on intermediate outcomes

Perceived returns to schooling.

It is thought that changes in students' perceived returns to schooling can affect schooling behaviour, such as progression and completion. As discussed elsewhere, the program did find an increased effect on years of schooling which the authors say was due to the increase in perceived returns to schooling observed at follow-up 4 to 6 months after the interventions. Though they cannot rule out other factors which may have contributed to this effect, on average, the participants revised their expected returns to schooling by RD$364 (Dominican Pesos). The authors find that 54% of the treatment group revised expectations of their own potential returns to schooling between baseline and endline, compared to only 27% in the control group (Jensen, n.d.).

Effect on students' plans to progress to high-school

Loyalka et al. (2013) conducted a multivariate analysis for both the information intervention and the counselling intervention and found no impact on either learning outcomes or on students' plans to go to high school compared to the control groups.

Further, the counselling intervention may have had a small negative effect on dropout. While primary outcomes such as this are discussed elsewhere, it is worth also mentioning in this section as the authors comment that students' may have concluded from the counselling sessions that achieving the grades to enter high-school or college were too difficult which could in turn affect plans to progress through school as well as dropout (Loyalka et al., 2013).

Student and Parent knowledge of financial aid eligibility

Dinkleman & Martinez (2011) report that exposure to the programme (ie. Watching the DVD) increases the proportion of students who report that they will finance future education with a government loan (4.6%) and decreases the proportion who do not know how to finance education (4.2%). Knowledge of loan opportunities increased by 50%.

The authors do not compare Family and Student treatments for this outcome but do so for primary outcomes. They mention that one explanatory factor to take into consideration for the primary outcomes when comparing the two groups is that in the family intervention, the students' had a choice of whether they would watch the DVD or not (as opposed to the student intervention where participants were made to watch it in school together with their peers). Only 60% of students in the Family intervention actually reported watching the DVD and the authors comments that the likelihood is that most of these would already be higher-ability students. This should be taken into consideration when

Contextual factors influencing effectiveness

Jensen

"In the Dominican Republic, while 80 to 90 percent of youths today complete (compulsory) primary schooling, only about 25 to 30 percent complete secondary school. Yet the mean earnings of workers who complete secondary school is over 40 percent greater than workers who only complete primary school" (IE:1), "further, more education workers are likely to receive greater non-wage benefits" (IE:1, footnote 2)

Loyalka

The IE Provides brief contextual information about China's economy and education. I will now provide a summary of this information. The tagged text contains the full passage. ; 'China's economy gradually shifts from one based on low-wage industries towards one based on higher-valued ser- vices, the demand for skilled labor will outpace that for unskilled labor. To meet the new requirements individuals will need to be equipped with higher levels of schooling. However,recent studies have shown that only about two-thirds of the students from poor, rural areas in China enter high school (Yi et al., 2012a)...most of these children enter the labor market and take unskilled jobs." (Loyalka et al., 2013:1013)

Dinkleman

IE: In Chile enrollment in and attainment of tertiary education is strongly correlated with family socioeconomic status in Chile. In 2006, only 12.7% of 18-25 year olds from the lowest income decile were enrolled in tertiary education, compared to 53.3% of the top income decile. (p.2)

Household Level

Programmes to reduce or eliminate user fees

Implementation

Targeting

Four of our included programmes are targeted on school type. Three of these are targeted at primary schools, one of which also targeted junior or middle secondary and one which targeted junior or middle secondary and secondary. Two of these also had a gender element. The Free Uniforms Program was targeted at primary schools. UPE specifically targeted primary education with a gender element, up to four children per household could receive fee elimination and of those, at least two had to be girls. The BOS programme, Indonesia was targeted at all public and private primary and junior secondary high schools. Guidelines strongly suggest the provision of BOS to poor students (Soharyo, 2006). The Girls' scholarship program was specifically targeted at girls in middle secondary and secondary school.

Four programmes are specifically targeted at economic status. The No- Fee Policy is targeted at the economic status by region as is the Three Tuition Policy reform. Gratuidad fee reduction initiative and the School Incentives Project are targeted at socio-economic status at the household level.

Kenya Child Sponsorship Programme is targeted at children within schools that have lost both parents, however this is not always the case[152].

Targeting Implementation Strategy

Programmes that are targeted by economic or socio-economic status have more complex targeting strategies.

The No Fee Policy has a three tier selection process used by national governments. Firstly a 'poverty score' is allocated to each province based on the 2001 census, this collates information on income, employment, education, health and living condition. Provincial governments can adjust their 'poverty score' by changing the weighting given to each of the five previously mentioned areas. Within these quintiles the population should roughly "equally poor" and contain around 20% of the student population. Secondly, the national government ranks these quintiles with the top two given no-fee status, this was extended to the third quintile after 2007. In the third stage, provincial governments decide which schools to allocate funding on the basis of community wealth (Garlick, 2013).

The Gratuidad fee reduction initiative is targeted at households on the basis of census data. The census has 74 questions encapsulated within 6 question areas that capture information on location of the household, its infrastructure and services available to it, it also collects demographic information of household members including family structure and members, their health and education, earnings and employment status. This information is used to categorise households into six levels, households that are ranked into levels one and two are included in the programme (Barrera-Osorio, 2007).

The Three Tuition Policy targeted was first targeted at schools chosen by the government of China and predetermined quotas of students were assigned to schools. Teachers chose students to participate based on their understanding of the student's socio-economic status (Hou & Zhou 2014).

Interviews were used for the School Incentives Project to identify students for participation based on family income. Interviews were conducted with families employed in carpet factories (Edmond and Shrestha 2014).

In the three programmes that are targeted at primary schools there little clear information on how targeting was implemented. It is not clear how UPE implemented family demographic targeting. The Free Uniforms Program is described as a research led expanding an existing programme in rural areas to urban areas but no further information is available. The BOS program in Indonesia was implemented to all primary and junior schools and was targeted on the number of students in each school (Soharyo, 2006). The targeting of poorer students was devolved to schools, particularly the principle (Soharyo, 2006). As The Girls' scholarship program was targeted at all girls in middle and high schools no further targeting mechanisms were required.

Randomised lottery is used to target non-orphan children in programme schools in the Kenya Child Sponsorship Programme with a replacement policy if the selected child is absent on the first day of school, a letter is sent home to encourage attendance on the first day of term (Evans, 2012).

[152] The sample of children in the programme evaluation used in this review exclude children that have lost both parents. Children included in the sample are randomly assigned within existing programme schools.

Implementation Fidelity

There are several process factors and issues of fidelity (if the program was implemented as it was designed), these issues occur across studies or are at times unique to the program.

Gaming behaviour from elected officials was considered an issue as they aimed to gain as much funding as possible from the central government for the programme. There is evidence of gaming behaviour in households that hoped to gain a poverty that would allow their inclusion in the programme (Garlick 2013 and Castaneda 2005).

Concerns about gaming strategies or pressure to be included within the targeting process are apparent in The No Fee Policy and in the Gratuidad fee reduction initiative. In The No Fee policy gaming is not considered an issue for the first two levels of targeting: the use of census data to create poverty score and the ranking of these by the national government. Provincial officials however report that there was lobbying in the third level of targeting, when provincial governments decide which schools to allocate funding. It is not clear to what extent this lobbying was effective (Garlick, 2013). There was however a high rate of compliance, with quintile 2 schools being 94 percentage points more likely to eliminate fees than quintile 3. Garlick (2013) argues that this makes the poverty score an effective tool for targeting poor schools. Potential gaming strategies were also of concern in the in Gratuidad fee reduction initiative. Elected mayors have to power to decide the areas in which the census used to target the intervention is implemented (Castaneda, 2005). Further to this, mayors are responsible for social programming but funding for the Gratuidad fee reduction initiative comes from central government, there is thus a reason to amplify the number of those eligible to increase the amount provided from central governments (Castaneda, 2005). In addition to gaming behaviour by elected officials there is potential for gaming behaviour in households that take the survey. In some cases people have asked for a home visit or provided the address for a house that they have rented for the purpose of achieving a lower score. There is also evidence of households attempting to influence census administrators in order gain lower scores (Castaneda, 2005). These issues are highlighted for the census implemented in 1996-2000, a new census was used after 2000 and the study used to evaluate this programme considers the time frame between 2005 and 2006. The extent to which this gaming behaviour is prevalent with the new census is therefore unclear. The authors argue that there is low likelihood of gaming with the new census as the scoring formula was kept secret. In addition there would be evidence of "stacking"; a greater number of participants that just met inclusion instead of just missing it which there is no evidence of (Barrera-Osorio, 2007).

Mis-targeting of schools and poor students within them. The targeting process may have resulted in the mis-targeting of schools and a failure to include schools from poor communities. This procedure also excluded salafiyah or religious schools. Targeting of poor students by schools was not considered culturally appropriate by implementing schools. In lieu of this, general subsidy was provided to all students (Soharyo 2006, Garlick, 2013, and Giese et al. 2009).

As previously mentioned in the theory of change, user fees are understood to act as barrier to education. This barrier is considered to be particularly prevalent in low income households or areas. Therefore the effectiveness of methods used to target such demographic groups may impact on the programmes educational outcomes. Targeting issues at the school in the BOS program included the use of outdated census data, ranking errors, failure to consider inequalities within provinces and failure to consider small rural schools) may have resulted in mis-targeting of schools and the exclusion of schools within poorer communities (Soharyo,

2006) . In addition, the programme was intended for all schools but the targeting mechanism ruled out salafiyah, religious schools, due to the changeable nature of their student population (Soharyo: 2006). A study found that the targeting of poorer not fully implemented as it was not considered culturally appropriate by schools. It was more likely for schools to provide a general subsidy for all students regardless of their economic background (Suharyo, 2006). The targeting mechanism in the No Fee Policy effectively targeted the intended treatment group with 98% of quintile 1 and 2 schools being included and only 6% of quintiles 3, 4 and 5 being included (Garlick, 2013). There are however concerns about the effectiveness of the quintile system for targeting schools that serve poorer neighbourhoods. The census data used to calculate the poverty score in each province is outdated (Mestry, 2013). There may be ranking errors between schools resulting in unequal inclusion of schools that provide for the same communities (Giese et al., 2009). Schools serve poorer communities but are close to better off areas may be ranked with a higher quintile and be excluded from the program. The provincial measure of quintiles may hide the inequalities within provinces and between schools.

Change in accountability from parents to those implementing fee reduction or removal caused three primary issues; poor management in the provision of goods or funds, misuse of funds and insufficient funds to schools. Poor management resulted in delays to the opening of schools and students not receiving the uniforms. Misuse of funds included payment of teacher Honoria and non-educational goods. Payments that did not reflect varying educational costs between primary and secondary schools and between rural and urban contexts resulted in insufficient funds (Giese et al. 2009, Mestry, 2013, Hidalgo et al, 2013, Juan 2007, and Marishane 2013).

The removal or reduction of fees represents a change of accountability that once lay with parents, to those implementing fee reduction or removal. Three key issues stem from this problem; poor management in the transfer of goods or funds, misuse of funds and insufficient funds in schools. A lack of information, communication and funding delays from provincial Education Departments to schools in the No Fee Policy resulted in these schools being unable to operate for a time (Giese et al., 2009: 12 and Mestry, 2013). Implementation by provincial governments was an issue for the Free Uniforms Program; in two of five provinces tailors were contacted too late to make the uniforms and as a result only 2 of 52 schools in these provinces received uniforms (Hidalgo et al., 2013). Independent monitors to the BOS programme found that with 97% of funds were used in accordance with policy, but 30% were used for teacher honoraria. In the No Fee Policy funds were used to purchase materials that were not for educational purposes in order to avoid having to account for unused funds (Marishane, 2013). Funds in this program are also reported to be insufficient to cover the costs of education and did not reflect variations in the costs of education between rural and urban schools, and between primary and high school (Juan, 2007, Marishane, 2013 and Giese et al., 2009).

A failure to monitor and inspect schools resulted in poor policy enforcement and the continuation of fee charges in programme schools (Setoaba, 2011).

A lack of policy enforcement is an issue in the No Fee Policy, schools continued to charge fees and this is reported to result from a failure monitor and inspect schools (Setoaba, 2011).

Process Factors

Ineffective targeting mechanisms for programmes to reach targeted low-income households

As previously mentioned in the theory of change, user fees are understood to act as barrier to education. This barrier is considered to be particularly prevalent in low income households or areas. Therefore the effectiveness of methods used to target such demographic groups may impact on the programmes educational outcomes.

Reducing user-fees can reduce the quality of education, this is caused by less parental involvement in schools, increased student-classroom and student-resource ratios and increased delinquent behaviours (Garlick 2013, Nkosi, 2011, and Ministry of Education and Sports of Uganda, 2005).

As suggested in the theory of change reducing and eliminating user fees can result in a reduction in quality of education. This reduction in quality may be caused by insufficient or a reduction in funds to the school and increased enrolment in no-fee schools. There is evidence of decreased quality in both no-fee and fee paying schools in South Africa (The No Fee Policy). The programme also resulted in a statistically significant increase in average class size (Garlick 2013). A study of three schools that were declared non-fee found that there was less parental participation in the school affairs, an increase in enrolment and an increase in delinquent behaviours of students including vandalism, theft and misuse of school property (Nkosi, 2011). Several factors hampered school governing bodies in this program including a lack of clarity and understanding regarding the use of funds, restrictions on suppliers that can be used for goods and services and an inability to use funds within school development plans (Marishane, 2013)

The UPE programme had issues with the quality education available; increased enrolment increased the resource-to-pupil ratio with the textbook-to-pupil ratio at 1:3, the classroom-to-pupil ratio to 1:55 and the desk-to-pupil ratio to 1:3 in 2003 (Ministry of Education and Sports of Uganda, 2005)

Misleading public messaging about the programmes meant that some parents did not pay for any associated costs of education, or reduced their fee-paying behaviour even if their children were in non-programme schools. This resulted in both a shortfall of funds for programme schools and reductions in funds for non-programme schools (Soharyo, 2006, Giese et al. 2009, and Ekaju, 2011).

Communication with participants, families and communities caused issues for both the BOS program, the No Fee Policy, South Africa and UPE. Misleading television messages to the public on the BOS programme resulted in many thinking that it would provide education free of charge rather than reduce educational costs. As a result the announcements were pulled and revised. An appraisal of the programme found that some of the print material, brochures, could not be located at any level from governing bodies to schools and communities and it could not be ascertained as to how the print material was used (Soharyo, 2006). The No Free Policy also had unclear public messages which caused misunderstandings for both schools and parents (Giese et al., 2009). This issue is apparent in the UPE programme; government information following it's launch lacked clarity and caused the misunderstanding that all costs of education would be covered, though non-tuition costs were still required. In order to maintain schools and keep them open, principles report having to be "creative" to keep receiving payments from parents (Ekaju, 2011).

The No Free Policy's misleading public messages that "education is free" negatively impacted school income for fee paying schools. There was both a decline in fee paying behaviour in parents and the per learner allocation from the government to fee-paying schools increased by only 10% between 2005-2007 which was not in line with inflation (Giese et al., 2009).

Limited funds and time caused poor training of implementers and schools, this hampered programme implementation. Misunderstanding of the programme exacerbated unclear public messaging and inadequately prepared schools for implementing the programme (Soharyo, 2006, Giese et al. 2009, Nkosi, 2011, Marishane, 2013, and Setoaba 2011).

Poor training was an issue from implementation in both the BOS programme and No Fee programme. Training for the BOS programme was hampered by a limited funds and time, poor materials on the programme, high numbers of participants and the late stage that the training was carried out in. Misunderstanding programme design and implementation by key implementers further exacerbated issues of unclear and misleading public messages mentioned earlier (Soharyo, 2006). Training with schools in the No Fee policy was hampered by a lack of basic knowledge in financial terminology and management systems that had been presumed by implementers (Giese et al., 2009). While it is noted that capacity building activities did not adequately prepare schools for implementing the programme (Setoaba, 2011). The issues experienced with training resulted in a lack of capacity in schools for financial management and administration (Marishane 2013 and Setoaba, 2011, Nkosi, 2011).

Additional Programmes in Treatment and control groups

Additional programmes or programme elements in treatment and control groups are reported in Child Sponsorship Programme, The Girls' scholarship program, The No- Fee Policy and the Three Tuition Reform Policy. The Child Sponsorship Programme Schools provided schools with two nurses that visited the school several times a year, agricultural trainers that ran clubs to grow food on the school grounds and a grant for classroom construction and learning materials. Students in these schools made up both the control and treatment group (Evans 2012). Sponsored children (the treatment group) also occasionally received letters of support and small gifts from sponsors (Evans 2012).

Authors note that in addition to The Girls' scholarship program many schools and students will have also received small scale interventions from charities and NGOs. He argues that they would not have been large enough to impact the outcomes reported (Gajigo, 2012).

In South Africa, the School-fee exemption policy (introduced in 1996) was already in place when the No Fee policy was implemented (Garlick, 2013). This policy required learners to be exempt from paying fees if they were below a certain household income level designated by means testing. Some students had exemption without means testing these include: fostered children, those in “kinship care”, those living in child headed households and those in receipt of social grants. Fee paying schools do not receive funds to off-set these fee-exemptions, fee paying students therefore cover the additional costs (Branson et al., 2013).

Changes to education policy in China render it likely that treatment groups for the Three Tuition Reform Policy also experienced other changes to education provision. These include changes to education funding; school budgets were included in county budgets rather then town budgets. They also include an effort to increase elementary education from five to six years (Hau and Bo, 2014).

Uptake of the intervention

There is take up information on two of the interventions. Both the No-fee Policy and UPE have high take up. Branson et al (2013) report significant increases in students that pay no fees between years that the programme was expanded; in 2007, 28% of students report paying no fees increasing to 64% in 2011. These figures however includes students that are included in the programme, those with scholarships and those with fail to pay fees. Take up is not directly noted on the UPE programme, however increasing pupil-resource ratio as a result of increased enrolment is an issue drawn on, see other contextual factors for more information.

Context

Education Policy

Municipalities in Colombia have devolved power to regulate fees that public schools can charge. In relation to the Gratuidad fee reduction initiative, the government of the city of Bogotá issue a resolution the controls the maximum fee and the additional items that that schools can charge (Barrera-Osorio, 2007).

The government of Kenya introduced a policy in January 2013 to cover the school fees and additional material costs of attending school. This did not cover include school uniforms which became the most significant cost of attending school. Wearing uniforms is a requirement for attending school, students that failed to wear uniforms were sent away from school. Politicians have voiced their disapproval of this rule but there is no clear policy on it (Evans, 2012).

In 1996 the South African Schools Act was introduced making education mandatory between ages 7 and 15 or for completion of such grades. This Act also increased the powers of school governing bodies that decide school policies (Mestry, 2013). The School-Fee Exemption Policy also introduced in 1996 saw the partial and full removal of school fees on income based means testing and other targeted children (Branson et al. 2013; p.3). The No Fee School Policy introduced in 2007 resulted in the removal of fees for selected schools, see program description for further details (Borkum, 2012).

In 1995 the Primary Education and Teacher Development Project was introduced in Uganda, under which the Government hired new teachers. Education Sector Investment Plan was introduced in 1998 in response to increased enrolment and decreasing quality of education caused by the UPE policy (Grogan, 2009). The Thematic Curriculum introduced in 2007 made English the language of instruction for both urban and rural areas, where previously was just in urban areas. Before the implementation of UPE Capitation Grants were given to schools (Grogan, 2009), these were often subject to elite capture with only 13% of non-wage expenses reaching schools, poor areas were therefore much more reliant on tuition fees (Reinikka and Svensson, 2004). Devolvement of Capitation Grant allocation and publicising grants increased the amount received in schools from 12-80% in 1999 (Grogan, 2009). Education reform in Uganda is often in partnership with the aid community, with political commitment to these policies (Essama-Nssah, 2011). Such reforms are have resulted in local and educational stakeholder feeling marginalised from the process (Higgins and Rwanyange 2005 in Essama-Nssah 2011).

Prior to the start of the BOS programme the Government of Indonesia introduced nine years of compulsory education in 2003.

Education policies implemented prior to the start of The Three Tuition Policy in China include a compulsory attendance law (1986) encouraging enterprises, administration bodies and local communities to establish schools. In 1985 schools were permitted to charge households for the attendance of children (Hau and Bo, 2014).

In Nepal where the School Incentives Project Evaluation was carried out education is free between grade one and eight (Edmond and Shrestha 2104).

External Event

Two external events occurred in the implementation of the No Fee policy, there was firstly a teacher strike in 2008, this may have resulted in an increase in fees to reflect increases in teacher salaries (Branson et al., 2013). Secondly, the policy, through its implementation experienced wide-spread public criticism in local media and debates (Setoaba, 2011).

Other Contextual barriers and facilitators

Other contextual barriers and facilitators are reported for the No-fee Policy, UPE, Three Tuition Programme and the Tuition waiver programme.

The South African schooling system reflects inequalities regarding race, class and gender. Perceived low returns for education are considered a key causal factor in the poor performance of students compared to other African countries.

For the No-fee Policy in South Africa there are existing inequalities on race, class and gender. The poor quality of black, public schools is seen in limited qualified teachers and resources (Mestry, 2013). In addition low returns to education are considered a key causal factor in poor performance of students compared to other African countries (Borkum, 2012).

Strong political will for the implementation of the Universal Primary Education (UPE) Reform, resulted in its implementation 6 months after the original government pledge. This resulted in insufficient physical and human resources to support the programme.

There was strong political will for UPE which resulted in its implementation 6 months after the original government pledge, however this resulted in implementation without sufficient physical and human resources to support the programme (Ministry of Education and Sports, 1999).

Pre-existing gender inequalities and the cost of education is significant barrier to education in China, though many rural families recognise the long-term value of education.

Similarly to the No-fee Policy the Three Tuition Programme contended with pre-existing gender inequalities, with rural families traditionally valuing education of sons over daughters (Hau and Bo, 2014). Further to this, the total cost as part of household expenditure increased three fold between 1991 and 2004 impacting education for children in poor rural communities (Hannum, Behrman, et al.,2008 in Hau and Bo, 2014). There is, however, a high return rate for education in China and many families in rural areas understand education to be “the only way out” of rural poverty for rural children (Hannum, Kong, et al., 2008 in Hau and Bo, 2014).

Outcomes along the causal chain as measured by counterfactual analysis

Two included studies measured outcomes along the causal chain as part of their counterfactual analysis. Garlick (2013) reports on class size for the No Fee Policy. These outcomes provide insights to the mechanisms that will affect program impact. The causal chain for reducing fees suggests that reducing fees without sufficient funding to accommodate increased enrolment increases class size and reduces the quality of education. The average class size increased from 39.3 to 39.8 students in Garlick’s (2013) analysis. The authors report that this difference was statistically significant. Edmond and Shrestha (2014) reports on a child labour outcome, time spent weaving, a reduction in child labour may impact the child’s available time and increase participation in education. Edmond and Shrestha, (2014) found that the stipend was associated with a 59% reduction in weaving.

School Level

Pedagogy

In the following section, we report the results of the synthesis of qualitative findings from our included pedagogy programmes, presented using the hypothesised programme theory as an overall framework for structuring the synthesis. This first section looks at intervention and implementation features that may be associated with relative success and failure, followed by a summary of the contextual barriers and facilitators to success.

We identified nine additional documents, process evaluations and qualitative evaluations related to the twenty programmes included in our analysis. For many programmes, there were no additional documents and therefore the main source of information for this section were the impact evaluations themselves. In general, the interventions included here were more likely to be fairly small-scale experiments, rather than nationwide policy. As a result, there is generally only a limited amount of process and qualitative information available.

The table at the end of this section provides a detailed overview of the included interventions’ design.

Process and implementation

Materials and equipment were distributed as expected and were well-maintained (Abeberese, Kumler and Linden, 2011)

Abeberese, Kumler and Linden (2011) reported that materials and equipment for the Sa Aklat Sisikat (SAS) programme in the Philippines were distributed where and when expected and that they were well-maintained.

Materials and equipment were delayed and/ or not of the desired standard (Lucas et al., 2014; KAPE, 2004; Spratt and Ralaingita, 2013)

Lucas et al., (2014 – Uganda and Kenya) note that most schools did not receive the classroom mini-libraries until April 2010, six months after the beginning of the intervention. The Read-Learn-Lead (RLL) programme in Mali was particularly affected by problems with the delivery of materials, with a substantial proportion of schools not receiving the intended materials. 27% of 2012 RLL treatment-school Grade 1 teachers and 43% of 2012 RLL treatment-school Grade 2 teachers reported that they had not received teachers’ guides (Spratt and Ralaingita, 2013). There were also some problems with the School Readiness programme (SRP) in Cambodia, where there were not enough curriculum documents or teaching aids to go around – something that later proved to be a substantial problem when it came to implementation (KAPE, 2004).

Teachers attended training sessions (Brooker et al., 2013; Abeberese, Kumler and Linden, 2011, He, Linden and MacLeod, 2007).

Several studies reported that attendance rates for training had been high. Attendance for the teacher training sessions under the Kenya based Health and Literacy Intervention (HALI) project: literacy programme were above 95 percent on average (Brooker et al. 2013). In the Philippines, the vast majority of teachers under the Sa Aklat Sisikat (SAS) Reading programme took and passed training, with successful completion of training based on attendance, classwork, and a final assignment. 130 teachers completed the training course, with 115 obtaining a certificate that provided professional points in the Civil Service Career system. 16 teachers failed to attend any of the sessions. These teachers were offered a recovery training session; 9 of them attended and 7 were absent (Abeberese, Kumler and Linden, 2011). Similarly, the Pratham PicTalk intervention (Year 2) in India had good teacher attendance at training overall and the programme offered materials and second-hand instruction from a supervised trained teacher in the same area (He, Linden and MacLeod, 2007).

Issues with the quality of the teacher training may have prevented teachers from delivering new content appropriately (KAPE, 2004; Mouton, 1995; Spratt, King and Bulat, 2013; RTI International, 2015).

A large proportion of treatment schools in the Read-Learn-Lead (RLL) programme in Mali did not receive the appropriate training. 11% of 2012 RLL school principals, 51% of RLL treatment school Grade 1 teachers, and 44% of 2012 RLL treatment-school Grade 2 teachers reported that they had not received RLL training (Spratt, King and Bulat, 2013). The RLL programme was designed to build upon national training in the relevant curriculum approach. However, the anticipated training was not provided and therefore the programme was weakened (Spratt, King and Bulat, 2013). RTI International (2015) note that there were some instances of trainers turning up late for scheduled teacher-training for the Primary Math and Reading (PRIMR) Rural Expansion Programme in Kenya.

A study of the School Readiness programme (SRP) in Cambodia expressed doubt about the effectiveness of training, finding that there was too much content to be covered in too little time, mixed quality of trainers (KAPE, 2004). Mouton (1995) found that training in the South African intervention's 'Suggestopedic' method did not provide all teachers with the necessary knowledge and skills to implement it correctly. For example, exercises were not always carried out effectively, games and songs, small-group work and student participation were under-utilised. The study concludes that this was a likely limiting factor on the intervention's effectiveness.

Problems with implementation meant that some interventions did not begin on time (Piper & Mugenda, 2013; Spratt, King and Bulat, 2013)

Some programmes experienced difficulties ensuring that they were delivered as planned. Piper & Mugenda (2013) report that the Primary Math and Reading (PRIMR) Initiative in Kenya was delayed from the outset, which meant that materials reached schools later than planned and early data collection occurred when the intervention had only just begun. Spratt, King and Bulat (2013) report that the implementing agency was unable to fully deliver on its ambitious implementation plan, with the result that the Read-Learn-Lead (RLL) programme in Mali was extended for a third unanticipated year.

Teachers prevented students from taking books home (Abeberese, Kumler and Linden, 2011)

A studies of the Sa Aklat Sisikat (SAS) programme in the Philippines reported that one key problem the programme faced was that sometimes teachers prevented students from taking books home, primarily because they thought they would be safer if kept in schools (Abeberese, Kumler and Linden, 2011)

Teachers welcomed new programmes (Abeberese, Kumler and Linden, 2011; KAPE, 2004)

Some studies found that teachers responded positively to the interventions. Teachers participating in the Sa Aklat Sisikat (SAS) Reading programme in the Philippines were reported to be open to the innovations introduced into classrooms (Abeberese, Kumler and Linden, 2011). A study of the School Readiness programme (SRP) in Cambodia found that focus-group discussions with teachers generally indicated both acceptance and satisfaction with the instructional methodologies introduced by the intervention (KAPE, 2004).

Teachers opposed some of the changes that a programme wanted to make (Piper and Mugenda, 2013).

One case is illustrative of the kind of opposition from teachers that programmes can face. The Primary Math and Reading (PRIMR) Initiative in Kenya chose to use grade one materials for both grades one and two, as class two materials were deemed too difficult a starting point for the grade two students. However, many teachers did not accept this and this resistance was echoed by some parents and this reception had a clear negative impact on take-up (Piper & Mugenda, 2013).

Teachers valued text messages with instructional tips and motivation (Jukes and Dubeck 2015)

The Kenyan Health and Literacy Intervention (HALI) project used text messages to teachers to provide instructional tips and motivation for lesson plans. Teachers reported they valued the support the messages provided and a study found that the text message mechanism was "successful in creating a sense of community, making teachers feel valued and listened to, providing a mechanism for feedback and intervention improvement" (Jukes and Dubeck, 2015: 31).

There were insufficient resources provided for implementation staff supporting the teachers (Piper & Mugenda, 2013, KAPE, 2004; RTI, 2015)

In another case, the tutors participating in the Kenyan Primary Math and Reading (PRIMR) programme found it difficult to support their assigned teachers while also handling the other duties for which they were responsible (Piper & Mugenda, 2013). A further study looking at the Primary Math and Reading (PRIMR) Rural Expansion Programme, also in Kenya, reached the same conclusion with Teachers' Advisory Centre tutors sometimes skipping school visits due to the additional burden of their administrative duties (RTI, 2015).

A study of the School Readiness programme (SRP) in Cambodia found that teachers did not feel adequately supported by school directors. Many directors were unaware of the resource requirements of the programme and therefore did not set aside sufficient allocations of the available funds (KAPE, 2004).

Teachers lacked the necessary experience and capacity to implement new instructional approaches (Mouton, 1995; Piper and Korda, 2011; Abeberese, Kumler and Linden, 2011).

Some studies reported that the skills and capacities of participating teachers limited the effectiveness of programmes. Mouton (1995) reports that many teachers participating in the English and Operacy programme (EOP) in South Africa had a very basic level of English, lacked the skills to properly engage their students and were unable to properly apply the intervention methodology.

In one case, the intervention in question was well-implemented in many respects, but the results clearly indicate that students in treatment groups performed worse than those in the comparison groups. The authors hypothesise that teachers were unable to successfully adopt the student-centred methodology as they required more time to implement it correctly or that their experience just did not prepare them for the new approach (Abeberese, Kumler and Linden, 2011).

The implementers of the Early Grade Reading Assessment (EGRA) Plus programme in Liberia observed that one key skill that teachers lacked was lesson-planning and responded by providing tightly scripted daily lesson plans, something that was seen as a key step in making the intervention more effective (Piper and Korda, 2011).

Teachers were concerned about the time and additional work required by programmes (Jukes and Dubeck, 2015; San Antonio et al., 2011; RTI, 2011).

The time taken up over and above that for the normal curriculum was something reported by various studies. One study found that teachers expressed concerns about the time they were having to spend preparing for lessons and that this may have reduced the time available for other subjects (Jukes and Dubeck, 2015). Some teachers that had to self-study the Module-based professional development for teachers (MBPDT) in the Philippines also reported that it was difficult to make the necessary additional time to do the studying (San Antonio et al., 2011). Some teachers participating in the Early Grade Reading Assessment (EGRA) Plus programme in Liberia complained that it imposed too much extra work on them, with report cards in particular viewed as taking up too much time and as a result were often unused (RTI, 2011).

High teacher turnover was sometimes a problem (Piper & Mugenda, 2014; Spratt, King and Bulat, 2013**)**

Two studies reported that teacher turnover was a problem, though programmes did adapt to deal with the problem. The Primary Math and Reading (PRIMR) Initiative in Kenya experienced high teacher turnover in some schools and therefore needed to train new teachers to ensure that programme activities could continue (Piper & Mugenda, 2014). The Early Grade Reading Assessment (EGRA) Plus programme in Liberia also reported that turnover and transfer was a problem, though the implementers worked with local education authorities to minimize out-transfers of RLL-trained teachers over the period (Spratt, King and Bulat, 2013).

Actual lesson time was often less than that scheduled by programmes (Piper and Mugenda, 2014; KAPE 2004).

Actual lesson time was often less than that scheduled by programmes. A study of the Early Grade Reading Assessment (EGRA) Plus programme in Liberia found that classroom ‘time on task’ was inadequate for several reasons. Firstly, teacher attendance was irregular with teachers arriving late or leaving early. Some rural schools were only open two hours a day. Students also missed or dropped out of school to work instead. As a result, the authors estimate that reading instruction probably took place three or four times a week instead of the planned five times (Spratt, King and Bulat, 2013). Another study reported that the large number of holidays during the School Readiness programme (SRP) in Cambodia disrupted the program’s momentum, as did late registration of students (KAPE, 2004). In Kenya, the scheduling of the Primary Math and Reading (PRIMR) Initiative clashed with other extracurricular activities with the result that there was reduced lesson time for pupils (Piper and Mugenda, 2014).

The national government may have reduced the supply of school inputs to schools involved in a programme (Spratt, King and Bulat, 2013)

One study reported some evidence that schools participating in the Read-Learn-Lead (RLL) programme in Mali received fewer textbooks and other inputs from the Ministry of Education, with the result that RLL resources that were intended to be 'additive' actually replaced inputs provided to other schools (Spratt, King and Bulat, 2013)

Teachers implemented programme activities and used instructional aids in class (Brooker 2013; Lucas et al., 2014; Abeberese, Kumler and Linden, 2011; Spratt, King and Bulat, 2013; Jukes and Dubeck, 2015)

As a measure of intervention compliance, (Brooker 2013: 46) assessed whether teachers had used instructional aids provided for the Health and Literacy Intervention (HALI) project: Literacy programme in Kenya. They found that 90 percent of intervention teachers were observed to have used the teaching aids in class. The Health and Literacy Intervention (HALI) project: Literacy programme in Kenya provided weekly text messages containing brief instructional tips and motivation for lesson plans. The response rate to messages containing questions was over 80 percent and the messages were deemed to have helped create a sense of community amongst teachers and have provided a mechanism for feedback and intervention improvement (Jukes and Dubeck, 2015).

Lucas et al. (2014) undertake analysis of both the Ugandan and Kenyan Reading to Learn Interventions (RtL) and examined whether materials were used and whether programme activities were undertaken as intended. They find that the likelihood of observing other reading materials, student-made materials, and wall charts and visual aids increased across both countries. They also observed increases in both countries in recommended textbooks. In addition, the authors carried out 'implementation analysis' of student, teacher, and head teacher activities. They find that, although programme instructions were uniform, the methodology was applied differently by schools and teachers. Scores across the two countries are similar with 25 percent of schools awarded 'high' scores of 7 to 11, 50 percent with 'medium' scores of 5 or 6, and 25 percent with 'low' scores of 4 of less.[153]

Spratt, King and Bulat (2013) found that compared to comparison teachers, Read-Learn-Lead (RLL) teachers engaged students more during reading instruction and reported that classrooms were found to be more participatory and friendly, though teachers were observed to pause less frequency to allow students to catch up. Another study of the Health and Literacy Intervention (HALI) project: Literacy programme in Kenya found that lessons were popular with students and increased student engagement and participation (Jukes and Dubeck, 2015).

[153] Implementation analysis scores were determined by the following 11 criteria: *Teachers* (1) Teachers are effectively using the five RTL steps in the correct sequence; (2) Teaching is done procedurally and with logical understanding and is not mechanical; (3) Teachers are innovative and committed to implementing the approach; (4) Teachers are motivated to support learners in numeracy and literacy outside teaching time; *Classroom Learning Environments*: (5) Appropriate learning materials are used; (6) The classroom library is utilized; (7) Children are reading age appropriate texts; (8) There is enhanced peer support among learners; *School Leadership*: (9) Head teachers provide technical support; (10) School and parents have a supportive relationship; (11) Functional school development plans prioritize lower grades.
Schools with scores 7-11 were considered 'high', scores of 5-6 'medium', and 0-4 were 'low' implementers relative to ideal implementation.

A study of the Sa Aklat Sisikat (SAS) Reading programme in the Philippines concluded that reading frequency in schools was higher as a result of the programme (Abeberese, Kumler and Linden, 2011).

Over a half of the teachers did not implement the new methods properly (Mouton, 1995)

Another study examining the English and Operacy programme (EOP) in South Africa concluded that only 13 of 36 teachers were adjudged to have implemented the new method properly (Mouton, 1995).

Teachers felt the programme improved their professional competence (San Antonio et al., 2011; Spratt, King and Bulat, 2013).

23 of 25 teachers participating in the Module-based professional development for teachers (MBPDT) programme in the Philippines reported that they felt the programme had increased their professional competence by enhancing their knowledge and skills in teaching mathematics (San Antonio et al., 2011). A study on the Early Grade Reading Assessment (EGRA) Plus programme in Liberia found little evidence that participants listened to radio shows that were a small component of the intervention (Spratt, King and Bulat, 2013).

The programme improved student and teacher attitudes (Kerwin and Thornton, 2015).

Students participating in the Northern Uganda Literacy Project (NULP) were significantly more positive towards school and believed more in their own ability. Teachers were also marginally more positive towards teaching and had higher standards for student performance (Kerwin and Thornton, 2015)

The programme succeeded in switching teaching from learning-by-rote English to more participatory native-language classes (Kerwin and Thornton, 2015).

The evaluation of the Northern Uganda Literacy Project (NULP) found that more time was spent on reading and writing and less time on speaking and listening activities that probably reflected rote memorization through call-and-response. Much more time was also spent speaking the native language Leblango instead of English (Kerwin and Thornton, 2015).

Contextual

Programme implementation was disrupted by political events and flooding (Abeberese, Kumler and Linden, 2011; Piper and Mugenda, 2014; Spratt, King and Bulat, 2013)

A number of interventions were disrupted by events outside their control. In the Philippines, flooding caused schools taking part in the Sa Aklat Sisikat (SAS) Reading programme to close for a period, although it did not prevent them from ultimately completing the intervention (Abeberese, Kumler and Linden, 2011). Kenya experienced extensive political violence in Kenya at the beginning of 2008 and Piper and Mugenda (2014) note that the Primary Math and Reading (PRIMR) Initiative was interrupted in 2013 by the national elections and the political activities that preceded it, with schools closed for a week and teachers heavily involved in political activities. Spratt, King and Bulat (2013) note that an attempted coup in Northern Mali by Islamist radical groups occurred in March 2012. The authors altered their analytical approach to account for the fact that they could not collect endline outcomes in these areas.

Context of high growth in enrolment put pressure on schools' resources and capacity (Lucas et al., 2014 [Uganda and Kenya]; Piper & Mugenda, 2014; Spratt, King and Bulat, 2013; Nonoyama-Tarumi and Bredenberg, 2009)

Four studies referred to the high recent growth in educational enrolment and graduation as a result of educational reform and in particular policies promoting universal access to primary education (Lucas et al., 2014 [Uganda and Kenya]; Piper & Mugenda, 2014; Spratt, King and Bulat, 2013; Nonoyama-Tarumi and Bredenberg, 2009) note that schooling capacity in Cambodia has not increased in line with the higher enrolment rates and the result has been overcrowded classrooms, textbook shortages, an overtaxed teaching force, and declining levels of instructional quality. Spratt, King and Bulat (2013) also refer to the fact that, though primary school access has increased, the overall quality of education has not.

Intervention design stipulated that content should be taught in students' mother tongue to promote learning (Lucas et al., 2014 [Kenya]; Kerwin and Thornton, 2015; RTI, 2015)

Particularly for reading, language and literacy interventions, the language of instruction was often something that typically featured in the intervention design and theory, with some interventions intended to be taught in the languages spoken in students' homes in order to promote learning (Lucas et al., 2014 [Kenya]; Kerwin and Thornton, 2015; RTI, 2015).

Students were often taught and tested in a language other than their mother tongue (Lucas et al., 2014 [Kenya]; Jukes and Dubeck, 2015, Brooker et al., 2013; Spratt, King and Bulat, 2013).

Language of instruction is an important factor in many of the countries where included programmes were implemented. In Kenya, national policy stipulates that for grades one to three, classes should be taught in the mother tongue language spoken in students' homes, whereas from grade four English should be used (Brooker et al., 2013). However, in many cases, English is used in primary grades because of the lack of instructional material for Kenyan mother tongue languages (Jukes and Dubeck, 2015). The fact that English is used in later exit exams also means it is often prioritised in earlier grades (Brooker et al., 2013). Lucas et al. (2014) note that some primary school teachers actively punished the use of local languages, while others employed a mixture of English, Swahili and other languages. Spratt, King and Bulat (2013) comment on the fact that reforms in Mali have prioritised early grade instruction in curriculum schools, but that French-language instruction is still the norm. This means that multiple curricular approaches coexist and this might be hypothesised to affect student learning (Spratt, King and Bulat, 2013). Many of these interventions emphasise the importance of teaching in the languages used by students at home. However, this was not always put into practice – for example, in Kenya, despite the Reading to Learn (RtL) intervention being designed to be taught in Swahili, teachers often used English (Lucas, et al., 2014). The Primary Math and Reading (PRIMR) Rural Expansion Programme found that careful consultation with experts and users of native languages helped address challenges associated with teaching in native languages not necessarily spoken by all children (RTI, 2015).

Intervention implementation disrupted by teacher strikes (Piper and Korda, 2011; Piper and Mugenda, 2014; Spratt, King, and Bulat, 2013; RTI, 2011)

Three studies reported instances in which teacher strikes disrupted programmes. Piper and Mugenda (2014) report that there were extended strikes twice over the course of a Kenyan intervention. Spratt, King, and Bulat (2013) comment that teacher strikes contributed to a shortened school year to variable degrees in different regions, reducing schools' ability to implement the full course of intervention lessons. Piper and Korda (2011) note that during the Liberian Early Grade Reading Assessment (EGRA) Plus programme, the academic year was delayed due to a volunteer-teacher strike caused by the government's dismissal of all unqualified volunteer teachers. This delayed the intervention and, as nearly 30 percent of programme teachers were volunteers, also resulted in the need to train replacement teachers, increased the number of students in the average class, forced some schools to teach multiple grades together, and damaged the morale of both teachers and their trainers. There was also a strong disincentive for remaining volunteer teachers to work for free for EGRA Plus while there were other programmes offering significant financial incentives (RTI, 2011).

Low levels of teacher knowledge and experience. Non-traditional instructional practices prevalent (Dixon, Schagen and Seedhouse, 2011; Berlinski and Russo, 2013; He, Linden and MacLeod, 2007 [Year 2]; Mouton, 1995; He, Linden and MacLeod, 2009).

Several studies noted the limited nature of traditional pedagogical approaches. One study reported that the intervention was undertaken in a context in which lecture-style teaching with little student interaction was the norm (Berlinski and Russo, 2013). Dixon, Schagen and Seedhouse (2011) note that the schools covered by the Synthetic Phonics Intervention were implemented in schools were teaching was typically by rote. Others noted the relatively low knowledge levels of the teachers being targeted (He, Linden and MacLeod, 2007 [Year 2]; Mouton, 1995) or the low quality of the instruction practices themselves (He, Linden and MacLeod, 2009).

Many schools suffer from resource constraints, with limited availability of material and large classes (Abeberese, Kumler and Linden, 2011; Mouton, 1995; Piper and Mugenda, 2013; KAPE, 2004; Kerwin and Thornton, 2015).

One study on the Sa Aklat Sisikat (SAS) Reading programme in the Philippines noted that schools lacked age-appropriate books, with the few that had libraries typically containing only books donated from developed countries with subjects and writing styles that were not age-appropriate. The authors note that it was unsurprising that teachers used them infrequently (Abeberese, Kumler and Linden, 2011). Piper and Mugenda (2013:15) note that the nonformal schools that made up a large proportion of participant schools are characterised "by tuition, poor infrastructure, frequent transfer of pupils from one school to another, lack of adequately trained staff, non-standardized managerial operations and high teacher turnover."

A study of the School Readiness programme (SRP) in Cambodia found that the resource intensive nature of the intervention did not function in what were typically resource-poor environments (KAPE, 2004). Mouton (1995) noted that the correct implementation of the pedagogical principles of the English and Operacy programme (EOP) in South Africa was hindered by class sizes that made small-group work impossible and a lack of electricity to operate tape recorders.

Programmes implemented in low resource settings (Dixon, Schagen and Seedhouse, 2011; Tan, Lane, and Lassibille, 1999; Lucas et al., 2014 [Kenya]; Piper and Mugenda, 2014; Mouton, 1995; Nonoyama-Tarumi and Bredenberg, 2009; Kerwin and Thornton, 2015)

Several studies noted the poverty of the districts in which interventions were implemented (Lucas et al., 2014 [Kenya]; Piper and Mugenda, 2014; Mouton, 1995; Dixon, Schagen and Seedhouse, 2011; Tan, Lane, and Lassibille, 1999). Others note that student performance

(Lucas et al., 2014 [Kenya and Uganda]; Tan, Lane, and Lassibille, 1999) and school completion (Tan, Lane, and Lassibille, 1999) are particularly low in intervention areas. Other studies also noted the low levels of pre-school provision and participation (Nonoyama-Tarumi and Bredenberg, 2009), and of parental participation and the value placed on schooling (Lucas et al., 2014 [Kenya and Uganda]). Mouton (1995) notes the legacy of 'Bantu' education in intervention schools, whereby black students received poorer educational conditions than their white peers.

Table A i4.1 a: Pedagogy intervention design

Study ID	Topic	Curriculum	Teacher resources	Student resources	Teacher training	Teacher monitoring, mentoring or feedback	Other
Abeberese, Kumler, Linden, 2011	Language arts (reading)	Reading was incorporated into existing curriculum	N/A	Age appropriate storybooks in both English and Filipino	2 days	Implementing agency monitored schools to ensure program fidelity and support teachers' usage of the new books.	N/A
Berlinski and Russo, 2013	Maths	New maths curriculum, more active learning	Lesson plans and teachers' manual	Treatment group 1 received a student workbook. Student group 2 received a student workbook and an interactive whiteboard.	10 hrs/week over 4 weeks	N/A	N/A

Study ID	Topic	Curriculum	Teacher resources	Student resources	Teacher training	Teacher monitoring, mentoring or feedback	Other
Dixon, Schagen and Seedhouse, 2011	Language arts (literacy, reading and spelling)	Synthetic phonics approach	Lesson plan and lesson pattern	Worksheets, flashcards, storybooks	Peripatetic teacher – unclear what training they received	N/A	*Peripatetic teacher
He, Linden and MacLeod, 2007 (Year 1)	Language arts (English language)	English language curriculum	Manual for use with flashcards	Flashcards, PicTalk machine allowing students to point on pictures and hear words pronounced	Peripatetic teacher – unclear what training they received	Peripatetic teachers attended weekly training sessions for feedback and to prepare materials for coming week	*Peripatetic teacher Maintenance team for PicTalk technology
He, Linden and MacLeod, 2007 (Year 2)	Language arts (English literacy)	English language curriculum	Manual for use with flashcards	Flashcards, PicTalk machine allowing students to point on pictures and hear words pronounced	Training was undertaken but no information on length	Pratham monitors circulated amongst schools on a regular basis to assist teachers	Maintenance team for PicTalk technology
He, Linden, and Macleod, 2009	Language arts (reading)	Reading and comprehension curriculum	N/A	Flash cards, story books	Training was undertaken but no information on length	Supervisors met instructors twice a week, and zonal heads meet supervisors once every 10 days to ensure	Community libraries set up

Study ID	Topic	Curriculum	Teacher resources	Student resources	Teacher training	Teacher monitoring, mentoring or feedback	Other
						consistency in training and implementation	
Irwing et al., 2008	Maths (mental arithmetic)	Unclear – no information	N/A	Unclear whether abacuses were provided	Not clear if teachers were peripatetic or trained in-school	N/A	Not clear if teachers were peripatetic or trained in-school
Jukes and Dubeck, 2015	Language arts (literacy)	Literacy curriculum	Teaching manual, lesson plans, instructional materials	Materials for classroom use	3 day initial workshop, with 1-day and 2-day follow-ups at the end of year 1 and year 2 respectively	Teachers received text message tips with monetary incentives for responses. Teachers filled out summary sheets documenting successful lessons and suggestions for improvement	N/A
Kerwin and Thornton, 2015	Language arts (literacy)	Mother-tongue curriculum	Teacher guides with literacy class plans.	Writing slates, reading and writing primers	5-day residential workshop, followed by three further intensive trainings and six Saturday	N/A	Parent-teacher meetings were held and parents were trained in how to interpret their child's literacy report card and

Study ID	Topic	Curriculum	Teacher resources	Student resources	Teacher training	Teacher monitoring, mentoring or feedback	Other
					in-service training workshops over school year		use a simple home reading assessment tool. Clocks were also provided to schools.
Leme, 2010	General - structured teaching methods	Curricula for structured methods	Lesson plans, homework materials, access to a website	Textbooks	For training, generally the providers meet every 2 or 6 months with all the teachers	Supervision of teachers. Monitoring visits. E-support through online forum.	Private institutions delivered intervention
Lucas et al., 2014 (uganda)	Language arts (reading)	Reading-to-learn curriculum	N/A	Reading and numeracy materials, books, lockable storage facilities	12 days of in-service training	Monthly mentoring meetings and teachers invited to quarterly meeting to share ideas with peers and receive refresher training	Implementing agency met with school management committees to promote intervention. Mini-libraries wereestablished in some communities

Study ID	Topic	Curriculum	Teacher resources	Student resources	Teacher training	Teacher monitoring, mentoring or feedback	Other
Lucas, et al., 2014 (Kenya)	Language arts (reading)	Reading-to-learn curriculum	N/A	Reading and numeracy materials, books, lockable storage facilities	12 days of in-service training	Monthly mentoring meetings and teachers invited to quarterly meeting to share ideas with peers and receive refresher training	Implementing agency met with school management committees to promote intervention. Mini-libraries were established in some communities
Mouton, 1995	Language arts (language)	English language curriculum	Unclear though it seems that some resources were provided	Unclear though it seems that some resources were provided	3 weeks	Four week-long monitoring visits. Some teachers received follow-up visits to support and motivate them	N/A
Nonoyama-Tarumi, & Bredenberg, 2009	Language arts (Khmer language)	'Bridging' curriculum developed	Documentation containing new games and activities	Physical upgrading of classrooms, stationery, materials for making teaching	14 days	Regular visits to support teachers in programme implementation	Formalised student assessment for monitoring purposes

Study ID	Topic	Curriculum	Teacher resources	Student resources	Teacher training	Teacher monitoring, mentoring or feedback	Other
				aids, classroom decorations			
Pallante, 2013	Language arts (reading and literacy)	Reading and literacy curriculum	Teachers were given graphic organisers to help them prepare lessons.	Materials (not described) were also provided to promote best practice.	Four workshops over the course of the school year	Teachers' classes were monitored and they received coaching reports and sustained support	Teachers were trained to use student test results to identify children at risk, monitor progress and tailor teaching accordingly.
Piper and Korda, 2011	Language arts (reading)	Reading curriculum	Lesson plans, teacher manual, pocket charts	N/A	1 week training at start of intervention, shorter refresher training in second school semester	One support visit by coaches every month	School report cards, radio outreach promoting reading, regional reading competition. Capacity building for Ministry of Education staff.
Piper and Mugenda, 2014	Language arts and maths	New curriculum for English,	Teacher lesson plans. Reading	Reading and maths materials integrated with	Teachers & head-	Coaches provided regular feedback, professional	Continuous assessment of students. Open-

Study ID	Topic	Curriculum	Teacher resources	Student resources	Teacher training	Teacher monitoring, mentoring or feedback	Other
	(Kiswahili reading, English reading, maths)	Kiswahili and maths	and maths textbooks.	lesson plans. 1 textbook per pupil	teachers: 10 days	development and in-class support. Monthly teacher reflection meetings	to-the-public reading and maths contests. Encouragement and reminders for teachers via email. Programme materials reviewed and updated.
RTI International, 2015	Language arts and maths (Kiswahili & English reading, maths)	New curriculum for English, Kiswahili and maths	Reading and maths textbooks.	Reading and maths materials integrated with lesson plans. 1 textbook per pupil	Teachers & head-teachers: 10 days	Coaches provided regular feedback, professional development and in-class support. Monthly teacher reflection meetings	Reading and maths exhibitions with parents and community invited to visit schools
San Antonio et al., 2011	Maths	N/A	Printed instructional modules	N/A	Ongoing training through modular learning, one per week for 5 weeks	Ongoing training through modular learning, one per week for 5 weeks. School-heads and supervisors followed up with teachers.	N/A

Study ID	Topic	Curriculum	Teacher resources	Student resources	Teacher training	Teacher monitoring, mentoring or feedback	Other
Spratt, King and Bulat, 2013	Language arts (literacy – local language)	Literacy curriculum	Lesson plans, teachers' guides	Materials for activities, student books	Teachers and school head trained. 3 to 6 days in each of 3 programme years	Support and monitoring visits incorporating classroom observation, student spot assessment, and advisory discussion with school educators	N/A
Tan, Lane and Lassibille, 1999	General - instruction tailored to learners' needs	N/A	Teacher resources	No explicit information	1 week	N/A	N/A

Notes:

N/A = not applicable/ no relevant intervention componenet
Curriculum: plan introducing new content or methods of instruction
Teacher resources: for example, lesson plans, activity guides
Student resources: for example, wallcharts, textbooks, technology
Teacher training format of teacher training
Teacher monitoring, mentoring or feedback
*Peripatetic teacher: trained teacher brought in to school to deliver intervention
Other: other intervention component of note

Computer Assisted Learning (CAL)

Process and implementation

Information on the fidelity of programme implementation and service delivery was reported for seven of the twelve programmes. Some discussed fidelity to implementation in a general sense, others reported issues with participant targeting, or with the hardware or software provided. We summarise the findings from this analysis below.

***Overall implementation compliance reported as high (e.g. teachers followed instructions, took training, materials were delivered etc.) (*Berlinski and Busso, 2013; Lai et al. n.d.; Lai et al. 2012; Lai et al., 2013; Mo et al., n.d; Mo et al., 2014; Humpage, 2013) (Impact Evaluations)**

General information on fidelity and service delivery was available for seven of the twelve programmes. Based on both notes from the volunteers who monitored Lai et al.'s (n.d.) language-focused CAL programme in China, the authors report that almost all schools closely followed programme instructions and that there were few violations. The same study refers to records kept by the teacher-supervisors, who delivered the intervention, to suggest that almost no CAL sessions were cancelled due to unforeseen events. In those cases where this did happen, schools arranged make-up sessions so that the programme was implemented according to schedule.

In Lai et al.'s (n.d.), Lai et al. (2012, 2013) and Mo et al. (n.d., 2014) CAL programmes in China teachers were only supposed to schedule the CAL intervention and help the students with hardware and software operations. They were not supposed to teach the students. The authors' class-room observations confirm that teachers did as such. They report that ''the sessions, in fact, were so intense that the attention of the students was fully on the computer and there was little communications among the groups or between any of the groups and the teacher-supervisor'' (Lai et al., n.d.:10).

Berlinski and Busso (2013) report that the CAL study in Costa Rica was implemented with high fidelity, that the materials and equipment were delivered as expected and that these remained functional throughout the experiment (p.4). They moreover report that the requirement that all schools teach 7th grade geometry during the second term was fulfilled by all but three classes, which were all in the control group.

Humpage (2013) reports that there was perfect compliance to treatment for the OLPC Pedagogical Support Pilot Programme in Peru. The authors notes: *" In this experiment, there was perfect compliance to treatment, in that all schools that were assigned to the treatment group received the Pedagogical Support Pilot Program (PSPP) training, while none of the schools assigned to the control group received this training. The training was school-wide, including all teachers, and took place over ten school days"* (p.67). From this it is not entirely clear, however, if all teachers received the necessary training. Humpage (2013) does report however that, when teachers in both treatment and comparison schools were asked about the training, there was a significant different between the two groups in terms of accompanied training they had received: *"In treatment schools, 43.3% of teachers report having participated in training with an accompanier, but to 11.8% of teachers in control schools also did (this difference is significant at the 1% level). Restricting the sample to teachers that were working in the same school in 2010, the difference increases from 31.5 to 42.8 percentage points. Teachers in treatment schools also report having spent significantly more days in training with an accompanier, and are significantly more likely to report having had "hands-on follow-up training."* (p. 68). These data are as expected: teachers in both treatment and control groups received the original OLPC training, but those in the treatment group received the additional PSPP training. Humpage (2013) further notes that the Ministry of Education reports that the trainers ‘‘implemented all components of the training as planned in all schools." (Humpage, 2013:61)

Some programmes faced technological issues, including insufficient, damaged and dysfunctional equipment, lack of internet access and software not being compatible with hardware limited **(Examples of Best Practices in Peru, n.d (Project document); Villanueva-Mansilla, 2012 (Qualitative study); David and Quispe, 2013; Sharma, 2014; Imbrogno, 2014 (Impact Evaluations).**

Hardware and software issues are reported for three of the programmes. Additional documentation (Examples of Best Practices in Peru, n.d.) reports that the Huascaran Programme faced technological issues; sometimes computers were not working, sometimes there was no internet connection, and in some cases the hardware was not appropriate to the programme software.

Cristia (2012) reports that the OLPC programme in Peru had as one of their selection criteria that school have internet access. Because schools typically did not have internet access, this requirement was dropped. The requirement that schools have electricity was maintained, however. Cristia et al. (2012) report that electricity access (among their sample?) was close to universal (p.12).

David and Quispe (2013) report that the OLPC programme in Peru faced problems with its hardware and software due to administrative and logistical problems according to the directors of the educational institutions. Data collected by the authors confirms this as it suggests that, of the 888 XO laptops provided to the treatment group, 181 were damaged to the extent that they were not usable. In addition, 38.85% of the 345 chargers provided were also unusable (p. 197).

One school only received 45 laptops for its 215 students. Additional documentation also reports that the amount of equipment delivered often did not match teachers and students (Villanueva-Mansilla, 2012:183). This documentation also reports that the laptops provided had a low ‘production level’; the software did not work as intended and the laptops were slow with many technical difficulties. They additionally report a lack of internet access, which formed an important component through which the programme aimed to achieve its intended outcomes. The Ministerio de Economía y Finanzas (n.d.) reports in terms of the OLPC programme in Peru that the appropriate infrastructure was not always in place (e.g. electrical outlets, work spaces). Villanueva-Mansilla (2012) also note that little thought was given to the access and connectivity of rural schools.

This finding was confirmed by David and Quispe (2013), who found that none of the schools in their sample had internet access. Finally, David and Quispe (2013) report that three schools in their study had support beyond the core programme elements. Two schools had ongoing ICT support, and one school provided children with education software beyond that already available on the laptop (remember that the OLPC laptops came loaded with 39 diverse applications).

Sharma (2014) furthermore reports that the OLPC Programme in Nepal faced similar issues. They report that 14 per cent of 906 grade 4 and 5 students surveyed in 2011 mentioned that their laptops had stopped working at least once. It has to be noted that a total of 34 per cent of these students noted that their laptop was accidentally dropped or stepped on. The authors additionally report that 18 of 28 schools reported instances where laptops stopped working. Most schools also reported it took at least five weeks to repair the laptops. Eight schools reported it took more than ten weeks. The authors note that: ‘*‘These delays in* *maintenance may have dissuaded teachers from regularly using the OLPC laptops in classrooms."* (p. 58). On the other hand, few laptops were lost or stolen.

Imbrogno (2014) notes that schools in both the MCT programmes in Mexico and Chile lacked computers. The study's research team rated schools in terms of four specific areas including basic inputs, infrastructure, implementation and learning environment. They report that: "Each treatment school was expected to have at least one computer per student in the computer labs, but that was in fact not the case in some of the schools. Schools which did not provide enough computers were obviously at a severe disadvantage concerning student access to the software part of the MCT curriculum. Students were forced to rotate and take turns in the computer lab, restricting the hours of access and amount of material covered in the math program." (Imbrogno, 2014:51-52).

***Regional and local government faced a lack of capacity for programme management* (Ministerio de Economía y Finanzas, n.d. (Process Evaluation)).**

The Ministerio de Economía y Finanzas (n.d.), reporting on the OLPC programme in Peru, reports that the regional and local governments and its implementation units had no clear plan on how to coordinate the programme's management. There were also additional costs that had not been budgeted for and a shortage of staff for the maintenance of equipment.

***Teachers did not receive sufficient training in delivering CAL and there appears to have been a lack of integration of the technology into existing teaching approaches* (David and Quispe 2014; De melo et al. n.d; Cristia et al., 2012; Sharma 2014 (all IEs); Villanueva-Mansilla, 2012 (Qualitative study); Examples of best practices in Peru, n.d. (Project Document).**

De melo et al. (n.d.:4) suggest that one of the reasons why the OLPC programme in Uruguay did not impact on test scores might have been that, although training was compulsory for school inspectors and teachers who applied as external consultants to support schools in the implementation process, teacher training was optional until after the period of evaluation.

De Melo et al. (n.d.:4) report that the OLPC programme in Uruguay was rapidly implemented because its governmental implementation unit reported directly to president rather than through the national educational administration (ANEP). This did mean that the coordination with teachers was not as fluent as desired. It is unclear to what extent this impact the actual implementation of the programme.

Design issues were also reported for the OLPC programme in Peru. Both the IE and additional documentation reported that the programme took a non-prescriptive approach as it was thought this would be a good approach to learning. As a result, teachers were never clear on how to incorporate the laptops into their teaching approaches. David and Quispe (2014) quote a teacher: ''We have the technology, but we were never told how to use it'' (p.7). Villanueva-Mansilla (2012) also report that computers were never fully integrated.

Cristia et al. (2012) report that the OLPC portal provided limited information and guidance for teacher on how to integrate computers into their regular pedagogical practices (p.6). David and Quispe (2013) report that only four schools were reported to have had training in how to integrate XO laptops into the curriculum. Moreover, their interviews with teachers and directors indicated that they felt there was insufficient support or training for teachers to help them to incorporate the use of the laptops into their lessons. Moreover, Villanueva-Mansilla (2012) report that insufficient training made teachers feel unqualified to use the laptops for teaching (pp. 6-9).

Sharma (2014) reports that similar problems existed for the OLPC Programme in Nepal. The author firstly reports that twenty-five percent of teachers reported not taking the original training and that the vast majority of these teacher did also not attend the refresher training session. Of the teacher who attended the training, about one third noted that the training was inadequate (p. 53). The authors secondly reports that *"the majority (59%) of teachers who were surveyed said that the 10-day residential and in-school training given at the beginning of school year was insufficient to enable them to properly integrate laptops into the regular classroom instruction process. "* (p. 56).

Additional documentation furthermore reports that teaching staff of the Huascaran Programme was often not prepared for the workshops as they did not have the experience needed for the training (mathematical and software knowledge). This meant that training was not as efficient as intended (Examples of best practices in Peru, n.d.).

Additional documentation (examples of best practices in Peru) reports that a lack of finance constrained the Huascaran programme. Budget constraints meant that the training workshops of the teachers were very short.

The majority of teachers took part in and passed the training **(Berlinski and Busso, 2013 (Impact Evaluation)).**

Berlinski and Busso (2013) note that the overall take up of the programme was high. The vast majority of teachers in the treatment arms took part in and passed the training.

Eighty per cent of parents attended the parent workshops (Humpage, 2013 (Impact Evaluation))

Humpage (2013) reports data from the Ministry of Education that "All parents were invited, and eighty per cent parents attended the workshops" (p.62)

Forty-two per cent of surveyed teachers report use of materials and laptops in non-programme grades (Sharma, 2014 (Impact Evaluation)).

Sharma (2014) reports that among teachers surveyed, 42 per cent reported they sometimes used materials and laptop in non-programme grades including in grade 4.

Laptops were used by non-receiving members of the household (Sharma, 2014 (Impact Evaluation))

Sharma (2014) reports that: *"almost 70% of grade two and three students, and more than 80% of the grade six students, who were surveyed said that at least one other family member uses their laptop at home. In the third round of the survey, 63% of the grade 4 and 5 students reported that others in their family also use their laptops. Their siblings are the ones who use their laptops the most. Forty-nine percent in year two said that they have taught their parents to use the computer. During teacher training, and in subsequent interactions with teachers and students, OLE Nepal has repeatedly emphasized that both teachers and students should encourage their friends and family members to use the laptops. Their philosophy was that expanding the user base for the laptops would positively contribute to the educational environment of the whole family and community. The findings discussed above suggest that the laptops are indeed having a significant positive spillover effect as envisioned by the program"* (Sharma et al., 2014: 62)

Intermediate outcomes: computer/ technology use

In the CAL interventions, one of the most important step along the causal chain is that students and teachers do actually use the computers. In the previous section, we have already reported issues with the provision of hardware and software. This section deals with the take-up of CAL by teachers and students. This information was provided for four programmes.

Use of computers was low among both students and teachers **(Barrera-Osorio and Linden 2009; David and Quispe 2013(Impact Evaluations)).**

David and Quispe (2013) also held interviews with teachers, students and educational authorities on the use and integration of the laptops in the Peruvian OLPC Programme. Most of the beneficiaries indicated that their use of the computers was minimal. Interviews suggest that computers were used in classrooms on average only once or twice a week. They were most often used for communication and less so for maths. The authors report that this might be due to the fact that teachers were not comfortable or certain about how the use the laptop in their teaching practices, and perhaps it interfered with their lessons. Interviews suggested that while some teachers were positive, there was also some unwillingness and reluctance amongst teacher to use the new technology; 67% of the Math teachers regarded the computers as neutral-negative, while 33 % regarded the computers as positive-very positive. This was different for Language teachers: 96% of language teachers regarded computers as positive-very positive (table 9 p.199).[154]

Barrera-Osorio and Linden (2009) administered baseline and follow-up questionnaires to both students and teachers (counterfactual analysis). They found that the CPE programme increased ''computer use among students and teachers by a surprising small amount, and [that] most of the use of computers by students is for the purposes of learning to use a computer rather than studying language'' (pp.15-16), which was one of the primary aims of the intervention. Reported use of the computers by teachers was additionally concentrated in the lower grades with older students' teachers reporting almost no computer use. It is for this reason the authors call the programme implementation 'a failure' (p.27).

Laptops were not used as frequently at home as intended **(Cristia et al., 2012; David and Quispe, 2013; Humpage, 2013 (Impact Evaluations))**

David and Quispe (2013) report that, while one for the key aspects of the OLPC programme was that children could also use the laptops at home, only one school in the Peruvian OLPC programme allowed students to do so. Cristia et al. (2012) report that in their study, about 40 percent of students took the laptops home in the week before their survey. Cristia et al. (2012) firstly note that principals might not have allowed children to do so as there would be no replacement if the laptops would be damaged or stolen, and secondly that parents were not adequately told that they would not be held financially responsible if a laptop would be stolen or lost. Hence, some parents preferred that the schools keep the laptops to avoid risk.

Counterfactual data collected by Humpage (2013) also reveals that the Pedagogical Support Pilot Programme (PSPP) in Peru, which was aimed at improving the original OLPC

---

[154] Implementation analysis scores were determined by the following 11 criteria: *Teachers* (1) Teachers are effectively using the five RTL steps in the correct sequence; (2) Teaching is done procedurally and with logical understanding and is not mechanical; (3) Teachers are innovative and committed to implementing the approach; (4) Teachers are motivated to support learners in numeracy and literacy outside teaching time; *Classroom Learning Environments*: (5) Appropriate learning materials are used; (6) The classroom library is utilized; (7) Children are reading age appropriate texts; (8) There is enhanced peer support among learners; *School Leadership*: (9) Head teachers provide technical support; (10) School and parents have a supportive relationship; (11) Functional school development plans prioritize lower grades.
Schools with scores 7-11 were considered 'high', scores of 5-6 'medium', and 0-4 were 'low' implementers relative to ideal implementation.

programme in Peru, the additional training under this programme also failed to have a significant lasting effect encouraging students to bring their laptops home (p.77-78).

***Students and teachers report high rates of computer usage, but around half of the use appear unrelated to the curriculum* (Cristia et al., 2012; de Melo et al., n.d; Ferrando et al., 2011 (Impact Evaluations)).**

Ferrando et al. (2011) report information on the use of computers. They report that the data pulled from the student, family and teacher questionnaires both in treatment and control groups suggests that the OLPC programme in Uruguay had positive effects on computer use amongst teachers and students; a high proportion of those who had been given a laptop had used it between once a week or more (73 and 87 per cent respectively); 80 per cent of students who received the laptop used it between once a week or more and once a month or more; 80 per cent of students surveyed brought the computer to school at least once a week, and around 45 per cent did so every day;87 per cent of teachers brought their laptop to school almost daily or at least a few times a week, while this percentage was 54 per cent amongst control teachers; and 98 per cent of students have teachers who use the laptops as a teaching tool, half of these teachers used them for mathematics specifically. There was also a higher number of teachers with access to computers.

However, de Melo et al. (n.d.: 15-16) report that laptop use was not widespread across all public schools and that students use their laptops most frequently for downloading information from the internet (68 per cent), suggesting that the laptops should not impact on reading and language outcomes.

In their study of the OLPC programme in Peru, Cristia et al. (2012), find that the programme generated a large increase in computer use both at school and at home (p.13). The authors report that about *"82 percent of students in the treatment group reported having used a computer at school during the previous week versus 26 percent in the control group"* (p.13). They also find large effects on computer use at home: *"42 percent of treatment students reported using a computer at home in the previous week compared with 4 percent of students in control schools"* (p.13). The authors moreover provide data from laptop logs which is in line with the reported information above and suggests that a sizeable share of students used the laptops intensely: *"Almost half of students started four or more sessions, 35 percent started between one and three and 15 percent did not use the laptop in the previous week"; "the average session lasted about 40 minutes" (pp.13-14).*

Use was concentrated in the time when schools were open and on days when computer use was heavier at school, computer use seemed to be heavier at home as well, suggesting some spill over from school to homes took place. The logs also shed light on how laptops were used. Standard applications (e.g. word processor, browser, calculator) account for 45 per cent of applications use; Games accounted for 18 per cent use; Music applications for 14 per cent use, and programming for 5 per cent. Finally, the rest of the applications accounted for 18 percent of use, and the most important were an application for recording sound and video and Wikipedia (8 and 4 percent, respectively). Cristia et al. (2012) also tested how resourceful students in treatment groups were in using the laptops. These results indicate that students in the treatment group displayed some useful skills in operating the laptop, though they showed certain limitations in mastering a range of applications (p.14).

Positive impact on teachers' computer skills and teachers' laptop use for lessons seen in the third week after programme faded after two years (Humpage, 2013(Impact Evaluation)).

Implementation and process data were collected three weeks after the start of The Pedagogical Support Pilot Programme (PSPP). These data show that "teachers began using the laptops more frequently and with a wider variety of applications by the trainers' second visit" (Humpage, 2013: 72-73). Humpage specifically notes that: "*Whereas in the first visit, only 64% of teachers could execute basic tasks on the XO like saving files to a USB or sharing files with other computers, at the beginning of the second visit*[155]*, 95% or more of the teachers could do these things" (ibid.).*

Humpage (2013) also assess counterfactual data on whether participating in the Pedagogical Support Pilot Programme (PSPP) had a lasting effect *from the end of 2010 to early 2012 on use of the laptops and on teacher and student opinions of the laptops.* This data shown no significant effect on teachers' self-reported computer skills (p.75).

Similar outcomes are reported for the use of laptops by teachers. While the data collected three week after the programme suggests an increase in the number of teachers who used laptops in their lesson plans[156], the long-term and counterfactual data suggests otherwise (see results under *"The programme did not increase laptop use among teachers and students for curricular areas, but resulted in use of more academic applications"*).

Humpage (2013) notes that: '*'These results suggest that the strong effects the training appeared to have on the variety of applications teachers used and the frequency with which they used them during the training faded after two years."* (p.77).

Training did not significantly reduce teacher-reported barriers to using the XO laptops (Humpage, 2013 (Impact Evaluation))

Humpage (2013) assesses whether participating in the Pedagogical Support Pilot Programme (PSPP) had a lasting effect *from the end of 2010 to early 2012 on use of the laptops and on teacher and student opinions of the laptops, including the various barriers teachers face when using the laptops.* Humpage (2013) reports: *"Table 3.6 reports the treatment effect on the likelihood that teachers report facing various barriers to using the XO laptops: problems with electricity access, with activating the laptops, with laptops breaking, connecting to the local network known as the "neighborhood", understanding applications, using the touchpad or mouse or an index of all six potential problems. The training did not significantly reduce teacher-reported trouble with any of these in the full sample or the 2010 teacher sample, although the effect is negative (indicating fewer problems) for five out of the six outcomes. Surprisingly, the treatment effect on the having trouble using the neighborhood network is positive and significant, indicating that teachers in the treatment group were 20.5 percentage points more likely to have trouble connecting to the local network, or to have never tried connecting. It could be that teachers in the treatment group are more likely to have had more experience with the local network, giving them more opportunities to have had trouble with it" (p. 74).*

Programme did not increase laptop use among teachers and students for curricular areas, but did increase use of academic applications (Humpage, 2013(Impact Evaluation)).

As reported above, the positive impact of teachers' computer skills seen in the third week after the programme faded after two years. Similar outcomes are reported for the use of laptops by teachers. While the data collected three week after the programme suggests an

[155] The second visit took place three weeks after the start of the training

[156] "*The percent of teachers that included the XO laptops explicitly in their lesson plans increased from 13% to 73%, and the average number of activities they planned with the XOs increased from 1.15 ctivities over the last three lessons in seven curricular areas to 11.18 activities."* (Humpage, 2013: 72-73)

increase in the number of teachers who used laptops in their lesson plans[157], the long-term and counterfactual data suggests otherwise.

Humpage (2013) assess whether participating in the Pedagogical Support Pilot Programme (PSPP) had a lasting effect from the end of 2010 to early 2012 on use of the laptops and on teacher and student opinions of the laptops. Although the author reports that the programme increased the likelihood of teachers having used a pc in the last week by 15.3 percentage points ($p < 0.05$), both teacher and student survey data, as well as data from computer logs reveals that the programme did not increase use of laptops by teachers and students for curricular areas. First of all the programme did not "have an effect on teacher-reported use for any of the curricular areas" (p. 77). In fact, the training had *"a significant negative effect on the intensity of XO use, defined as the number of applications used multiplied by the number of times they use them, reducing the number of reported application uses in the last week by 0.349 uses in the full sample ($p < 0.1$) and by 0.458 uses in the 2010 teacher sample (significant at the 5% level)"* (p.77). This is confirmed by data from the computer logs:

"The laptops' logs provide an objective source of data on how students use their laptops, capturing data on the most recent four sessions on the laptop. Looking at activity in the most recent week, 35% of children in the control group used their XO in the past week, compared to 31% of children in the treatment group (see Figure 3.2 for more detail). The results presented in Table 3.10 show that the treatment had a negative effect on the average number of sessions in the last week. The treatment effect was significant and negative for the 2010 teacher group, reducing the number of sessions by 0.390 sessions ($p < 0.05$). Looking at the treatment effect by grade, the effect is significant and negative for 4th graders and 6th graders in the 2010 teacher sample" (Humpage, 2013: 77).

Secondly, the author found mixed results on the programme impact of the types of applications used. While survey data shows a significant effect on use of applications that were covered in the training, in contract, there is no evidence from the computer logs that the training training increased the use of the applications covered in the training as a percentage of all application uses. These logs do, however, show a significant increase in the use of math and programming applications and a decrease in the use of music applications. The authors notes that "this could be interpreted as teachers using the computers for more academic pursuits." (p. 78).

Students report high rates of use of technology, prescribed software and class materials **(Berlinski and Busso, 2013 (Impact Evaluation))**

Berlinski and Busso (2013) use their endline student surveys to create indicators of class material and technology use. They report that all estimates for the use of technology and class materials are positive and large (p.21). They also report that students' workbooks and teachers' manuals were being used in almost all of the treated classroom. Both treatment arms used the prescribed software. This use was, as expected, lower in the computer lab treatment arms because schools using computer labs were supposed to use the lab only once or twice a week. There was no difference in use of the internet in the classroom between treatment and control schools

[157] "*The percent of teachers that included the XO laptops explicitly in their lesson plans increased from 13% to 73%, and the average number of activities they planned with the XOs increased from 1.15 activities over the last three lessons in seven curricular areas to 11.18 activities."* (Humpage, 2013: 72-73)

Students and teachers find educational software useful and easy to use but there was insufficient use thereof both in and outside classroom (Sharma, 2014; Imbrogno, 2014(Impact Evaluations))

Sharma (2014) reports that, in the case of the OLPC Programme in Nepal, the majority of teachers and students reported to like digital contents; find them easy to use; consider them appropriate to student level and interests; find them useful; and find they make teaching much easier. Teachers' positive attitude on the materials might have been influenced by their participation in the development of the materials. The author of the evaluation reports that teachers in the test schools were ''extensively consulted while developing the initial digital contents'' (Sharma, 2014: 48). The authors moreover report that the majority of students (93 per cent) report being able to use the laptops. Sharma (2014) also reports, however, that laptops were not always used for their intended purposes: only two-third of grade 4 and 5 students report to have used the laptops to read the educational digital materials developed as part of the programme; 33 per cent of grade 6 students report they did not use the digital educational materials provided by the programme; 43 per cent of grade 4 and 5 students reported they examined the digital materials only once outside of the classroom; 55 per cent of grade 4 and 5 students reported sometimes using the e-library; about one-third of students report to have read five or more books from the e-library; almost a quarter of students report that both them and their friends use the laptop for computer games and for taking photos.

Imbrogno (2014) notes that students in both the MCT programmes in Mexico and Chile report high levels of satisfaction with the new curriculum and its implementation. Imbrogno (2014) reports that *"Most of the students rate the ease of use of the Tutor, teacher help with the Tutor, and effectiveness of the Tutor highly or very highly"* and that *"Students feel the MCT helps them learn mathematics, they enjoy using it, and they find the teacher to be supportive"*(p. 69). Nonetheless, Imbrogno (2014) also reports that in both the MCT programmes in Mexico and Chile the standard for total MCT software usage hours by students was not met by any classroom in any school. They additionally note that very few students used the software in their own time: *"The MCT curriculum calls for 40% of class time to be devoted to the software. In practice, this would constitute two days of a typical school week, which over a six month time period would mean roughly 25 hours spent in the computer lab in lieu of traditional math lectures. Note that this standard for total MCT software usage hours by students was not met, on average, by any classroom in any school. Conversations with the school principals revelaed that the twice per week standard for computer use was rarely met. Students often only visited the lab once each week during class time, and very few of them accessed the software in their own time."* (p.51)

Treatment students report to have better computer skills than control students (Sharma, 2014 (Impact Evaluation)).

In the context of the OLPC Programme in Nepal, Sharma (2014) reports: *"Twenty-four percent of the students in grade 4 and 5 in program schools in 2011 say they can use the internet well, while only two percent say so among students in comparison schools. Similarly, 66 % of the program school students said they type well while only 9% said so in comparison schools. Eighty-nine of the program school students say they know how to use a computer—the corresponding figure for comparison school students is 15%. Though the benefits of these skills in the immediate future are unknown, one would expect these sets of skills to be useful later in life.''(p. 64).*

The number of programme sections mastered correlates to improved test scores (Imbrogno, 2014 (Impact Evaluation))

Imbrogno (2014) reports that while the number of total hours of usage is not significant in predicting post-test scores ('not significant in Mexico, marginally significant in the negative direction in Chile, p. 67), the number of programme sections mastered does correlate to improved test scores: *"Across the 310 students in Chile, the sections mastered (skills mastered, units) variable has an average of 24.3 (252.3, 8.2) and standard deviation of 12.8 (154.2, 3.9). The regression results show that an increase in usage of the MCT software by one standard deviation of sections mastered (skills mastered, units) improves posttest scores by 0.28 (0.31, 0.25) standard deviations. That is a consequential increase in achievement on the posttest. In Mexico, the sections mastered (skills mastered, units) variable has an average of 21.2 (190.3, 7.1) and standard deviation of 12.9 (144.1, 3.8). Our results there show that an increase in usage of the MCT software by one standard deviation of sections mastered (skills mastered, units) improves posttest scores by 0.40 (0.28, 0.30) standard deviations" (p. 66-67).*[158]

Infrastructure and implementation school ratings correlate to student MCT usage (Imbrogno, 2014 (Impact Evaluation))

Imbrogno's (2014) research team rated schools in the MCT progammes in both Chile and Mexico in terms of four specific areas including basic inputs, infrastructure, implementation and learning environment and correlated these outcomes with test scores. The authors reports that: *"infrastructure and implementation school ratings are highly correlated with student MCT usage. Though this result is based on just ten observations*[159] *and is therefore not the strongest in this paper, we believe it has practical importance for policy considerations… Every school was expected to have one computer per student and reliable connectivity to the internet, but in reality this was not seen. Those schools which experienced this ideal infrastructure possessed an environment which allowed their students to excel, while those who adopted the MCT curriculum but did not have the ability to properly use it saw their students lag behind… Students will also master more skills using the MCT when the frequency of lab use and the conditions in the lab (learning environment) are high…. The schools which most closely followed the recommended implementation activities saw their students complete and master more of the MCT software curriculum"* (p.68)

Intermediate outcomes: motivational outcomes and classroom activity

It was hypothesises by some authors that CAL may lead to improved educational outcomes (incl. learning, attendance, drop-out, enrolment rates) through pathways such as interest in learning and self-efficacy. Information in this category was provided for five of the programmes.

***Students in some CAL programmes report increased levels of motivation and self-efficacy* and self-esteem (Ferrando et al., 2011; Lai et al., n.d; Lai et al., 2012; Lai et al., 2013; Lai et al., 2011a (Impact Evaluations)).**

Ferrando et al. (2011) report information on motivation in response to the OLPC programme in Uruguay. They report that the data from the student, family and teacher questionnaires both in treatment and control groups suggests that motivation might have been a channel in

[158] *"The prepared Bridge to Algebra curriculum consists of 14 units, 57 sections, and 552 skills."* (Imbrogno, 2014: 41)

[159] Ten observations refer to the ten schools in which the programme was implemented

student learning: *"a very high number both mothers and teachers express that ever since students have an xo, they show more motivation. Most of these students' teachers believe that the tool is good or very good for learning (94 per cent),that the fact that each student has a laptop for personal use is positive (91 per cent) and that motivation is particularly high in mathematics and language tasks (61 per cent). In the case of the parents, they indicate that their children show differences in behaviour ever since they received their xo: they look for more information for their homework (57 per cent), they are more motivated to attend school 36.6 per cent) and they are happier (34.4 per cent)"* (pp.38-39)

Lai et al. (n.d.) and Lai et al. (2012, 2013) reports that students in the language CAL and CAL boarding school programmes in China enjoyed the programmes. In addition, both Lai et al. (2012, 2013) and Mo et al. (n.d.) note that the CAL sessions in both programmes in China were very intense, signifying the time of the sessions was well spent.

The researchers of the various CAL interventions in China assessed the effect of CAL on interest in learning and self-efficacy as part of their counterfactual analyses. First of all, Lai et al. (n.d.), who studied the language CAL intervention in China, report that "*Compared to students in the control group, the students in the treatment group "like school" more and had higher levels of self-efficacy of Mandarin studying and self-confidence. In particular, students in the treatment group had significantly higher levels of self-efficacy of Mandarin studying than those in the control group (row 1, column 1), yet the differences between the treatment and control groups in the other two indicators were insignificant (row 1,columns 2 and 3)."* (Lai et al., n.d.: 20)

Lai et al. (2011a), who implemented a math CAL intervention in migrant schools in China, also found that c*ompared to students in the control group, the students in the treatment group "liked school" significantly more, and that self-confidence significantly improved in the treatment group compared to the comparison group. On the other hand, they find no significant effects in self-reported math study efficacy.* One possible reason for this, the authors note, might be the programme's remedial nature as it focused on repeated exercise rather than the students' ability to think creatively and solve problems.

***Students in a CAL programme in Costa Rica report a more active learning environment* (Berlinski and Busso, 2013 (Impact Evaluation)).**

Berlinski and Busso (2013) report outcomes of classroom activity that provide some insight into the workings of the programme. They conclude that the intervention generated a more active learning environment and an increase in classroom activity. That is, students in the treatment group report "explaining concepts to the class more often, preparing more exercises for others to solve, and frequently discussing possible solutions or arguments with other students" (p.)

Treatment classes devoted more time to discussion and less to teacher lecturing, and students are less likely to work individually. This is according to design, as the programme's new curriculum prescribed more time is devoted to exploration and formalization and less to practice. The authors show that *"in all treatment groups were stimulated in ways consistent with the objective of achieving mathematical competence. In particular, the class observer recorded whether students make, explain and validate mathematical conjectures, explain relations between concepts, manipulate propositions, or discover mathematical rules from observing and analyzing patterns. The first scale looks at students prescribed learning practices while the second looks at whether or not teachers purposefully foster those practices. We see positive point estimates for all groups with larger magnitudes and statistical significance in the technology arms"* (p.22)

Seventy-nine per cent of teachers report increase in workload after programme (Sharma, 2014 (Impact Evaluation))

Sharma (2014) reports that 79 per cent of teachers delivering the OLPC programme in Nepal reported that their workload increased significantly as a result of the programme. The author reports that: *"It is reasonable to assume that they reverted, to a large extent, to traditional ways of teaching and ignored the materials in the laptop. Under such circumstances, the effect of laptops with educational content on student learning would be minimal.*" (p. 53)

***Students in a CAL programme in Chinese boarding schools did not report higher levels of study motivation or self-efficacy* (Lai et al., 2012(Impact Evaluation)).**

Lai et al. (2012) report that the CAL programme in Chinese boarding schools did not have any significant impact on non-cognitive outcomes. They report that compared to students in the control group, the students in the treatment group did not "like school" more. According to the data, boarding school students did not report higher levels of math study efficacy or metacognition after the CAL program. The authors of this study also report that the programme's remedial nature and lack of focus on problem solving skills might explain this.

The programme did not increase teachers' or students' enthusiasm for the laptops (Humpage, 2013 (Impact Evaluation)

Humpage (2013) assess whether participating in the Pedagogical Support Pilot Programme (PSPP) had a lasting effect from the end of 2010 to early 2012 on use of the laptops and on teacher and student opinions of the laptops. In terms of teachers' opinions on the programme, the authors notes: *"In the teacher survey, teachers were asked whether they agreed or disagreed with a series of statements about the XO laptops, such as, "The laptops are just for playing," or "Children learn more working on a laptop than on paper." The estimated effect of the training was negative for both estimates; for the 2010 teacher sample, the training significantly decreased teachers' score on an eight-point index of positive opinions about the XO laptops by 0.824 points (p < 0.01), suggesting that the training was not successful in one of its main objectives to increase enthusiasm for the laptops."* (p.75).

The author also notes: *"In the student survey, students expressed their preferences for working on a laptop over various alternatives for learning and for play. The training did not have a significant effect on an index of positive opinions about the laptops, but the coefficient estimates are negative overall and for each grade in both samples. The coefficient estimates suggest that older children may have more negative opinions about the laptops. A potential explanation is that the novelty may wear off for students who have used the laptops for a number of years"* (p.76). It is noted however, that the programme did *"significantly increase fourth graders' families' odds of owning a personal computer by 9.4 percentage points in the full fourth grade sample and by 10.1 percentage points in the sample of fourth graders whose teachers were at the same school in 2010 (p < 0.01 for both estimates)"* (p.76).

Intermediate outcomes: spillovers

Programme resulted in English test score spill overs to non-programme students in grade 4 but not to non-receiving members of treated students' household who reportedly used laptops (Sharma, 2014 (Impact Evaluation))

Sharma (2014) reports that the OLPC programme in Nepal resulted in different types of spillovers. The authors first of all also measured test score outcomes for grade 4 students, who did not receive the intervention, in both programme and non-programme schools. They find that grade 4 students in programme schools score significantly better in English than students in non-programme schools. A possible explanation of these spillover effects might be that, among teachers surveyed, 42 per cent reported they sometimes used materials and laptop in non-programme grades including in grade 4.

The authors moreover note that these spillover outcomes are not dependent on having a sibling in a programme grade: ''*The test scores of grade 4 students in program schools who have siblings with laptops were not statistically different from those without siblings with laptops, as shown in Table 26''* (Sharma et al., 2014: 62). This is despite the reported use of laptops by non-receiving household members. *This use by non-receiving members " suggest that the laptops are indeed having a significant positive spillover effect as envisioned by the program. However, there has been no spillover onto test scores yet. "* (Sharma et al., 2014: 62).

Contextual Barriers and Facilitators

Many public school math teachers have little or no training in Math (Imbrogno, 2014 (Impact Evaluation))

Imbrogno (2014) reports that, in Chile, ''many of the public school math teachers have little or no formal training in math'' (p.44). He also reports that the situation is ''much the same in Mexico'' (*ibid*.).

New Schools and Infrastructure

***The PACE-A program faced issues with the delivery of the teacher training component* (USAID, 2010).**

Results from an audit by USAID (2010) suggests that the PACE-A programme faced issues with the delivery of its teacher training component. It is reported that the majority of the primary education teachers received only a portion of the training course, while only three per cent received the full course. Furthermore, only 38 per cent received the key workshop essential to operating a class and 21 per cent did not receive basic training[160]. In addition, a lack of training standards resulted in training inconsistencies across implementing agencies and their training curricula. The document also reports that these issues are likely to have largely[161] resulted from record keeping, monitoring and internal review issues by implementing partners. There were additional monitoring issues.

Firstly, heavy workload and other priories meant that the PMU monitoring and evaluation officer had not been able to perform field testing (USAID, 2010). Secondly, due to difficulties of travelling to remote locations, CRS staff visited the projects once every few months rather than every month, as initially intended (Burde and Linden, 2013:4).

***Integration of new schools with existing MoE schools was not always effective* (USAID 2010).**

The audit administered by USAID (2010) also found that the process of class integration with MoE schools in the PACE-A programme was not always effectively carried out. In some

[160] These teachers did have 2.2 years of teacher experience (USAID, 2010)
[161] USAID (2010) notes that in some cases teachers (and other users) had difficulties completing forms due to limited literacy.

cases, this negatively affected students. As the process involved the relation of classes to the nearest MoE school, often far away from the original communities, students ended up walking long distances each day (USAID, 2010:10).

***Distance to school found to have a detrimental effect on test scores* (Burde and Linden, 2011, 2013).**

As noted above, Burde and Linden's (2011, 2013) programme theory suggests that distance has an impact on educational outcomes as traveling long distances to school takes up significant ivnestments in time, transportation, and alternative housing strategies. The authors estimate the effects of distance on academic outcomes and find that enrolment and test scores decrease by 16 percentage points and 0.19 standard deviations per mile. They additionally find that distance affects girls more than boys as *"girls' enrolment falls by 6 percentage points more per mile (19 percentage points total per mile) and their test scores fall by an additional 0.09 standard deviations (0.24 standard deviations total per mile)"* (Burde and Linden, 2011: abstract).

However, the authors also note Burde and Linden (2013:28) note that while the Afghan government and other donor countries have focused on improving access to schooling through building schools, there might be more important impediments that prevent girls from going to school. The opportunity costs of a girl going to school, for example, are extremely high. Such impediment, as well as other cultural and religious factors could prevent a girl from going to school regardless of the number of schools available.

***The program was implemented in a context effected by decades of war, characterised by low school enrolment rates and a large gender gap (*Burde and Linden, 2009)**

Burde and Linden note that Afghanistan has been crippled by decades of war. As of 2007, half of school-age children were unenrolled, and of primary school-age children, only 37 percent attended school (2009:7). There is a significant gender gap.

Descriptive findings: Improvement or Replacement of School Infrastructure

***Implementing agency performance rated as unsatisfactory* (World Bank, 2001).**

Additional documentation reported that implementing agency performance of the IRP evaluated by Lokshin and Yemtsov (2003, 2004) was unsatisfactory (World Bank, 2001). In addition, additional documentation reports that the IRP faces some financial management issues. The World Bank (2001) reports that as since the Project Implementation Unit (PIU) worked in the very difficult environment and had to rapidly implement the project, more than 80 % of the credit was disbursed by 1997. Since the endine assessment of Lokshin and Yemtsov's (2003, 2004) study was in 1997, this probably did not affect the study's outcomes. It should be kept in mind, however, that the programme might not have been sustainable beyond 1997, and therefore, the impacts of the programme not lasting in the long term. The World Bank (2001) reports an additional reason why the project was less effectively implemented towards the end of the project. While the programme was supposed to end in December 1997, it actually ended in June 2006. Financial management issues additionally causes the World Bank to freeze the Special Accounts funds in a local bank. This in turn affected project implementation (World Bank, 2001:14).

***Lack of sufficient organisational and financial management systems* (World Bank, 1996; World Bank OED, 1997; World Bank, 2001).**

The SIF programme evaluated by Newman et al. (2002) had issues with their organisational and management capacity. Social sector ministries were weak and sectoral policies virtually non-existent (World Bank OED, 1997:1).

The information and financial management systems of SIF, evaluated by Newman et al. (2002), were reported to be inadequate (World Bank, 1996). Additional documentation to Newman et al. 's (2002) study, reports that the SIF also faced financial management issues as municipalities did not always keep track of their counterpart financing obligations, leading to overcommitments (World Bank, 2001, p.21). Moreover, The World Bank (2001) notes that one problem affecting the SIF evaluated by Newman et al. (2002) was the lack of counterpart funding from municipalities. This lack of counterpart finance also delayed the project implementation. These problems were perceived to be a problem wih the SIF by project beneficiaries, ''when in fact it was a problem of poor planning on the part of municipal governments'' (p.18).

Another factor was significant fluctuation in the amount of money channeled through SIF. It particularly led to an increase in administrative costs as a percentage of the total costs. Adminstrative capacity largely involved fixed cost (personel, offices, vehicles) that could not be adjusted to a budget decrease. Greather predictability would have allowed the SIF better control of their administrative costs (World Bank, 2001:18-19).

***Delays to project implementation due to lack of counterpart funding and politicisation* (World Bank, 1996; World Bank, 2001).**

The information and financial management systems of SIF, evaluated by Newman et al. (2002), were inadequate (World Bank, 1996). Additional documentation to Newman et al. 's (2002) study, reports that the SIF also faced financial management issues as municipalities did not always keep track of their counterpart financing obligations, leading to overcommitments (World Bank, 2001, p.21). Moreover, The World Bank (2001) notes that one problem affecting the SIF evaluated by Newman et al. (2002) was the lack of counterpart funding from municipalities. This lack of counterpart finance also delayed the project implementation. These problems were perceived to be a problem wih the SIF by project beneficiaries, ''when in fact it was a problem of poor planning on the part of municipal governments'' (p.18).

Additional documentation for the study evaluated by Newman et al. (2002) reports that the SIF demonstrated its support through the programme by agreeing to finance it (incl. SIF salaries nad administrative costs), increasing its scale and by agreeing to incorporate SIF's staff into the Civil Service Reform Program (World Bank, 1993:7). A different document reports that in the period between 1991-1993 the SIF faced increased levels of politicisation that led to delays of project implementation in some areas. Despite this politicisation the SIF retained its technical quality, which is reflected in UDAPSO's finding that beneficiary communities regarded SIF as a non-political technical institution (World Bank, 1996:7-8).

***All libraries in the Akshara Library Programme reported to function as expected* and were utilised at intended levels of intensity (Borkum et al., 2013).**

On the other hand, Borkum et al. (2013) reported that all libraries in the Akshara Library Programme functioned as expected (p.2).

Borkum et al. (2013) note that while the ''average fraction of children visiting the library in each month of the study [fluctuated] between 50 and 90 percent''(p.5), overall, ''81 percent of children visited the libraries at least once a month during the academic year'' (ibid.). In addition, if one excludes exams and holidays, children on average visited the library 2.4

times and borrowed 1.3 books a month (ibid.). Usage of libraries and books was slightly higher in hub-schools, with an average of 2.65 visits and 1.35 books per month compared to 2.12 visits and 1.16 books per months in spoke schools.

***High staff turnover led to a deterioration of project management* (The World Bank, 2001 (Project Document); World Bank, 1993 (Project Document).**

Several studies also report issues of staff turnover. As the original closing date of the IRP neared, most qualified staff took on jobs in other PIUs, resulting in a deterioration in the quality of project management. This in turn led to delays in procuring the final three contracts and several extensions. The document also note that this implementation progress was directly related to the IRP's ownership. In the first several years of the programme there was strong ownership. As the project neared its closing data, lack of ownership adversely affected the project performance (The World Bank, 2001:9)

Additional documentation for the study evaluated by Newman et al. (2002) reports that the SIF faced similar issues (World Bank, 1993). It reports that SIF faced weak management between 1991-1993, as well as management changes that impacted on project implementation as a result of a loss of institutional memory. In addition, it is reported that ''SIF was often criticized for an overly engineering view of the world, with insufficient attention paid to the software aspects of the intervention. This was certainly the case with the education investments of the SIF'' (World Bank, 2001: 19).

***There was no change in household expenditure on education* (Lokshin and Yemtsov, 2003).**

The authors note that *''the rehabilitation of school infrastructure fails to produce significant changes in parents' assessment of schooling conditions''* as *''the the differences in changes in expenditures between the treatment and control group villages are insignificant''* (p.17).

***Civil unrest and polical instability a barrier to project implementation* (World Bank, 1998; World Bank, 2001).**

The World Bank (2001) notes that the IRP in Georgia as evaluated by Lokshin and Yemtsov (2003, 2004) was affected by ongoing civil unrest, which ''which stopped all transportation along the major land links to Russia and especially during the early years of the projects'' (p.13). This affected the programme as private sector bidders were unwilling to bid on works in or deliver goods to Georgia, resulting in several failed procurements and problems with the transport of goods. The World Bank 2001 notes that: ''the PIU believes one of the major accomplishmentsin those early years was in successfully arranging for goods to be shipped to Georgia through areas such as Chechnya without any losses in transit'' (p.9).

A different World Bank documents (1998) reports that political instability was generally a problem that was feared to affect the implementation capacity prior to the intervention. The project incorporated specific measures to prevent this as much as possible[162]. The World Bank (2001) notes that the country saw increased political stability between 1994 and 2000, resulting in increased foreign investments.

***Power outages, poor quality telephone lines and the lack of reliable delivery services formed a significant barrier to the implementation of the project* (The World Bank, 2001).**

---

[162] Project used a non-bureaucratic implementing agency and the maintenance of management procedures was emphasised in providing technical assistance (Word Bank, 1998:3).

The implementation of the programme was initially affected by the country's poor infrastructure. The World Bank (2001) reports that power outages, poor quality telephone lines and the lack of reliable delivery services formed a significant barrier to the implementation of the project.

Descriptive findings: Sanitation & hygiene interventions

***Sanitation school improvements not fully compliant, with few meeting government standards* (Freeman et al., 2012).**

The hygiene and sanitation interventions implemented by Freeman et al. (2012) seemed to have had some issues in terms of its implementation. The authors note that, ''although there were significant differences between intervention and control groups at follow-up[163], a substantial proportion of school improvements did not meet standards necessary to be considered fully compliant'' (p.384) as results from the school surveys that fewer than 40% of pupils in schools from either intervention arm reported that soap was always available; approximately 60% reported that water was always treated; and >75% reported drinking water was always available (p.384). In addition, it is noted that schools that received latrines approximately halved their pupil-latrine ratio, few, however, achieved Government standards.

***The share of female teachers increased in schools with latrine construction* (Adukia, 2014).**

As mentioned above, Adukia's (2014) programme theory asserts that latrine construction may also improve educational outcomes through impacting on female teacher attendance. This is because the absence of proper sanitation facilities affects female teachers as much as female students. The author finds that *''the impact of latrines is not generally higher in schools that had a higher initial share of female teachers, which suggests either that female teachers did not reduce their absenteeism relative to male teachers or that female teachers' work attendance did not impact student enrollment''* (p.4). The author does find, however, that latrine construction raises the share of female teachers per school, in particular when latrines are sex-specific. This in turn may positively impact female students and encourage their enrolment (p.4).

***WASH knowledge and behaviour improved substantially in treatment schools* (Freeman et al., 2012).**

Freeman et al. (2012) provide a counterfactual analysis of the uptake of hygiene knowledge and practices by pupils. The authors find significant and substantial differences in pupil WASH knowledge between intervention and control groups following the intervention[164]. In addition, key hand washing times and scores on a hand washing demonstration significantly increased in treatment schools (p. 384). Furthermore, the hygiene promoted seems to have promoted schools to take-up hygiene measures other than those provided as part of the intervention: ''Intervention schools – where no water supply improvement or soap was provided – significantly improved in consistent provision of drinking water, hand washing water and soap, as compared to control schools'' (p.394). Nonetheless, fewer than 40 % of pupils in treatment schools reported that soap was always available.

[163] Table 2 on page 385 of Freeman et al. (2012) shows there were significant differences in terms of hygiene and sanitation between treatment and control groups at follow-up (e.g. drink water availability)
[164] This information is provided in Table 2 on page 385 of Freeman et al. (2012)

***Sex-specific latrines appear important for sanitation programmes to benefit girls (*Adukia, 2014).**

The programme theories of both studies in this sub-category stresses the importance of latrines in enhancing security and in turn, promote enrolment, attendance and learning. It is also noted that sex-specific latrines might be specifically important for improving eduational outcomes for girls. This is confirmed by Adukia's (2014) analysis of the impact of latrine types. While the initial results only provide limited evidence that the impact of latrine is larger for girls than boys, this becomes clearer when looking at the impact of sex-specific latrines. The author reports that while "Unisex latrines have a greater impact on pubescent-age boys than pubescent-age girls. Pubescent-age girls benefit little from a unisex latrine but benefit greatly from sex-specific latrines''. The results also show that privacy and safety become important at older age as the construction of unisex latrines exacerbates gender disparities at older age and as younger girls benefit more from unisex latrines than older girls. Nonetheless, *''sex-specific latrines have some additional impacts at younger ages, which is consistent with concerns of bullying at younger ages for both boys and girls"* (pp. 3-4).

***Headmasters displayed a lack of awareness of sanitation issues (*Adukia, 2014).**

Adukia (2014) held interviews with headmasters and students as part of his/her study. The results reveal a lack of awareness surrounding sanitation issues among headmasters. While headmasters reported that latrines were not necessary for the regimented hygiene routines during the shool day, students reported that access to latrines was an important issue that affected their education. Nationally, however, sanitation has been an important part of India's national agenda since the 1950s, resulting in an increase in the number of latrines. The SSHE campaign evaluated by Adukia (2014) formed an important part of this.

***Lack of latrines led to girls' concerns safety and privacy* (Adukia, 2014).**

Adukia's interviews with young girls also highlighted safety and privacy concerns during interviews at home. Many reported that, in the absence of latrines, they were forced to use various school premises for defecation (e.g. behind a tree, behind a school sign). One student described an atmosphere of fear amongst girls of being assaulted when isolated from view. This 'discouraged her and her friendsfrom eating, drinking, and relieving themselves during the school day'' (p.9).

***Post election violence caused considerable disruption to programme implementation* (Freeman et al., 2012).**

Freeman et al (2012) that post election violence in 2008 caused considerable disruption to programme implementation. They additionally assess the extent of migration and destruction of property in their study communities. The survey finds severe destruction of property in some communities compared to others. Nonetheles, ''there was no statistical difference between intervention packages (P = 0.08)'' (p.384). In addition, the authors report that while some or severe migration occurred in all geographical strata, it was greater in in Nyando/ Kisumu East (47%) than Rachuonyo (24%) or Suba (29%), (P = 0.02)" (p. 384).

Teacher Level

Teacher Incentives

In the following section, we report the results of the synthesis of qualitative findings from our included teacher incentives programmes, presented along the hypothesised programme theory which is used as an overall framework for structuring the synthesis. This first section looks at implementation features that may be associated with relative success and failure of the included teacher incentives programmes, followed by a review of the contextual barriers and facilitators to success.

Targeting/participants

The included studies present very limited information about how the evaluated interventions were targeted, so it is not possible to provide a detailed account of the targeting criteria and mechanisms used. What follows is therefore an overview of the general targeting approach reported in the included studies and associated documents, and some detail on targeting criteria from those studies where this was reported.

Four of the included studies report that the programmes were targeted at particular school types. The ALI experiment in Mexico was targeted at Federal Upper Secondary schools with a technical orientation and agricultural and industrial focus (Behrman et al., 2012). The SNED programme in Mexico was targeted at state-funded and state-subsidised primary and secondary schools (Gallego, 2008a). The Andra Pradesh Randomised evaluation was targeted at state funded primary schools (Muralidharan and Sundararaman, 2011) and the Seva Mandir Teacher incentive programme was targeted at Non-formal Education Centres in India (Duflo et al., 2012).

The remaining three programmes used a particular feature of the population as targeting criteria. The ICS teacher incentive programme in Kenya targeted schools based on their performance status – selecting schools that scored below the district average on test scores (Glewwe et al., 2010). The META programme selected schools based on their location, focusing on rural schools (Glewwe et al., 2010). The Carrera Magisterial targeted teachers based on seniority (teachers had to have a minimum level of seniority to be eligible to apply) and for secondary school teachers, based on contractual status (teachers had to have a contract to work more than 19 hours per week) (Santibañez, 2007).

Implementation

Basic implementation features

Duration

Four of the programmes - the ALI experiment in Mexico, ICS Teacher incentive programme in Kenya, META programme in Peru an Andra Pradesh Randomized Evaluation in India were implemented for a period of 2 to 3 years, with the Andra Pradesh Randomized Evaluation extended for another three years to a total of 5 years ((Behrman et al., 2012, Muralidharan and Sundararaman, 2008, Glewwe et al., 2010, Cueto et al., 2008). The three remaining programmes were ongoing at the time of evaluation (Gallego, 2008a, Santibañez, 2007, Duflo et al., 2012)

Scale

Most of the evaluated programmes were either experimental or small scale/pilot interventions. The Ali experiment in Mexico and Seva Mandir Teacher Incentive programme in India were implemented in less than 100 schools (20 and 57 respectively) (Behrman et al., 2012, Duflo et al., 2012), The ICS teacher incentive programme in Kenya, Andra Pradesh Randomized Evaluation and META programme in Peru were implemented in 100, 200 and 250 schools respectively (Glewwe et al., 2010, Muralidharan and Sundararaman, 2011, Cueto et al., 2008). Only two of the programmes were nationally-implemented programmes. In Chile, 90 per cent of all primary and secondary schools were eligible to participate in the SNED programme (Gallego, 2008a). In Mexico, 51 to 60 percent of all basic education teachers participated in the Carrera Magisterial at some point between 1991 and 2002 (Santibañez, 2007)

Implementation agency

The majority of the included programmes were implemented by the government, with a fair proportion implemented by NGOs. The Ali experiment and Carrera Magisterial programme in Mexico, SNED programme in Chile, and the META programme in Peru were implemented by the Ministry of Education (Behrman et al., 2012, Gallego, 2008a, Santibañez, 2007, Cueto et al., 2008). The Andra Pradesh Randomized Evaluation in India was implemented by the by the Azim Premji Foundation (a leading non-profit organization working to improve primary education in India) on behalf of the Government of Andhra Pradesh, with technical support from the World Bank (Muralidharan and Sundararaman, 2011). The ICS Teacher incentive programme was implemented by the dutch NGO International Christelijk Steufonds (Glewwe et al., 2010), and the Seva Mandir Teacher incentive programme was implemented by the Rajasthani NGO Seva Mandir (Duflo et al., 2012).

Programme implementation, delivery and take up

Several of the included evaluations and associated documents reveal aspects of programme implementation and delivery that may have acted as either barriers to or facilitators of programme take up, adherence and programme effectiveness. These factors are briefly outlined below.

Delays in information provision and lack of clarity about the intervention

Three studies mention that information about the intervention was not provided in a timely manner or was not sufficiently clear. Behrman and colleagues (2012) note that schools participating in the ALI program were not informed until well into the first semester of the first year about the intervention and that it took some time for the students to be informed about the programme. They hypothesise that this may account for the programme effect being larger in successive years when the programme was more established. Findings of in-depth interviews conducted with 792 teachers participating in the SNED programme in Chile found that some schools were unaware about the intervention and therefore did not apply. Other schools did not realise that schools were grouped according to region and socio-economic characteristics. Schools that had limited infrastructure and facilities, low baseline test scores or many “at-risk” students may thus have felt that the bonus was not achievable, and consequently did not apply (Gallego, 2008b). Similarly, Glewwe and colleagues (2010) suggests that in the first year of the ICS Teacher incentive programme in Kenya, some teachers have not fully understood all the implications of the incentive scheme. Anecdotal evidence from the first year ceremony suggested that teachers did not realise that students not taking the test or dropping out would reduce their chances to win the prizes. The authors also refer to reports from the field that indicate that teachers took the programme more seriously in the second year, after having seen it work in the first year.

Poor implementation fidelity and delivery

One study reports that the incentive programme was not implemented as intended and two studies note that the way the intervention was delivered made it difficult for participants to succeed. Santibanez (2007) notes that during the first few years of Carrera Magisterial in Mexico, the evaluation system was not fully functional and the wage promotions thus weren’t based on a formal assessment as intended. The author notes that while Carrera Magisterial was intended to reward teachers on merit, in practice, it resulted in an across-the-board salary increase for most of the teachers in the programme during these first few years. They confirm this with an analysis that shows that teacher test scores did not differ significantly between teachers that were promoted and those that were not, and highlight that promotion to higher levels seemed to be drive by non-merit factors such as seniority, education levels and geographic location rather than teacher performance (Santibañez, 2007).

This study also noted that the way the Carrera Magisterial programme was implemented made it difficult for a large number of teachers to succeed. The authors highlighted that not all teachers that enrolled in the programme were considered for promotion. Teachers had to fulfil six criteria, some of which (such as the requirement to undertake professional development) were difficult to meet. The authors do not have evidence about why this criterion was difficult to fulfil but speculate that this might have been due to a lack of teacher training opportunities in the country, or due to teacher's choices. The authors report that over time, the enrolment rate for the intervention declined considerably (Santibañez, 2007). Teachers participating in Chile's SNED programme reported that participation in the programme was time-consuming because the programme generated additional paperwork that the teachers had to completed, yet no extra time or meetings were dedicated to the programme. The authors also report results of an analysis that 38 per cent of the schools have never been awarded the SNED bonus. They hypothesise that one of the reasons for this is because teachers in some schools may not respond to the programme with higher effort (Gallego, 2008b, Contreras and Rau, 2012).

Lack of transparency about decision process used to award incentives

Two studies note that the tool used the measure teacher performance or decisions to award the incentives were not sufficiently transparent. Behrman and colleagues (2012) note that in the first year of the ALI experiment, there would have been considerable uncertainty about the difficulty and format of the exams that were used to determine bonuses, noting that this might account for limited effectiveness of the programme in the first year. In-depth interviews of teachers participating in the SNED programme in Chile found that teachers felt that the decision process about how bonuses are awarded was not clear, resulting in anxiety, expectations and frustration for those who did not receive it. As a result, some teachers reported to have given up on applying for the bonus in subsequent years (Gallego, 2008b).

Resistance and logistical problems with monitoring component

Two studies mention problems with the implementation of the teacher monitoring components. Interviewed teachers participating in the SNED programme in Chile disliked the idea of being assessed or monitored, even if there were incentives to do so (Gallego, 2008b). Cueto and colleagues (2008) report logistical difficulties in setting up the daily monitoring of classes and collection and processing of reports that needed to be submitted for payments. The authors do not specify what these were but they note that these problems discouraged continuation in the programme (Cueto et al., 2008).

Buy in from participants

Three studies report on support for the programme and resulting increased effort by participating teachers. Interviews conducted during the second year of the ICS Teacher incentive programme found that all teachers supported the use of incentives to motivate teachers and most teachers reported on changes in school activities and attitudes as a result of the programme. More than 80 per cent of teachers felt the prizes were justly awarded in the first year, three quarters of the teachers reported that they increase homework assignment as a result of the programme, almost 90 per cent reported conducting more test preparation activities, and almost 70 per cent felt that cooperation among teachers increased (Glewwe et al., 2010). Results of interviews with teachers participating in the SNED programme likewise suggest that none of the teachers objected to the idea of incentivising performance. Rather, they were considered highly important motivators that not only increase the self-esteem of teachers but improve commitment and quality of work and optimize the internal climate in the colleges (Gallego, 2008b). The Andra Pradesh Randomised evaluation in India also conducted a survey to analyse stakeholder reactions to the programme and found that teachers there was a general acceptance of the programme and the idea of performance-based pay. The study finds that almost all teachers in incentive schools (95%) had a favourable opinion about the programme and nearly 75 per cent of teachers in incentive schools reported that their motivation increased as a result of the programme. The study also finds that over three quarters of teachers were in favour of the idea of having at least some performance-based pay, and 20 per cent of teachers were in favour of at least 20 per cent of teacher's pay being determined by performance (Muralidharan and Sundararaman, 2008)

Additional interventions

Four evaluations mention that treatment and/or comparison group participants received additional interventions. These may reduce the comparability and generalisability of the study findings. To the extent that participation in these programmes was not randomly distributed among the treatment and comparison groups, these additional interventions may also bias the treatment effect estimates.

Behrman and colleagues (2012) report that between 11 and 13 per cent of the treatment and comparison group students in the ALI experiment in Mexico also received the Oportunidades cash transfer and/or were part of another scholarship programme). The authors note that this may account for the limited effects of the intervention on drop out rates since the Oportunidades cash transfer is conditional on student attendance. The authors note that the distribution of cash transfer recipients did not differ between treatment and control group (Behrman et al., 2012). Contreras and Rau (2012) mention that one year after the launch of the SNED programme in Chile, a programme called Jornada Escolar Completa was implemented in about 19 per cent of the schools. The programme increased the school day from a half day to a full day in these schools (Contreras and Rau, 2012). Glewwe and colleagues (2003 and 2010) note that some treatment and control group schools included in the evaluation of the ICS Teacher incentive programme in Kenya had also received textbooks and modest grants. Some treatment group schools also received a deworming programme and some comparison group schools received a pre-school support programme (Glewwe et al., 2010, Glewwe et al., 2003). Duflo and colleagues (2012) notes that treatment and control group teachers included in the evaluation of the Seva Mandir teacher incentive programme in India attended bi-monthly teacher training sessions. In these sessions, the implementing agency Seva Mandir reviewed the attendance rates and berated teachers that were seen to have a poor attendance record (Duflo et al., 2012).

Outcomes along the causal chain as measured by counterfactual analysis

Four studies report additional process and intermediate outcomes that did not meet the inclusion criteria of the review, but may be useful in helping to interpret the findings of the review.

Two studies measured student behaviour outcomes. The evaluation of the ALI programme in Mexico measured student behaviour in terms of average hours per week of maths study, average hours per week studying non-maths subjects, fraction of students paying attention more than 75 per cent of time, never or almost never texting while doing homework, never or almost never watching TV while doing homework, helping classmates and reporting putting much effort into studying. The authors of the study do not find any statistically significant effects of the intervention on student behaviour, with the exception of the average number of hours that 10th grade students spent studying maths which seems to be 0.138 fewer hours on average compared to the control group (standard error: 0.091) (Behrman et al., 2012). The Andra Pradesh Randomized Evaluation evaluated the effect of the teacher incentive on the proportion of students using textbooks and asking questions. The study finds that there was no statistically significant difference in the behaviour of treatment group and control group students (Muralidharan and Sundararaman, 2008)

The evaluation of the SNED programme in Chile evaluated whether the programme affected the characteristics of teachers in the treatment schools such as whether the teacher had any qualification, a university or post-university degree, attended any training course, worked in another college, the number of years spent teaching in the same class and the seniority level in the same college. The authors find that that there were no differences between treatment and control group on most of these variables, but find that following the intervention, the percentage of teacher working in other colleges was lower in treatment schools compared to control schools (Gallego, 2008a)

The study of the SNED programme in Chile also finds that the programme was associated with a significant increase in the percentage of parents that were informed about the academic results of their children (Gallego, 2008a)

The Andra Pradesh Randomised evaluation in India conducted a household survey at the end of the fifth year of the programme to evaluate whether household adjusted their own inputs into their child's education as response to the programme. The authors do not find any statistically significant differences between treatment and control group in household expenditure on education, student time allocation and household perceptions of school quality, suggesting that parents did not adjust their education inputs as a result of the programme (Muralidharan, 2011)

Finally, the study measured the effect of the external measurement and monitoring on test score outcomes by comparing outcomes of the control group with a "pure" control group that was not externally monitored throughout the year. They find no statistically significant differences between the treatment and control group, indicating that the external monitoring aspect of the programme did not have a major effect on student performance. They hypothesise that this might be due to teachers being informed that specific school or teacher performance information would not be shared externally on an identifiable basis, so the external monitoring did not create incentives for teachers in the control group to perform better (Muralidharan and Sundararaman, 2008)

Other process and intermediate outcomes reported by studies

The authors of the evaluation of the Carrera Magisterial reported that there is evidence that teachers dedicated extra time to test preparation and hypothesise that this additional effort might have been partly responsible for the positive effects observed in some programs (Santibañez, 2007).

Context

The included studies report very little about potential policy-related or contextual barriers and enablers of programme effectiveness. The one noteworthy event was a teacher strike during the evaluation period of the META programme in Peru. However, the authors note that they have taken this into account in their analysis (Cueto et al., 2008).

Other contextual information provided in studies helps paint an understanding of the conditions in which the evaluated programmes were implemented.

Teacher accountability structures and norms

Three studies report details about existing teacher accountability structures and norms in the countries where the programmes were implemented.

Muralidharan and colleagues (2008) who conducted the Andra Pradesh Randomized Evaluation in India note that teacher unions in the country are strong and disciplinary action for performance is rare (Muralidharan and Sundararaman, 2008). Glewwe and colleagues (2010) who evaluated the ICS Teacher incentive programme likewise noted that unions are very strong and therefore it is difficult to fire teachers, that decisions to hire, fire and transfer teachers are made centrally by the Ministry of Education, and that although parent committees exist, these are usually weak. The authors note that sometimes parents provide gifts to teachers if schools do well in national exams, but most committees do not do so and they only appeal the authorities about teachers' poor behaviour only in extreme cases. However, the results of the national primary school leaving exam are front page news in Kenya and often posted in headmasters' offices, which may act as a form of incentive for teachers (Glewwe et al., 2010). Duflo and team (2012) who evaluated the Seva Mandir Teacher Incentive programme noted that firing of teachers by the implementing agency was rare. They also note that Udaipur, where the intervention was implemented, is sparsely populated. As a result, teacher attendance monitoring prior to the introduction of the programme was poor and absences high (Duflo et al., 2012).

Three studies comment on accountability norms in the countries where the programmes were implemented. Cueto (2008) comments that the culture of accountability in Peru is quite weak (Cueto et al., 2008). Muralidharan and colleagues (2011) note that best practice in teaching in India typically teaching to maximise test scores on high-stakes tests and that norms of teacher effort in India are quite low. (Muralidharan and Sundararaman, 2008). Both Muralidharan and colleagues (2011) and Duflo and colleagues cite a nationally representative survey that found that about a quarter per cent of teachers in India were absent from the classroom during normal school hours and that less than half of them were engaged in any teaching activity (Duflo et al., 2008, Muralidharan and Sundararaman, 2011).

Existing incentive structures do not reward performance

Four studies provide more details about existing incentive structures, noting that teacher incentive programmes were implemented in context where performance is not typically rewarded.

Cueto (2008) notes that teacher salaries in Peru are low compared to other countries and their value in terms of their purchasing power has halved since 1942. The government has increases salaries in 2003, 2004 and 2005, but none of these increases were linked to incentives. He also notes that until the META programme in Peru was implemented, the existing education policy did not permit performance-based incentives and did not facilitate teacher mobility (Cueto et al., 2008).

A study cited in Muralidharan and colleagues (2008) also reports that teachers reporting high levels of job satisfaction were more likely to be absent because they felt that there was no difference in professional outcomes for those who tried hard and those who shirk (Muralidharan and Sundararaman, 2008). Mizala and Urquiola (2013) comments that the Chilean education system rewards seniority, not performance (Mizala and Urquiola, 2013).

Glewwe (2010 similarly notes that teacher salaries in Kenya depend primarily on years of experience and teachers' education, without any opportunity for performance-based salary increases or promotions (Glewwe et al., 2010).

Student performance standards

Three studies comment on the existing student performance standards. Gallego notes that there was significant variation in the Chilean national test scores, especially in schools with similar socio-economic populations (Gallego, 2008b). Muralidharan and colleagues (2011) cite a recent Annual Status of Education Report which shows that more than half of 6 to 14 year old children living in rural areas could not read at the second-grade level, despite a high enrolment rate of 95 per cent (Muralidharan and Sundararaman, 2011). Cueto similarly notes that despite an increase in school enrolment to about 96 per cent in 2006, several reports indicated that there is a gap between what students should know and what they do know (Cueto et al., 2008).

Teacher hiring

In the following section, we report the results of the synthesis of qualitative findings from our included teacher hiring programmes, presented along the hypothesised programme theory which is used as an overall framework for structuring the synthesis. This first section looks at implementation features that may be associated with relative success and failure of the included teacher incentives programmes, followed by a review of the contextual barriers and facilitators to success.

In addition to the nine papers describing the eight included studies, we also identified eight additional documents that present qualitative, process and project information for six of the included programmes. All these documents were used to inform our qualitative synthesis of intervention and implementation features associated with interventions' relative success and failure.

Process and implementation

Teacher recruitment and allocation

Some teacher hiring programmes were able to fill the vacancies created by programmes relatively quickly (Duflo, Dupas and Kremer, 2007 and 2012; Muralidharan and Sundararaman, 2013).

In two cases, programmes were able to recruit teachers fairly quickly. Muralidharan and Sundararaman (2013) report that the majority of schools started the process of hiring additional contract teachers within a week and had made appointments within a month. A schools had completed their teacher recruitment after two months. Similarly, Duflo, Dupas and Kremer (2007 and 2012) report that most schools had appointed additional teachers within a few months of programme initiation.

Though vacancies were created and funding provided for teachers, posts were not always filled with some remaining open and others initially being filled but then teachers leaving without being replaced (Bold et al., 2013).

(Bold et al. (2013) report that not all schools managed to appoint teachers to the vacancies that the programme created, with only around 87% of schools able to hire a teacher at some point in the programme. Furthermore, some teachers were appointed but then left and the vacancies created were not always filled. As a result, the contract teacher vacancies created by the programme were filled in only around 70% of total programme months.

Teachers were not always allocated to the right schools (Chin, 2005).

Operation Blackboard was intended to allocate additional civil-service teachers to one-teacher schools. It was meant to reduce school closures and the incidence of multigrade teaching, but increase time devoted to child-centred learning as opposed to textbook learning. However, Chin (2005) reports that there was a large amount of teacher misallocation. State governments did not allocate teachers properly and as few as one in four were actually sent to a one teacher school.

Additional teachers enabled schools to reduce class sizes substantially (Duflo, Dupas and Kremer, 2007 and 2012)

Duflo, Dupas and Kremer, (2007 and 2012) report that class sizes in treatment groups were significantly reduced with the number of students reduced from 82 to 44 on average.

Reallocation of teachers within schools can undermine the desired effect of additional teachers (Bold et al., 2013).

Bold et al. (2013) found that the pilot of the Kenyan National Teacher Programme resulted in very small reductions in class sizes. Although teachers were typically placed in the correct grade, they were also asked to cover other grades. Furthermore, existing teachers were often reallocated within schools to spread the teaching load evenly (Bold et al., 2013).

The recruitment process for additional civil-service teachers bypassed existing waiting lists and resulted in delays, court cases and rushed recruitment (Dyer, 2012).

The Indian Operation Blackboard programme's teacher recruitment operation was not well conceived or implemented (Dyer, 2012). New posts were advertised but not offered to those already on waiting lists for teaching jobs. As a result of this mistake, there were court cases, delays and finally rushed recruitment. Furthermore, this process meant that the programme ended up hiring teachers that had previously been set aside because their credentials did not meet state standards. Presumably this meant that a lower standard of teacher was hired than might have been possible.

The introduction of new teacher recruitment processes disrupted existing ones and initially led to delays in recruitment (Estrada, 2013).

Estrada (2013) reports that the Mexican programme disrupted existing recruitment mechanisms. Some states already had competitive appointment systems and recruitment became delayed because teachers now could not apply for jobs in their states until they had first taken the national exam. In the majority of cases the first national exams went well, though there was an instance in one state of the exams being 'blocked' (not defined, but possible union action) with the result that they were postponed.

Additional materials and classroom equipment

Additional materials and classroom equipment was not delivered to schools or used (Chin, 2005)

Chin (2005) reports that even after the Operation Blackboard's implementation had been completed, many schools still lacked the equipment that should have been provided. Only 56% of a sample of participant schools that received visits had functioning teaching aids. Furthermore, it was common for them to be locked away rather than left out to be used, because teachers feared being blamed for loss or damage.

Civil-service teachers were able to 'pull rank' and obtain greater access to additional materials and physical classroom infrastructure compared to their contract teacher peers (Duflo, Dupas and Kremer, 2007 and 2012).

For the Extra Teacher Programme in Kenya, budget was made available for additional infrastructure and resources to facilitate the increased capacity necessary for splitting classes in two. Inputs such as classroom facilities were supposed to be equally allocated across sections. However, in practice, established civil-service teachers may have pulled rank and obtained better physical classroom infrastructure (Duflo, Dupas and Kremer, 2007 and 2012).

Programme implementation

NGO-led implementation of the National Teacher Programme pilot was better than a government-led version (Bold et al., 2013).

The pilot of the National Teacher Programme in Kenya was made up of two arms – one NGO-led and one government-led. Bold et al. (2013) report that the NGO-led version of the programme was far better implemented than the government version. In principle, district staff should have made routine visits to all schools. In practice, the Ministry's ability to call on district officials to carry out such tasks was limited. It is unclear whether the NGO was more successful, though World Vision employed permanent staff and paid 'volunteers' to monitor and implement such programme activities. Overall, Schools in the NGO treatment arm were 15 % more likely to have received a monitoring visit that in the government arm, although this was not significantly correlated with test scores. 'NGO managed schools' were also more successful than 'government managed schools' in filling teacher vacancies and had higher rates of teacher attendance. Overall, teacher presence in the classroom was higher in NGO managed schools than government managed schools (73 % versus 63 % respectively), though this was not significantly correlated with test scores.

Salary delays occurred and led to poorer performance (Bold et al., 2013).

Delays in paying the teachers recruited under the National Teacher Programme pilot in Kenya were also common. These were more severe in government managed schools than in NGO managed schools. Bold et al. (2013) also find that the salary delays were significantly and negatively correlated with test score improvement.

Teacher hiring processes can be prone to local capture with selected teachers often those with a connection to existing teachers (Bold et al., 2013).

Bold et al. (2013) report that the hiring process was compromised by nepotism for government managed schools. The percentage of contract teachers who were friends of existing teachers or School Management Committee members was two thirds in the government treatment arm, almost twice as high as in the NGO treatment arm. This difference was not significantly correlated with test scores.

There was little monitoring of implementation and it would not have been difficult for states to use the funds in unintended ways (Chin, 2005).

Chin (2005) notes that misappropriation of Operation Blackboard (OB) funds was certainly possible because there was little or no monitoring of implementation. Dryer (1999) goes further, reporting that the central government suspected that OB funds were temporarily diverted by states for their own schemes, which could account for delays in implementation.

Less than a third of prospective teachers taking Mexico's new standardised teacher exam passed (Estrada, 2013)

Estrada (2013) reports that a low proportion of candidates actually passed the standardised tests introduced by a Mexican programme designed to recruit only highly qualified teachers. In fact, less than a third of those taking the standardised test passed it.

Contract versus civil-service teachers

Contract teachers were worse paid and more likely to report that they received their pay on a very irregular basis compared to civil-service teachers. As a result, vacancies may have attracted less qualified candidates (Vegas and de Laat, 2003).

Vegas and de Laat (2003) report that contract teachers were more likely to report that they received their pay on a very irregular basis than their civil-servant peers. While 36 percent of contract teachers reported that they were paid 'very irregularly', only 8 percent of civil-service teachers reported the same.

Despite generally being more poorly remunerated than civil-service positions, contract teacher positions were still sought after (Duflo, Dupas and Kremer, 2007 and 2012; Bau & Das, 2014).

Bau and Das (2014) find that a large part of teachers felt discriminated against due to lower salaries and comparatively poorer non-monetised befits as well as in the assignment of certain responsibilities. They may have been less motivated as a result. However, contract teachers still received better pay than private school teachers and this may have been a mitigating factor (Bau & Das, 2014). In Kenya, despite low pay and lack of job security, Duflo, Dupas and Kremer (2007 and 2012) report that contract positions were still actively sought after by unemployed teachers.

Context

Unions have opposed new teacher hiring approaches, either because they threaten existing jobs or because they will mean lower pay, fewer privileges and less job security (Bold et al., 2013; De Pascual Pola, 2009).

Bold et al. (2013) report that there was extensive union action in opposition to the pilot of the National Teacher Programme. Union action was ongoing during the intervention and by June 2011, 4 months before the impact evaluation ended, the government acquiesced to union demands to absorb the contract teaching positions that the programme had created into civil service employment. The authors found a strong and significant relationship between union identification and changes in test scores. The difference in test scores between a teacher who felt represented by the union and a teacher who did not, accounts almost exactly for the difference in test scores between NGO and government treatment arms. This evidence is only suggestive but the authors argue that the results are consistent with the hypothesis that the national controversy surrounding the contract teacher scale-up spread to the contract teachers employed by the government in the experiment and negatively affected their performance, while teachers employed by the NGO were largely immune to the political struggle between the government and the teachers union. De Pascual Pola (2009) reports the Mexican programme selecting teachers according to their achievement in a competitive exam also provoked some opposition from unions. In order to appease the unions, some places are still awarded without using the new system.

Limited involvement of local stakeholders may result in low buy-in for an intervention (Dyer, 1999).

Dyer (1999) reports that Operation Blackboard in India received low buy-in from local stakeholders because they did not feel involved. States resented the scheme as they were given no involvement in the design and had the difficulties of implementation without any ownership (Dyer 1999: 56)

Providing materials

Process and implementation

Many schools did not receive text books (Sabarwal et al., 2014, Das et al., 2013).

Information on programme fidelity was only available for two of the evaluated programmes. Sabarwal et al. (2014)'s formal assessment shows some non-compliance in both treatment and control groups of the evaluated textbook programme; while 15 % of the treatment schools did not receive any textbooks at all, 46 % of control schools received textbooks from sources other than the programme (p.10). There were also reports of the textbooks not being delivered to schools, but rather to a other central location in the distric.

Sabarwal et al. (2014) furthermore find that even when books reach the schools, few books actually reach the students as many administators stored the books rather than distribute them among students. The authors find that number of core textbooks stored per students present in school is significantly higher in treatment than in control schools.

Sabarwal et al. (2014) note that the ongoing national decentralisation process may have contributed to somne of the issues with text book decentralisation. Both the District Education Officers and local council officers were active, ''leading to some confusion on the exact chain of command and roles and responsibilities of different agents'' (p.5). It is reported that ''process data from the impact evaluation reveals that none of the DEOs or Local Councils had a completely clear picture of who was responsible for book pick-up and distribution, official signatories, or monitoring of textbook delivery. Likewise with regard to the actual disbursement of books, knowledge was spotty and written records were rarely found'' (p.18).

In addition, books were delivered directly to only 20 % of treatment schools and as many as 25 % of the head teachers claimed to have paid textbook retrieval from personal resources without reimbursement (Das et al., 2013).

*Half of the grant was spent on text books, the remainder on classroom construction (*Glewwe et al., 2007, 2009).

Glewwe et al. (2007, 2009) note that grant schools spent about half of the grant on textbooks. The remainder of the grant was mostly spent on classroom construction. Since classroom construction takes time, the effects evaluated are likely to *''largely reflect textbook purchases''* (2007:24-25).

The majority of the grant was spent on stationary, classroom materials and practice materials (Das et al., 2013).

Das et al. (2013, table 2) report that *''the majority of the grant money was spent on student stationery, such as notebooks and writing materials (over 40 percent); classroom materials, such as charts (around 25 percent); and practice materials, such as workbooks and exercise books (around 20 percent)''* (pp. 35-36). Around 10 percent was furthermore spent on durable items such as plates, cups, and schoolbags in the first year.

Spending on text books was very low as these are provided by the government at no cost. The authors also report that spending patterns across various categories were almost identical between the first and second years of the project. This rules out explanations based on diminishing returns to the items procured or the durable nature of school materials (e.g. that the grants in the first year were spent by schools on items that households cannot substitute for, while in the second year, the grants were spent on more substitutable items).

Parents adjusted their contribution to school inputs in response to grant program for schools to purchase materials (Das et al., 2013, Glewwe et al., 2007, 2009).

As mentioned above, Das et al. (2013) evaluate a two-year grant programme in which the first year of educational materials provided through the school was unanticipated by parents, while the second year was anticipated. They hypothesise that, in the second year, households are likely to re-align their spending patterns in response to the anticipated inputs to the extent that these behavioural responses might mediate the impact of educational inputs. The authors report that the intervention was taken up as expected in the sense that parents did indeed expect the grant and adjusted their spending accordingly.

Using qualitative evidence from field reports, the authors firstly report that *''in many cases, parents were aware of the grant program, and waited to see what materials the school would buy with the grant before incurring their own expenditures on materials*'' (p.44). Furthermore, household interviews revealed that *"two months into the school year, most parents had not bought the materials that they thought would be provided by the school"* (Das et al.,. 2013: 44). In addition, the authors findd that household education spending in treatment schools does not change in the first year (relative to spending in the control schools) but that is significantly lower (at the 1 % level) in the second year (table 3, p. 38). They furthermore note that *"for each dollar provided to treatment schools in second year, household spending declines by 0.76 dollars".* The authors thus find ''considerable crowding out of school grant by households in second year''. Consistent with this are lower test scores in the second year compared to the first year of the grant programme.

Glewwe et al. (2007, 2009) also report behavioural change among some of the smaller programme schools. They note that the data from the household questionnaire and school committee questionnaire suggests that the SAP funding lead to a decline in harambee fundraising (which focuses on classroom construction) in small schools. This crowding out could be a possible explanation for little effect of textbooks on average scores (p.26).

Access and use of materials

Four studies provided information on both teachers and students' access to and use of the provided by materials.

About a third of students still reported they did not have access to textbooks (Glewwe et al., 2007, 2009).

Glewwe et al. (2007, 2009) administered a student questionnaire to shed light on students' use of the textbooks provided and purchased through the intervention. Students in textbook schools report better access to school-owned textbooks in textbooks in grade-subject combinations that received textbooks. Sixty-two per cent of textbook school students in year 2 report having access to a school-owned book in class for the grade-subject combinations that received textbooks, compared to 8% of students in the 50-school comparison group (Table 3, Columns 1 and 2).

By year 3, the difference had narrowed; 72% of students in textbook schools were issued textbooks, versus 28% in comparison schools'' (Glewwe et al., 2007:10). More students in textbook schools, moreover report that they could take textbooks home ("More than half of students in textbooks schools report that they can take home school texts on subjects for which textbooks were given (Table 3, cols. 3 and 4), compared to less than 10% of comparison school students" p. 10).

Books were kept in storage and not distributed to students (Sabarwal et al., 2014).

Sabarwal et al. (2014) furthermore find that even when books reach the schools, few books actually reach the students as many administators stored the books rather than distribute them among students. The authors find that number of core textbooks stored per students present in school is significantly higher in treatment than in control schools, and textbooks per pupil in classrooms were not significantly higher in treatment schools than in control schools (Sabarwal et al., 2014, table 11). There was no significant increase in the likelihood of students in treatment school using a textbook for a core subject, nor any impact on any of the more direct indicators of textbook use (table 9, p. 29).

There was an increase in suppy of learning materials (Das et al., 2013)

Das et al. (2013) preclude storage in the grant progamme they evaluate. They note that schools reported the same spending patterns for both two years of the programme, suggesting that storage was limited. In addition, the grant led to a near for one increase in learning materials in the first year (p. 44).

Teachers reported high levels of use and familiarity with the flipcharts (Glewwe et al., 2000, 2004).

Glewwe et al. (2000, 2004) first conducted interviews with 82 grade 7 and 8 teachers in flip charts subjects at 21 of the treatment schools. The authors report that even though teachers were told their answers ''would not affect future aid to their school, the teachers may have nonetheless felt an incentive to bias their usage estimates upward. Yet over ninety percent of the teachers gave specific answers to questions that required some experience using the charts'' (Glewwe et al., 2000, 2004:15)

Following these interviews, the authors report that most teachers were aware of the flip charts (98 %), reported to have used them (91 %) and to have found them useful (92 %). These 92 % of the teachers reported the average chart had been used on 10-20 per cent of the schools days in the year the research was conducted (1998). The authors note that: ''Given that the charts were shared between grades 6-8 at least, this represents reasonably high utilization of the charts'' (p.15). In no cases had the flip charts been lost or stolen. The authors additionally conducted a flip chart use survey. This survey revealed that charts were used an average of 13 days per 75-day terms in grade 9 comparted to 7 days each in grade 6 and 7. Furthermore, thirty per cent of grade 7 and 8 teachers reported that the charts were relatively more helpful to low-performing students. Finally, it is noted that, since grade 7 and 9 teachers had priority over the use of the charts, they accounted for roughly 60-75 per cent of the total use (based on a survey in which teachers reported the number of times they had used the charts', p. 3-4). The authors note that ''one potential hypothesis for the low estimated effect of flip charts is that the charts would have been more useful in lower grades'' (Glewwe et al., 2000, 2004:15)

There was an increase in teachers' use of textbooks, but this did not reach more than sixty-two per cent at its peak (Glewwe et al., 2007).

Classroom observation data was collected to see whether the progamme affected pedagogy. Textbooks were used in 62.2% of the classroom observation sessions in the textbook schools, 45.8% of the observation sessions in the comparison schools, stat signif diff at the 5% level). By year 3 this difference was smaller (45.6% in textbook schools vs. 37.4% in comparison schools) and not statistically significant'' (Glewwe et al., 2007:10). There was also an increase in teacher presence in the classroom. Finally, there was a small but significant increase in homework assignment in the schools that received textbooks.

Glewwe et al. (2007, 2009) report some changes in the design of the grant programme during the 18-month period they evaluated it. The authors report that there were some changes in the curriculum and new editions of textbooks appeared at the start of year two. It is furthermore notes that ''the changes were modest, but some teachers may not have wanted pupils to use a version of the text older than the one they were teaching from, so differences across schools in effective numbers of textbooks may have declined'' (Glewwe et al 2009: 117).

Sabarwal et al. (2014) find significant positive impacts of the programme on the likelihood of teachers having a lesson plan (table 5, p. 2014) and the likelihood of teachers teaching (table 7, p. 27). They additionally and surprisingly note that parents of children in treatment schools were substantially less likely to complain to head teachers about the lack of books (table 4, p. 24). These outcomes are surprising considering the few books delivered to the treatment schools. The authors note that these outcomes are, however, ''consistent with teachers in treated schools receiving textbooks. Because the number of teachers in a school is much lower than the number of students, distributing books to teachers will still leave enough books in storage for future "smoothing'' (Sabarwal et al., 2014:18).

The programme had a small impact on teachers encouraging the use of textbooks (Sabarwal et al., 2014)

It is not surprising then that, while the programme had a small impact on teachers encouraging the use of textbooks, there was no significant increase in the likelihood of students in treatment school using a textbook for a core subject, nor any impact on any of the more direct indicators of textbook use Table 10 page 30 (Sabarwal et al., 2014)

Das et al. (2013:46) find that ''there are not differences in teacher absence or teaching activity across treatment and control groups''

Das et al. (2013) discuss the ideas of a 'Gift Exchange Effect', where there is a drop in impact after the start of the programme because enthusiasm or input levels off. *''if such a "gift exchange" or "novelty" idea was empirically relevant, we should expect similar patterns to be present in the other experiments conducted in the same setting, with considerably higher impact when programs start, but then dropping off to no impact when schools get habituated to the programs. We find that this is not the case''* (Das et al., 2013:46).

The text books, which were in English (the third language of most students) might have been too difficult for most students (Glewwe et al., 2007, 2009).

Glewwe et al. (2007, 2009) suggest the textbook might have been too difficult for students. All textbooks were written in English, which is the third language for most students. Fifty of the 100 sample schools were randomly selected for visits in year 4, and the median students in grades 3-8 were asked to read the English books provided by the programme. Results show that the median students in lower grades (3-4) had difficulties reading the books. In higher grades (5-8) reading difficulties were less common. Nonetheless, even if students could read the words, many may have had difficulties using a text book in their third language[165]. Evidence from the grant programme lends further support to the hypothesis that textbooks were best suited for the strongest students[166] : *"estimates based on the 25 schools given grants offer further support for the proposition that textbooks are most useful for students with strong preparation. First, an interaction term between pretest scores and the program variable (Table 12, column 3) was positive and statistically significant at the 1 percent level. Second, estimates similar to those in Table 9 (not shown) show significant impacts for quintile 5 (5 percent level in level regressions and 10 percent in differenced regressions) but not for the other four quintiles"* (Glewwe et al., 2009:129-130).

Teacher hiring

In the following section, we report the results of the synthesis of qualitative findings from our included teacher hiring programmes, presented along the hypothesised programme theory which is used as an overall framework for structuring the synthesis. This first section looks at implementation features that may be associated with relative success and failure of the included teacher incentives programmes, followed by a review of the contextual barriers and facilitators to success.

In addition to the nine papers describing the eight included studies, we also identified eight additional documents that present qualitative, process and project information for six of the included programmes. All these documents were used to inform our qualitative synthesis of intervention and implementation features associated with interventions' relative success and failure.

Process and implementation

Teacher recruitment and allocation

Some teacher hiring programmes were able to fill the vacancies created by programmes relatively quickly (Duflo, Dupas and Kremer, 2007 and 2012; Muralidharan and Sundararaman, 2013).

In two cases, programmes were able to recruit teachers fairly quickly. Muralidharan and Sundararaman (2013) report that the majority of schools started the process of hiring additional contract teachers within a week and had made appointments within a month. A schools had completed their teacher recruitment after two months. Similarly, Duflo, Dupas and Kremer (2007 and 2012) report that most schools had appointed additional teachers within a few months of programme initiation.

Though vacancies were created and funding provided for teachers, posts were not always filled with some remaining open and others initially being filled but then teachers leaving without being replaced (Bold et al., 2013).

(Bold et al. (2013) report that not all schools managed to appoint teachers to the vacancies that the programme created, with only around 87% of schools able to hire a teacher at some point in the programme. Furthermore, some teachers were appointed but then left and the

165 *"In grade 3, only 16% of the median students could read the grade 3 English textbook, and only 28% of the grade 4 median students could read their English textbooks (Table 11). Difficulty literally reading the textbook is less common in upper grades – the figures are 67% for grade 5 and over 90% for grades 6-8. Yet even if students can read the words in the textbooks, many may have difficulty effectively using a text book in their third language (Table 11 columns 2-4)"* (Glewwe et al., 2007: 24)

166 "Strictly speaking, this is not a fully independent trial since the Group 2 schools are control schools for evaluating year 1 outcomes, and analyses of the year 2 data for grants and for textbooks use the same comparison schools (50-school comparison group), but it is independent in the sense that the 25 textbooks schools are excluded from the grants sample" (Glewwe et al., 2009:129)

vacancies created were not always filled. As a result, the contract teacher vacancies created by the programme were filled in only around 70% of total programme months.

Teachers were not always allocated to the right schools (Chin, 2005).

Operation Blackboard was intended to allocate additional civil-service teachers to one-teacher schools. It was meant to reduce school closures and the incidence of multigrade teaching, but increase time devoted to child-centred learning as opposed to textbook learning. However, Chin (2005) reports that there was a large amount of teacher misallocation. State governments did not allocate teachers properly and as few as one in four were actually sent to a one teacher school.

Additional teachers enabled schools to reduce class sizes substantially (Duflo, Dupas and Kremer, 2007 and 2012)

Duflo, Dupas and Kremer, (2007 and 2012) report that class sizes in treatment groups were significantly reduced with the number of students reduced from 82 to 44 on average.

Reallocation of teachers within schools can undermine the desired effect of additional teachers (Bold et al., 2013).

Bold et al. (2013) found that the pilot of the Kenyan National Teacher Programme resulted in very small reductions in class sizes. Although teachers were typically placed in the correct grade, they were also asked to cover other grades. Furthermore, existing teachers were often reallocated within schools to spread the teaching load evenly (Bold et al., 2013).

The recruitment process for additional civil-service teachers bypassed existing waiting lists and resulted in delays, court cases and rushed recruitment (Dyer, 2012).

The Indian Operation Blackboard programme's teacher recruitment operation was not well conceived or implemented (Dyer, 2012). New posts were advertised but not offered to those already on waiting lists for teaching jobs. As a result of this mistake, there were court cases, delays and finally rushed recruitment. Furthermore, this process meant that the programme ended up hiring teachers that had previously been set aside because their credentials did not meet state standards. Presumably this meant that a lower standard of teacher was hired than might have been possible.

The introduction of new teacher recruitment processes disrupted existing ones and initially led to delays in recruitment (Estrada, 2013).

Estrada (2013) reports that the Mexican programme disrupted existing recruitment mechanisms. Some states already had competitive appointment systems and recruitment became delayed because teachers now could not apply for jobs in their states until they had first taken the national exam. In the majority of cases the first national exams went well, though there was an instance in one state of the exams being 'blocked' (not defined, but possible union action) with the result that they were postponed.

Additional materials and classroom equipment

Additional materials and classroom equipment was not delivered to schools or used (Chin, 2005)

Chin (2005) reports that even after the Operation Blackboard's implementation had been completed, many schools still lacked the equipment that should have been provided. Only 56% of a sample of participant schools that received visits had functioning teaching aids. Furthermore, it was common for them to be locked away rather than left out to be used, because teachers feared being blamed for loss or damage.

Civil-service teachers were able to 'pull rank' and obtain greater access to additional materials and physical classroom infrastructure compared to their contract teacher peers (Duflo, Dupas and Kremer, 2007 and 2012).

For the Extra Teacher Programme in Kenya, budget was made available for additional infrastructure and resources to facilitate the increased capacity necessary for splitting classes in two. Inputs such as classroom facilities were supposed to be equally allocated across sections. However, in practice, established civil-service teachers may have pulled rank and obtained better physical classroom infrastructure (Duflo, Dupas and Kremer, 2007 and 2012).

Programme implementation

NGO-led implementation of the National Teacher Programme pilot was better than a government-led version (Bold et al., 2013).

The pilot of the National Teacher Programme in Kenya was made up of two arms – one NGO-led and one government-led. Bold et al. (2013) report that the NGO-led version of the programme was far better implemented than the government version. In principle, district staff should have made routine visits to all schools. In practice, the Ministry's ability to call on district officials to carry out such tasks was limited. It is unclear whether the NGO was more successful, though World Vision employed permanent staff and paid 'volunteers' to monitor and implement such programme activities. Overall, Schools in the NGO treatment arm were 15 % more likely to have received a monitoring visit that in the government arm, although this was not significantly correlated with test scores. 'NGO managed schools' were also more successful than 'government managed schools' in filling teacher vacancies and had higher rates of teacher attendance. Overall, teacher presence in the classroom was higher in NGO managed schools than government managed schools (73 % versus 63 % respectively), though this was not significantly correlated with test scores.

Salary delays occurred and led to poorer performance (Bold et al., 2013).

Delays in paying the teachers recruited under the National Teacher Programme pilot in Kenya were also common. These were more severe in government managed schools than in NGO managed schools. Bold et al. (2013) also find that the salary delays were significantly and negatively correlated with test score improvement.

Teacher hiring processes can be prone to local capture with selected teachers often those with a connection to existing teachers (Bold et al., 2013).

Bold et al. (2013) report that the hiring process was compromised by nepotism for government managed schools. The percentage of contract teachers who were friends of existing teachers or School Management Committee members was two thirds in the government treatment arm, almost twice as high as in the NGO treatment arm. This difference was not significantly correlated with test scores.

There was little monitoring of implementation and it would not have been difficult for states to use the funds in unintended ways (Chin, 2005).

Chin (2005) notes that misappropriation of Operation Blackboard (OB) funds was certainly possible because there was little or no monitoring of implementation. Dryer (1999) goes further, reporting that the central government suspected that OB funds were temporarily diverted by states for their own schemes, which could account for delays in implementation.

Less than a third of prospective teachers taking Mexico's new standardised teacher exam passed (Estrada, 2013)

Estrada (2013) reports that a low proportion of candidates actually passed the standardised tests introduced by a Mexican programme designed to recruit only highly qualified teachers. In fact, less than a third of those taking the standardised test passed it.

Contract versus civil-service teachers

Contract teachers were worse paid and more likely to report that they received their pay on a very irregular basis compared to civil-service teachers. As a result, vacancies may have attracted less qualified candidates (Vegas and de Laat, 2003).

Vegas and de Laat (2003) report that contract teachers were more likely to report that they received their pay on a very irregular basis than their civil-servant peers. While 36 percent of contract teachers reported that they were paid 'very irregularly', only 8 percent of civil-service teachers reported the same.

Despite generally being more poorly remunerated than civil-service positions, contract teacher positions were still sought after (Duflo, Dupas and Kremer, 2007 and 2012; Bau & Das, 2014).

Bau and Das (2014) find that a large part of teachers felt discriminated against due to lower salaries and comparatively poorer non-monetised befits as well as in the assignment of certain responsibilities. They may have been less motivated as a result. However, contract teachers still received better pay than private school teachers and this may have been a mitigating factor (Bau & Das, 2014). In Kenya, despite low pay and lack of job security, Duflo, Dupas and Kremer (2007 and 2012) report that contract positions were still actively sought after by unemployed teachers.

Context

Unions have opposed new teacher hiring approaches, either because they threaten existing jobs or because they will mean lower pay, fewer privileges and less job security (Bold et al., 2013; De Pascual Pola, 2009).

Bold et al. (2013) report that there was extensive union action in opposition to the pilot of the National Teacher Programme. Union action was ongoing during the intervention and by June 2011, 4 months before the impact evaluation ended, the government acquiesced to union demands to absorb the contract teaching positions that the programme had created into civil service employment. The authors found a strong and significant relationship between union identification and changes in test scores. The difference in test scores between a teacher who felt represented by the union and a teacher who did not, accounts almost exactly for the difference in test scores between NGO and government treatment arms. This evidence is only suggestive but the authors argue that the results are consistent with the hypothesis that the national controversy surrounding the contract teacher scale-up spread to the contract teachers employed by the government in the experiment and negatively affected their performance, while teachers employed by the NGO were largely immune to the political struggle between the government and the teachers union. De Pascual Pola (2009) reports the Mexican programme selecting teachers according to their achievement in a competitive exam also provoked some opposition from unions. In order to appease the unions, some places are still awarded without using the new system.

Limited involvement of local stakeholders may result in low buy-in for an intervention (Dyer, 1999).

Dyer (1999) reports that Operation Blackboard in India received low buy-in from local stakeholders because they did not feel involved. States resented the scheme as they were given no involvement in the design and had the difficulties of implementation without any ownership (Dyer 1999: 56)

System Level

School Based Management Interventions

In the following section, we report the results of the synthesis of qualitative findings from included SBM programs, presented using the hypothesised programme theory as an overall framework for structuring the synthesis. This first section looks at intervention and implementation features that may be associated with relative success and failure, followed by a summary of the contextual barriers and facilitators to success.

Design

Targeting

Some SBM programs targeted disadvantaged areas of the country or worst performing schools (Gertler et al. 2012, Bando et al. 2010, Skoufias et al. 2006, Khattri et al. 2006, Santibanez et al. 2014), whereas other programs targeted schools that were comparatively better off in terms of human and social capital (Yamauchi et al. 2014, Pradhan et al. 2014; Carneiro et al. 2015).

For instance AGE was first introduced in the poorest municipalities of the poorest 23 states of Mexico, as defined according to the marginality index developed by the National Population Council. Coverage was extended to disadvantaged schools in the eight remaining Mexican states in 1998 (Gertler et al. 2012 p. 69). The worst performing schools in these states were targeted using an index based on CONAPO's community 'marginality' index, teacher- student ratios, the number of students per school and repetition and failure rates.

Similarly, Bandos' study is a randomized experiment evaluating the AGE programme in the state of Veracruz that targeted telesecundarias schools (television based lower secondary schools) that rank last in the standardized tests (ENLACE) and international assessments (PISA). The PEC programme in Mexico also targeted disadvantaged schools in poor urban and rural areas. To identify disadvantaged schools, PEC used a poverty index that the Oportunidades programme and Mexico's National Population Commission (CONAPO) constructed (Skoufias et al. 2006, p. 2). Participation in the program is voluntary and to qualify for the program schools had to prepare a school improvement plans which makes entry into the programme more likely for the schools with greater capacity and resources. In the case of the PEC- FIDE Program state program coordinators are instructed to target school zones in high poverty areas and those with low test scores in ENLACE.The TEEP programme in the Philippines was implemented in provinces identified as the most socially depressed in the Social Reform Agenda. Part of the IE suggests that schools were chosen based on perceived strength of their capabilities (Khattri et al. 2006 p. 291). Under BESRA, SBM was scaled up to non- TEEP divisions after the completion of TEEP. In reality the schools with higher human capital received the SBM grants first as they had to design good school improvement plans to receive the grant (Yamauchi et al. 2014, pp 8- 9). This is also the case for the school grants program in Senegal. Through this program, every elementary school in Senegal could apply for funds for a specific school project, with the best proposals being selected through a competitive process. Thus, the selected schools may be better organized to put together a good grant application. Comparing this set of schools with a nationally representative sample of Senegalese households, they find that the study schools have fewer students adn are more likely yto have electricity than the average school in Senegal, but are similaron other measures (Carneiro et al. 2014, pp 4-5, 12).

The Pradhan et al. study targeted rural public primary schools in six districts in Central Java and Yogyakarta because there were few large education projects active in the area, enabling the results to be relatively free from the risk of contamination from other projects. In addition, conditions in the area "were hypothesized to be ripe for community engagement to flourish - the area is peaceful, has reasonably high levels of existing social capital, and schools are relatively well equipped (high levels of electricity, adequate number of teachers, etc.)" (Pradhan et al. 2014, p.19).

The remaining programmes/ studies do not report on specific targeting other than the type of school or region. Beasley et al. (2014) evaluate a pilot project conducted in public primary schools in two regions of Niger that included the introduction of school grants to the COGES programme as a way to increase the involvement of schools committees. Based on the results of the evaluation, grant provision was planned to be scaled- up nationwide (Beasley et al. 2014, p. 7). Participating and control schools were randomly selected from public schools in these two regions. The Whole School Development Programme in Gambia targeted lower basic public and government-aided schools in regions 2, 3, 4, and 6 of Gambia. The remaining regions were excluded from the programme as they were either too urban or had been previously used as a pilot of the programme. The SanAntonio study targeted public secondary schools located in Camarines Sur in the Bicol Region as it is the biggest school division and would ensure more convenience of access. The PSI programme was introduced into Sri Lankan schools on a phased- in basis starting with one education zone per province in 2006. By the end of 2010, the PSI programme was scaled up to cover the entire country.

Process and implementation

Uneven implementation between states, with divergence from targeting criteria and the timing and amount of school grants (Skoufias et al, Santibanez et al. 2014; Carneiro et al. 2015).

Nationwide programmes such as the PEC reform devolved decision- making powers to state governments and schools over decisions such as the definition of the target population, amounts of the school grants and technical support provided to PEC schools. Based on findings of interviews with stakeholders involved in the PEC programme, Skoufias et al. (2006) report that schools included in the first year of PEC were not among the neediest in the country, which was a pre- condition established in the guidelines of the programme. Unlike most states that provide PEC grants at the beginning of the school year, the state of Colima distributed them at the end of the school year. Other states changed the numbers of schools included in PEC (usually reducing the amount of the school grants). Similarly, in the case of the PEC- FIDE program, Santibanez et al. (2014) report that program targeting did not follow the guidelines consistently. They found PEC- FIDE schools in low poverty areas and/ or with ENLACE test scores above the average.

Carneiro et al. also report divergence from the timing the project was implemented across states. By the end of year 2 of the study projects in the North had been running 7.6 months longer than in the South. (Carneiro et al. 2015, p. 29)

Grant disbursement

Grants may not have been disbursed as intended, with delays to completion of grant disbursement reported in both Mexico and Niger (Blimpo et al. 2011, Skoufias et al).

Two studies report that grants were not disbursed to the schools as intended (Blimpo et al. 2015, Skoufias et al. 2006). The Blimpo et al.'s study in Gambia reports that the disbursement of grants in the grant- only schools was not completed after the first year. In addition, principals found the disbursement process time consuming as the grants were required to be approved by the regional director (Blimpo et al. 2015,5). In PEC schools in Mexico the timing and amount of school grants varied. In the state of Colima distributed them at the end of the school year. Other states changed the numbers of schools included in PEC (usually reducing the amount of the school grants) (Skoufias et al. 2006)

On the other hand, the Beasley et al. study in Niger reports a reasonably high compliance rate with the data indicating that 492 out of 500 schools received the exact amount allocated to them, and six others received the grant but not in the correct amount (Beasley et al. 2012, p. 55). However, the program was originally intended to last three years (with three cycles of grant disbursement) but due to issues with the financial transfer mechanism at the central level and the political coup in 2009, the study was terminated after only one year.

Some schools dropped out due to changes in leadership, conflict (among administrators, teacher and parents) and the work load that accounting for the grant money imposed (Murnane)

Two studies reported on the extent to which schools joined/ participated in the SBM programme: Although the expectation of the PEC programme was that all schools that joined PEC in the first year would participate for five years, Murnane et al. (2006) report on the number of schools that withdrew from PEC by year and type. 82 PEC1 schools dropped out of the program after one year, and larger number dropped out in subsequent years. Some of these schools subsequently rejoined PEC Reasons for dropping out of the programme include changes in leadership, conflict among participating administrators, teacher and parents and the dismay at the work load that accounting for the grant money imposed.

School stakeholder interaction and participation in SBM

A main assumption of the theory of change for SBM interventions is that school stakeholders and parents work together to put SBM into practice and that teachers respond positively to parent participation and monitoring.

Increased parental/ community involvement in school matters such as attending school and committee activities, checking of homework or financial contributions represent one pathway through which SBM may impact learning outcomes. Several studies report on measures of parental/ community participation in school activities as a result of the intervention such as parents' time devoted to school activities and students' homework or parents' financial or in- kind support to the schools.

Treatment schools showed a higher rate of adoption of the SBM concept compared to the control group including higher rates of establishment of school committees (Blimpo et al., Aturupane e tal. 2014).

Blimpo et al.'s study report on a broad range of indicators of community participation, school management and school characteristics focusing around the take- up of the WSD program such as the establishment of school/ community participation committees and the adoption and implementation of a parent teacher association (Blimpo et al. 2015, p. 29). The authors note that one- year post intervention there is a higher rate of establishment of School Management Committees as recommended by the School Management Manual (SMM) in WSD schools. For example, 84% of the WSD schools had set up a curriculum management committee whereas only 51% of the control schools did so. The committees in the control group are often different in nature and reflect the organization in place prior to this research. Only about one-third of the schools in each group have adopted and actually implemented the new PTA constitution, with a 3-percentage point edge in the WSD schools (Blimpo et al. 2015, pp 16- 17.)

Aturupane et al.'s study report that although the PSI programme had no impact on whether the principal had conducted a needs analysis or had a long- term plan, it increased the probability that the principal had implemented some kind of project without financial support from the central or provincial government, had prioritised the school's needs and had increased the probability that a SDC had been formed.

SBM did not lead to an increase in parents' engagement with schools (Bando et al. , Pradhan, Aturupane et al. 2014).

Bando assesses parental inputs using counter factual analysis of take- up and finds a lack of evidence on changes in parental investment (measured as parents' voluntary contributions and hours spent by parents in activities related to parents association). The author finds no effect of the program on parent's voluntary contribution and the time spent on parent association activities (effect estimates in table I.30 and I.31). There is some heterogeneity in this effect, with the lowest 71th percentile of parental participation increasing their time spent, and the highest 20th percentile reducing the time spent.

Similarly, Aturupane et al. examine parents' expenditure on the education of their children using counter factual analysis. They find that the program had no significant impact on parents' expenditure for grade 4 and grade 8 students.

The Pradhan et al. study constructed two summary indices that included variables on parent- level and community- level inputs to education. They included variables such as parental financial or in- kind donations to the school and parents' involvement in and support for education. Community level input included variables measuring community support for schools and school committees and village council's collaboration with schools and overall support for education in the villages (See Online Appendix, pp. 8- 11 for a list of variables included in the two indices).

Pradhan et al. find that the grant treatment arm increased awareness of school committees (Likelihood that parents know committees exist). The increased awareness however does not translate into increased interaction between parents and school committees and only internal meetings of the school committee are impacted positively by the grant. The training treatment arm increased in- kind contributions of parents to the school committee but decreases the likelihood that parents ever visit to observe the class. The authors find no effect on parents' engagement with schools (Pradhan et al. 2014, p. 124).

There was an increase in parents' engagement with schools (Beasley, Skoufias, Gertler, Santibanez et al. 2014)

Using data on changes between June 2002 and June 2004 from surveys of students in 505 PEC schools, Skoufias et al. report on a range of measures indicating whether school are moving towards managing schools more effectively (2006, pp 9-11). Overall, students in PEC schools report improved school infrastructure and security, increased parental participation in schools and in students' homework and some improved teaching practices but unchanged involvement of school principals.

Gertler et al. conducted a series of focus group sessions with parents and carried out a qualitative survey of school directors' perceptions in 115 randomly selected AGE schools. The focus groups and survey confirmed the perception that the AGE program had improved interaction and communication with school directors and teachers. Both parents and principals reported that the AGE increased parental participation in schools, made parents more demanding in terms of attention to their children's learning needs and teacher effort, and increased parental involvement with homework. Parents also perceived that AGE had a positive impact on teacher effort in terms of increased hours spent in schools (Gertler et al., 2012,p. 70)

Beasley et al, find that, on average, parents increased their contribution to the schools and their participation in school management, but they could not go so far to enforce rules on teacher attendance. They find no overall impact on teacher supervision or remedial action for teacher absence (2014, p. 24).

Qualitiatve analysis of the PEC- FIDE program revealed that parents increased motivation or a changed understanding of their role in the school. Those interviewed explained that under PEC- FIDE, parents participated more and that those who had not participated in the past were now involved (Santibanez et al. 2014)

Parents were willing to try to improve school quality by participating in school committees, but they were not able to enforce rules and do so (Beasley)

Beasley et al. also report on measures of parent participation divided into non- assertive actions (financial contributions, in- kind contributions, whether school committee is in charge of monitoring pupil attendance and sanctioning pupils for poor attendance) and assertive actions (whether school committee is in charge of collecting, spending fees, infrastructure and supplies etc.) The authors also report on actions taken in opposition of teachers such as whether the COGES is in charge of monitoring teacher's attendance and has taken remedial action against teachers for repeated absence.

They find that, on average, parents increased their contribution to the schools and their participation in school management, but they could not go so far to enforce rules on teacher attendance. For example, their analysis shows a 7% increase in the average number of meetings in the year following the grant and a 27% increase in the proportion of school committees in charge of collecting fees. They find no overall impact on teacher supervision or remedial action for teacher absence (Beasley et al., 2014, pp 23- 24). The authors conclude that while parents were willing to try to improve school quality by participating they were not able to do so. Possible reasons include limited capacity and information to make investments that are likely to improve quality. The authors therefore argue that in contexts combining parents with weak authority/ capacity and teachers who prefer a centralised education system parent participation in SBM does no present a good strategy (Beasley et al. 2014; p. 39).

Few school councils functioned as a collaborative planning or shared decision making tool and were limited to signing off on decisions made by the principal (Santibanez 2014; Aturupane et al. 2014).

School councils fulfilled an important monitoring function that encouraged transparency and ensured that resources would actually be spent in (Santibanez 2014)

Similarly, Santibanez et al. also find limits to the extent parents and school councils are able to influence decision making. Qualitative findings of PEC-FIDE implementation did not find strong evidence to suggest that school councils engaged in collaborative planning or shared decision making. Although parents reported increased motivation, commitment and participation in school councils, planning and spending decisions were done by the principal in most cases. The authors therefore conclude that SBM's positive results are likely to be driven by the immediate benefits of increased school- level spending and not necessarily by producing better or more inclusive governance structures. Even though school councils were not consistently active in decisions about how to use PEC FIDE funds, in almost all cases, the council was aware that the resources had arrived and how they would be or had been spent. Therefore, as argued by the authors, one key effect of school council involvement is to be able to monitor that resources reach the schools.

Limited decision making power of parents and a lack of collaborative decision making has also been observed in the case of the PSI program in Sri Lanka. A qualitative study in one of the program districts found that the selection mechanism of SDC members is not democratic in most of the schools (more than 60%), and principals influence the selection of members for SDC, and SDC members of most schools have been selected by the principal. The authors also note that the SDC meetings is also not been conducted in participatory manner in most of the schools (more than 70%). It herefore appears that most of the principals still perform the key role in decision making in the schools.

Teachers may feel resistant to SBM as they perceive it as undermining their authority (Beasley)

Treating teachers as allies and investing in teachers working conditions and training may reverse a negative reaction to SBM (Beasley; Carneiro et al.)

Alongside this evidence, two studies reported a decrease in teacher effort as a result of the SBM intervention. In Niger, Beasley et al.'s (2014) observe a decrease in teacher presence in response to the grant. Qualitative feedback suggested that teachers preferred a centralized government and disliked that the communities were in charge of the grant as it undermined their authority. Other factors contributing to decreased teacher presence might also be related to the fact that schools did not spend the grant on expenses supporting teacher. One- teacher schools in turn, which invested in the teachers' working conditions and/ or made some transfer to the teacher actually increased teacher attendance. This might be potential evidence that teachers' negative reaction to parent participation can be reversed when parents behave like allies.

Carneiro et al. report that the school grants decreased teacher turnover, particular in the South where the program significantly affected the amount of training they got and how likely they were to remain tn the school from one year to the next.

In Sri Lanka, Aturupane et al. find no significant impact of the program on teacher absence and a negative effect of teachers' assignment of homework. The authors therefore conclude that there is little apparent change in teacher behaviour. However, the program did increase the probability that the principal had implemented some kind of project without financial support from the central or provincial government, had prioritised the school's needs, and had increased the probability that a SDC had been formed. Overall, the authors conclude that the results suggest that principals were doing little to implement activities that should lead to increased school quality.

The school principal plays a key role in motivating stakeholders to participate in school governance (SanAntonio, 2008)

Based on data from interviews of 40 experimental participants (school heads, teachers, parents, community representatives), SanAntonio (2008) find that different factors relating to the interaction between the school stakeholders either hindered or facilitated successful implementation of the ASC in the Philippines. Factors that motivated stakeholders to participate included receiving support from the principal and being recognized for their contribution as well as cooperative and dedicated ASC members. In contrast, reported factors that hindered advisory school council members from being actively involved in ASC included principals who appeared to lack self- confidence, who were tactless, authoritarian and did not manifest skills in evaluating the merits of ideas presented by the council or indifference and lack of dedication of other ASC members (SanAntonio, 2008, pp 57-59).

While there were no effects on student achievement, SBM led to higher levels of self-empowerment, commitment to work for school improvement and trust in school authorities among stakeholders compared to the control group (SanAntonio: 2018, p. 51).

The Advisory school council intervention in the Philippines provided an opportunity for school stakeholders to discuss school management concerns in monthly meetings with the school principal.

One study (SanAntonio, 2008) reports on the perception of participants on the operational effectiveness of the intervention. Based on results of a questionnaire of participants in the advisory school council intervention in the Philippines, the author finds overall high levels of satisfaction hastened by the willingness of principals to share information and power with the ASC. Respondents reported different measures of ensuring accountability to their respective constituents including consulting opinion leaders regularly, taking advice from fellow parents, students and community leaders. 91.9% of the respondents of the survey claimed that the power and authority vested in the ASC was adequate, 53% reported that the efforts of the ASC contributed to improvements in the teaching/ learning (pp 55- 57).

Grants were focused on construction and other material inputs, rather than books, learning materials or teacher training (Bando et al., 2015, Beasley et al. 2014, Skoufias et al. 2006).

The Bando study finds that, overall, parents prioritise spending program funds to make infrastructure improvements (restrooms, patios, sidewalks and playgrounds) rather than books and other learning materials. The author argues that learning outcomes are not likely to respond to these changes. Similarly, based on a detailed questionnaire on grant arrival and spending of 85 schools, Beasley et al. find that the spending was not used on activities likely to improve learning (at least in short run -). One year after the treatment, they find a small improvement in the infrastructure index of schools, largely driven by increases in the number of classrooms and the construction of walls around the compound. The authors find no overall impact on the educational materials available at the schools and fees charged to parents (Beasley et al. 2014, pp 24- 25).

The most common use of funds was on material inputs such as construction and office supplies rather than extra lessons and material. In the case of the PEC programme in Mexico, schools were provided with a five year grant. In the first four years, the programme required schools to spend 80 percent of the grant on supplies, infrastructure and other physical goods. In the final year, only 50 percent of the grant were spent on such goods and much of the grant funds teachers training and development (Skoufias et al. 2006)

The major share of funds went to teaching materials and teacher training (Carnoy et al).

Using PDE documentation, Carnoy et al. find that about 30 percent of the PDE funds had been devoted to buying electronic equipment (TVs, CD=- Rom players etc.), another 25 percent of the average budget went to learning and teaching material. Teacher training made up about one quarter of total PME expenditures (roughly same distribution in 1999, 2000 and 2001).

Skoufias et al. (2006) undertook a reflexive comparison of changes in parent school relations, student satisfaction and teacher performance in PEC schools using data on changes between June 2002 and June 2004 from surveys of students in 505 PEC schools. Overall, students report improved school infrastructure and security, unchanged involvement of school principals, increased parental participation in school and in students; homework, some improved and some unchanged teaching practices and increased expectation by parents and students that students would complete advanced education (Skoufias et al, 2006, p. 10)

Schools that focused spending funds on human resources improvements showed larger positive effects on test scores compared to schools that emphasized spending on manuals and other education materials (Carneiro et al. 2015)

Based on counterfactual analysis, Carneiro et al. find a North- South difference in program impact with larger positive effects for schools in the South of the country, where projects tended to focus on training human resources (teaching and management) compared to the North, where priority was placed on the acquisition of school material (e.g. textbooks/ manuals). This is also reflected in the amounts schools reported the project spent on principals, teachers, the management committee, and the students. The authors therefore suggest that the latter type of investment is likely to be more effective than the former type of investment.

Some teachers and principals found the administrative work and time spent on community engagement burdensome, potentially taking away time spent on pedagogical responsibilities (Murnane, Khattri, Blimpo).

Several studies report on the additional bureaucratic burden to schools in implementing SBM as a result of increased administrative work or increased engagement with the community, thus raising questions about the longer- term viability. If teachers need to reallocate a significant amount of time to the implementation of different SBM components, they have less time to dedicate towards pedagogical responsibilities and teaching which might in turn adversely affect student learning.

Although a goal of the PEC program is that schools should benefit from increasing autonomy and that state authorities reduce administrative burdens imposed on schools, Murnane et al. note that many state coordinators reported that administrative paperwork and expenditure reports result in being more complex and time consuming every year that take the time of principals and teachers and divert them from working on implementing their school improvement plans, instructional improvement and building stronger relations with parents and teachers (Murnane et al. 2006 p. 28). Blimpo et al.(2015) note that principals found the disbursement process time consuming as the grants were required to be approved by the regional director. The implementation Completion Report for the TEEP project indicates that increasing community engagement demanded that principals and teachers spend considerable time on community relations in addition to their administrative and pedagogical responsibilities, a commitment that several were beginning to find burdensome (Khattri et al.2012, p. 290).

Contextual factors influencing SBM effectiveness

Encouraging parents and community participation in school- management through placing financial resources directly under parental control or through increasing their participation in school committees is stated as an aim of several included SBM programmes representing a potential pathway to increased school quality. However, as pointed out by several authors, parents' authority and capacity are an important prerequisite for parents to undertake these tasks and hold school stakeholders accountable.

The social capital and education of parents may influence their ability to hold the schools accountable and participate effectively in school management (Beasley, Blimpo)

Lack of experience, poor knowledge of committee memebers and insufficient leadership qualities of principals may all act as a barrier to poor implementation of SBM

For example, Beasley et al.'s study of the COGES programme in Niger find that although parents increased participation and responsibility over school management this did not translated into increased school quality. Some aspects of school management improved such as cooperation between school stakeholders but overall accountability did not change and spending was often non- educational, intended to make a profit (agricultural projects, school festival playgrounds etc.). However, in the case were school committees were educated or had experience in another community organization (used by the author as proxies for parent authority) parents increased the monitoring of teacher attendance in response to the grant.

Blimpo et al. (2015) also analysed the importance of baseline human capital, measured as adult literacy, in mediating the effect of the WSD programme in Gambia. The authors conclude that the WSD intervention is likely to improve learning outcomes in areas with high baseline human capital but could be counterproductive in areas where the basic human capital is very low. They also found a large disconnect between parents evaluation of the schools and the actual performance of the schools and argue that this disconnect may explain the inability of the parents to hold the schools accountable and participate effectively in the management.

Enhancing the school council's capacity to make good spending decisions through training may be essential to ensure improved student outcomes and promoting parental involvement are important (Santibanez et al. 2015)

Based on qualitative findings, Santibanez et al. argue that although cash grants do serve as the catalyser for establishing a functioning school council, parents, teachers and principals might need more help understanding what exactly they need to do to improve student outcomes. It was not evident in PEC- FIDE schools that parents, teacher had enough information (e.g. prior test scores, student learning diagnostic results, information on best practices related to school materials spending) to make informed decisions about how to best utilize the grant.

Qualitative findings based on interviews in one of the program districts (Colombo districts) indicate that overall, the characteristics of community participation in the areas of decision making through SDC, attendance at the meetings related to decision making, and control over financial resources has been changed very slightly as a consequence of the PSI. Moreover, it seems that the schools are unwilling to get involved of the external community members may also lead to lack of trust. Further barriers include (1) Majority of principals (more than 60%) indicate that lack of experience and poor knowledge of the SDC members' on school management as a big challenge. (2) Other SDC members; parents and teachers (more than 70%) argue about poor leadership qualities of the principals displayed in school management as a huge challenge in decision making. (3) According to majority of principals and teachers (more than 60%),

Engagement with existing community institutions in the planning of educational activities may be important in contexts where such institutions are powerful (Pradhan).

In addition, the extent to which school committees are able to engage community institutions in the planning of educational activities may prove important in contexts where the institution is powerful such as in the case of the village councils in Indonesia. They are a democratic organization that has the power to draft village legislation, approve the village budget and monitor the village government. Pradhan et al.'s study found that nearly two years after implementation, measures to reinforce existing school committees structure (grant & training interventions) demonstrated limited or no effects. In contrast, measures that foster linkages between the school committees and the village councils lead to greater engagement of education stakeholders and, in turn, learning. The authors argue that increased community support was the main mechanism through which learning outcomes were improved.

The effectiveness of SBM in improving education outcomes may also be influenced by contextual factors.

Capacity of state departments of education to provide support and training to schools implementing SBM may play an important factor in determining the effectiveness of the SBM (Murnane)

Murnane et al. (2006) find heterogeneity in the effect of SBM on student dropout rates based on the Human Refined Development Index (2000). Participation in PEC had its largest impact on schools in the *Medium Outcome* states whereas they find no impact on student outcomes in the *Low Outcome* states. The authors note that one possible explanation for the lack of strong impact on the group of low LDI states is that their departments of education lacked the capacity to provide strong support to schools enrolling in PEC.

Centralised, hierarchical education systems may present a challenge for implementing some SBM program goals such as increasing the involvement of parents and the community (Beasley et al., 2014; Murnane et al. 2006).

The education system in Niger has traditionally been very hierarchical and centralised with little room for local community participation prior to the establishment of school committees (COGES) (Beasley et al. 2014). In Mexico, decentralisation to the state- level jurisdiction was implemented in the early 1990 but critics argue that state governments reproduced the centralised bureaucratic model in the administration of their educational system. Murnane et al.(2006) therefore argue that the capacity to undertake new responsibilities such as providing technical and pedagogic support to PEC schools and reducing their administrative burden may vary among states.

Low teacher quality, reduced instructional time due to widespread double- shift schools, and teacher compensation may have been a barrier to the success of SBM (Blimpo et al).

Furthermore, there are a number of other dimensions inherent in the education sectors that may have acted as barriers to the success of the SBM programme. For example, Blimpo et al. (2015) report that in the case of Gambia there appear to be other binding contextual constraints that may mediate the impact of the WSD programme. Some of these include low teacher quality, reduced instructional time due to widespread double- shift schools, and teacher compensation (p. 29).

The program in Niger was intended to last three years but was terminated after one year due to a political coup (Beasley et al.).

Out of the included studies only one reported on significant events that affected the implementation of the intervention. Although the programme was originally intended to last three years, the study had to be terminated after only one year due to a political coup in 2009 and issues with the financial transfer mechanism at the central level (Beasley et al. 2014, p. 10).

Awareness sessions may have been essential in clarifying objectives of the program and overcoming scepticism by schools that the program would decrease government support and hand over power to parents (Aturupane et al. 2014)

There may also be initial resistance to SBM programs due to misunderstanding of the objectives and scepticism of schools that it would result in decrease of government support or grant too much power to parents. In the case of the PSI program in Sri Lanka, qualitative evidence found that awareness sessions proved useful to clarify objective of program. . According to respondents following the awareness sessions, these sessions have been helpful in addressing scepticism of some principals whether PSI will over empower parents by providing a thorough understanding of the intentions of PSI for democratic decision-making and ensuring accountability while continuing the government's support to schools. (Wehella, p. 59).

Community Based Monitoring Interventions

In the following section, we report the results of the synthesis of qualitative findings from our included programmes, presented using the hypothesised programme theory as an overall framework for structuring the synthesis. This first section looks at intervention and implementation features that may be associated with relative success and failure, followed by a review of the contextual barriers and facilitators to success.

Process and implementation

Uptake of the interventions

All of the fours studies (of eleven) that reported on the take up of the interventions among intended community participants reported positive findings.

Strong attendance/interest from the community in CBM meetings, including among minority groups (Banerjee et al.; Pandey et al. 2011; Lassibille et al. 2011)

Social accountability interventions generally encourage inclusiveness of the whole community, as described in World Bank's participation sourcebook (World Bank, 1996). Of the five programmes that worked at the community level, two of the studies reported that the programmes actively encouraged participation from all in intervention activities (Pandey et al. 2011; Banerjee et al. 2010) and three of the studies reported that community meetings to disseminate information were well attended and had a good attendance from women (Banerjee et al.; Pandey et al. 2011; Lassibille et al. 2011)

Banerjee et al.'s study in India reported that meetings to discuss the role of parents in monitoring were well attended, with all treatment villages holding at least meeting and having on average 108 participants, 95 per cent of whom were present during the entire meeting. Lower castes were well-represented and meetings were held in high and low caste neighbourhoods where necessary to improve participation. Facilitating teams "went from hamlet to hamlet within a village, making sure to cover "low-caste" hamlets, carrying out conversations about education in small and large groups (which enabled women to participate, for example) and inviting local people to take the lead" (Banerjee et al., 2010: 20).

There was particularly impressive uptake of the third intervention arm treatment of Banerjee et al.'s study which trained village volunteers in a simple technique for teaching to read, where volunteers in 55 of the 65 treatment villages started reading classes, which were on average attended by 8% of the children in these villages.

Lassibille et al. (2010) reported that accountability meetings were well attended with 64 participants on average, almost half of whom were women.

Good uptake by parents of tool to assess child learning in India (Pandey et al. 2011)

Pandey et al. (2011) found significant uptake of the intervention tool provided to assess the grade-specific learning levels of child: around 60% reported using the learning tool.

Schools or the community followed up on information provided as part of the intervention (Pandey et al. 2011; Mizala & Urquiola, 2013)

Pandey et al. 2011 reported that a large percent of participants in their focus groups had discussed the disseminated information on the role of the community in oversight of education with others in the village and the discussions had lasted at least a few weeks. Qualitative interviews of committee members found that 87 per cent of committee members reported discussing campaign information with others in the village. Seven five per cent of parents reported using the learning booklet provided to assess the learning levels of their children. Of these, 43 per cent had gone on to discuss it with the service provider or school committee and a much smaller number with the Chair of the committee. Thirty-five per cent of participants went on to bring up teacher attendance as a concern with the school or committee (either the VEC or PTA) or education official.

The study of the SNED programme in Chile found that between 80 and 90% of schools reported that they had engaged in dissemination activities following publication of ‘winner schools’, for example informing their parents' association, sending notes home, or raising the issue during PTA-type meetings (Mizala & Urquiola, 2013), although this information was self-reported.

Parents’ knowledge following information campaigns

Parent and school committee knowledge of their roles and responsibilities and the status of education are important process outcomes for obvious reasons; in order for parents and committees to participate in collective action towards schools, they must first be aware of the local education situation and know how they can act on it.

Limited improvement in parental and/or education committee knowledge of monitoring institutions following information campaigns

The two included studies from India reported findings on parental and/or education committee knowledge of monitoring institutions following information campaigns (Banerjee et al., 2010; Pandey et al., 2011). The findings from these studies were mixed.

In Banerjee’s study in India, across all three treatment arms only 2.6% of parents were more likely to know about the existence of VECs, which in total adds up to only 7% of parents in treatment villages there was no change in parent’s knowledge of Village Education Committee (VEC) existence or its role following any of the information campaigns or in VEC knowledge of their role. This is a surprising result given that there was strong participation in village meetings as previously discussed.

Pandey et al.’s study, also in India, found a small improvement in awareness of roles and responsibilities, but importantly, this was mainly for the VEC or Parent-Teacher Association (PTA) chairs, and for low-caste groups the impact was significantly lower. Interestingly, 73 per cent of a focus group undertaken with a selection of the community members in the treatment group said that more frequent meetings to disseminate information would make the intervention more effective, suggesting perhaps three meeting rounds was not enough to increase knowledge (Pandey et al., 2011).

Small increase in parents and school committee knowledge of the status of education in their village after the intervention compared to the control group (Banerjee et al. 2010; Andrabi et al. 2013)

Two studies (out of the ten) reported findings on parental and/or education committee knowledge of education status in the village, Banerjee et al.(2010) and Andrabi et al. (2013).

Banerjee’s study found that parents and VEC members were slightly more aware of the status of education in the village following the intervention.

Andrabi et al. (2013) found that in villages that received a report card, there was an increase in accurate perceptions of school quality.

Parent and school committee participation in school oversight and management

Parental and/or school committee participation in schools was measured in four of the eleven studies, covering four unique programmes.

Parental participation in schools did not increase as a result of the social accountability intervention (Banerjee et al. 2010; Zeitlin et al. 2012; Nguyen and Lassibille (2008); Pandey et al. 2011)

Banerjee et al. (2010) found no impact of any of the interventions on parental involvement in schools or involvement of the Village Education Committee (VEC) in monitoring of schools. Parents were no more likely to have visited the school or to have volunteered time and/or money in the treatment villages than the control villages. VECs in treatment villages were no more likely to have raised a concern.

Zeitlin et al. (2012) found no significant effect of the program on frequency of and attendance at Parent Teacher Association and School Management Committee meetings.

Nguyen and Lassibille (2008) reported that teachers in treated schools were no more likely to communicate with the parents on student matters. Similar to this finding, there was no improvement in terms of the school head teacher communication with teachers and with the community on school matters.

Pandey et al. (2011) did find some positive trends in school committee participation, mainly improvements in member participation in school visits and inspections, however this was not consistent across the five participation outcomes they measure and the impact was very uneven across groups.

Minority groups excluded from using information provided to them as part of CBM initiatives (Pandey et al. 2011)

Pandey et al. (2011) held separate meetings in low-caste and high-caste neighbourhoods or with women only where people were unwilling to come to the common meeting. However, they found both quantitative and qualitative information to suggest that lower castes were excluded from acting upon the information provided to them. In the campaigns in both Uttar Pradesh and Madhya Pradesh, there was an increase in awareness of school committee/ entitlements though not in participation of low caste members. Among lower caste groups, focus groups participants mentioned that people were afraid to talk to the teachers and that they more frequently found that they experienced anger from a teacher when they raised issues related to their child's learning compared to non-low caste groups. This suggests that there are additional barriers to participation among typically excluded groups.

Parent response to information campaigns: switching schools

The rate of switching between schools is an important consideration as it provides an indication of whether parents are demanding more from their current school as a result of more information or just 'switching' their children into better-performing schools. Three studies assessed this.

Parents did not switch their children into better quality schools as a result of improved information about school quality (Banerjee et al., 2010; Andrabi et al., 2013; Camargo et al., 2012).

Banerjee et al. (2010) find no evidence of parents moving their children to private schools, even though the private sector was very active in Uttar Pradesh (where the intervention took place). In the third treatment arm, the number of children dropping out of private/NGO schools actually increased, which the author suggest may be a result of parents may feel that feeling that the additional reading classes that are part of that treatment may be an adequate alternative.

Andrabi et al. (2013) report data for the number of children who switched schools or dropped out in the village as a fraction of children enrolled at baseline, and found no overall change in the switching rate in treatment villages.

Camargo et al. (2012) results suggested that there was no evidence of parents sorting their children into better schools following release of test scores. They assessed this by looking at important changes in student characteristics between treatment and control groups over time, which would be expected if parents had sorted their children into better performing schools. They do suggest that there may not have been enough time between release of scores and the follow up survey for parents to have switched their child into a different school.

Education sector response to information campaigns

Substantial changes in private school fees as a result of more information in Pakistan (Andrabi et al. 2013)

Andrabi et al. (2013) reported that following the introduction of village-level education report cards in Pakistan, there were substantial changes in school fees charged by private schools. Private schools decreased their annual fees relative to control village private schools by an average of Rs. 218 in response to the report card intervention, around twenty per cent of the baseline private schools fees.

Information campaign reduced leakage of funds from the education system but did not eliminate it in Uganda (Reinikka & Svensson, 2007)

Reinikka & Svensson (2007) report that on average schools only received on average 24 percent of the total yearly grant from the central government in 1995. In 2001, after the newspaper campaign reporting education grant had been running for four years, this had increased to more than 80 per cent. They associate this with exposure to the newspaper campaign, that is, the closer the school to a newspaper outlet, the larger the reduction in diversion of education funds. However, this still left 20 percent of schools not receiving their entitlements and about 30 percent of schools receive less than two-thirds of their entitlements.

Implementation fidelity and service quality

Very little information was presented on implementation fidelity. There was no suggestion from the impact evaluations that any of the nine included programmes or experiments were not implemented as intended, although this information may have been out of the scope of these impact evaluations. Nguyen and Lassibille's (2008) evaluation of the AGEMAD programme in Madagascar reported that implementation was successful, and all treatment schools received the tools they were supposed to and organised the meetings to discuss the school development plan as intended. Duflo et al. (2007) reported that School Committees in all 70 schools sampled for SBM received training on monitoring of their new teacher and contract renewal decisions as intended.

Contextual Barriers and Facilitators

As previously discussed, very little information was presented on the contextual barriers and facilitators to success of the evaluated social accountability interventions outside of the impact evaluations themselves. Nonetheless, the limited information can offer some insights for understanding our results.

Teacher responsiveness to community and structure and responsibilities of parent committees

One of the key assumptions of the theory of change for social accountability interventions is that teachers are responsive to community demands for improved performance. This is to some extent moderated by the responsibilities and powers of parent or community education organisations, for example, whether school committees or parent teacher associations have the power to reward or punish teacher performance, but also whether the context allows for genuine punishment or reward even if such power does exist.

Teacher incentive structures may limit the effectiveness of CBM initiatives (Pandey et al. 2011; Andrabi et al. 2013; Zeitlin et al. 2012)

Pandey et al.'s study in India (2011) was the only study that directly asked participants about their opinion on perceived barriers to the effectiveness of the information intervention. The common theme of these groups was the lack of responsiveness of the teachers and it was reported that raising concerns regarding learning was frequently met with a negative or angry response. Reported themes included "teacher is dominating", "it is difficult to talk to the teacher", "teacher does not listen", and "teacher does not care".

Alongside this qualitative evidence, authors in three studies (Pandey et al. 2011; Andrabi et al. 2013; Zeitlin et al. 2012) discuss anecdotally poor teacher incentives structures that limit their responsiveness to parent demands, and thus the power of accountability mechanisms.

Andrabi et al. (2013) reported that in Pakistan, teachers unions are strong and teachers are rarely penalised in any way for poor performance. The authors use the example of teacher absenteeism in public schools in the country as evidence of this, where rates of absenteeism are around 17 per cent even in pre-announced school visits (Andrabi et al., 2007).

In Uganda, Zeitlin et al. (2012) mention that teaching vacancies take time to fill in the country, particularly for senior positions, and therefore school committee members may not want to fire underperforming teachers even if they were inclined to do so.

Pandey et al.'s (2011) baseline survey in India found that certain types of teachers displayed lower effort than others, with civil service teachers having lower attendance and activity rates than contract teachers. In the states of Uttar Pradesh and Madhya Pradesh more educated teachers had lower attendance and/or activity rates. They suggest that this is because teachers from higher socioeconomic strata often have protection from local elite groups and less incentive to perform well.

Extent of power of the school committees may play a role in determining the effectiveness of CBM initiatives (Pandey et al. 2011; Banerjee et al. 2010)

There was generally little information in the included studies about the types and extent of power that school committees have. However, Pandey et al. (2011) suggest that the structure of the school committees may contribute to the success (or failure) in improving participation, specifically committee size and length of term. If committee members are elected for a year only, they may not have sufficient time to support a cause. He uses this as an explanation for the fact that in Uttar Pradesh, where Parent- Teacher Associations (PTAs) are small, village level bodies that are elected for 5 years, there were greater improvement in participation and awareness than in Madhya Pradesh, where committees are at the school-level, relatively large in size and elected for a year only. In addition, while in Madhya Pradesh school committees must verify teacher presence for a teacher to receive her salary, in Uttar Pradesh committees control the tenure of contract teachers only and in Karnataka committees neither have the power to track teacher attendance or control teacher tenure. They suggest that this may be why there was an impact on teacher effort indicators in Madhya Pradesh but none in Uttar Pradesh and Karnataka.

This finding is also reflected in the Banerjee study in India which also targeted the interventions at school committees in Uttar Pradesh. They discuss that the committees can petition for resources that the school is entitled to ask for, can complain about the teachers or the level of education in the schools, but, ultimately the decision on whether anything will be done about an issue is not in their hands.

Responsiveness of the education provider to demands for better education may vary by school provider type (Camargo et al. 2012; Andrabi et al. 2013)

There is some limited evidence to suggest that the responsiveness of the education providers may vary by school type.

Three of the eight programmes covered both public and private schools (the campaigns in Brazil, Pakistan and Chile). The studies in Brazil (Camargo et al., 2012) and Pakistan (Adnrabi et al. 2013) found far larger improvements in student test scores in private schools following publication of information on school performance.

Camargo et al. (2012) suggest that this may be because school administrators in private schools face market incentives to keep a certain level of reputation, and are thus more responsiveness to publication of negative test scores. They however did not have the data to explain whether it mainly due these market pressures on school administrators or due to improvements in student's effort.

Similarly, Andrabi et al. found large improvements in student test scores for children in initially poorly performing private schools only, with only very small improvements in public schools. They present data to suggest that this was a result of schools changing their investments, with no accompanying (visible) improvement in household investments into their child's education.

Parents' human and/ or social capital may moderate the effectiveness of information campaigns (Andrabi et al. 2013; Pandey et al. 2011; Lassibille et al. 2013)

High rates of illiteracy among parents is a reoccurring context across the included studies, although it is not consistently reported by authors as a barrier or facilitator. Andrabi et al. (2013) report that in Pakistan, high rates of illiteracy among parents (37.3% in sample villages) may be leading them to get more involved in making demands of schools as they are less able to increase their involvement in their children's education at home.

Pandey et al.'s evaluation of an information campaign in three states in India found greater impacts on student achievement in villages with low literacy rates. They explain this as being because villages with more illiterate parents should have a greater demand for schooling.

On the other hand, Lassibille et al. (2013) suggests that wealthier and more literate parents are better able to use the information provided by the report cards, and, presumably, better able to monitor school activities.

The quality or capacity of teachers is often an underlying issue, but the intervention does not directly address supply issues (Lassibille et al. 2010; Pandey et al. 2011)

Existing capabilities of teachers is a potential barrier to improvements in education outcomes from social accountability interventions, as these interventions do not address underlying shortcomings in teacher quality or capacity. There is limited discussion in two of the included studies to suggest this may be the case.

The evaluation of the AGEMAD initiative in Madagascar (Lassibille et al. 2010) found small improvements in children's performance in Malagasy and Maths, but no improvements in French test scores. The authors attribute this to the fact that teachers in Madagascar often do not master French well and so the standard of teaching in this subject is lower.

Pandey et al. (2011)'s study in India also finds variations in impacts across grades and subject areas even where they find improvements in teacher effort, which they suggest is in part due to poor quality teaching. Thus, even if the intervention succeeds at encouraging participation of communities in schooling which leads to increasing teacher effort at school, the effect that this will have on student performance is constrained by the quality of the teaching.

Concerns with elite capture did not appear to have materialised in Kenya (Duflo et al. 2012)

Duflo et al. (2012) report that there was a concern that empowering parents within the Parent Teacher Association (PTA) to hire and monitor their own teacher could lead to a bias towards hiring of locals or favouring of certain ethnic groups for positions. However, they found that the program suggested schools were 12 per cent less likely to hire contract teachers from the local area and the share of relatives of existing teachers among contract teachers fell by around half.

Public- Private Partnerships

In the following section, we report the results of the synthesis of qualitative findings from our included programmes, presented using the hypothesised programme theory as an overall framework for structuring the synthesis. This first section looks at process and implementation features that may be associated with relative success and failure, followed by a review of the contextual barriers and facilitators to success.

Process and implementation

***Inefficient administration led to an underutilisation of WB loan (*King, Orazem & Wohlgemut (1997:489), Process Evaluation)**

The PACES program was burdened by a lack of communication between schools and administrators, which led to the underutilisation of the World Bank loan (World Bank, 2001)[167].

***Delays to payment of vouchers put strain on programme implementation and school quality (*King, Orazem & Wohlgemut (1997:489), Process Evaluation.**

First of all, inefficient administrative processes were reported to cause delays in the payments of vouchers and put a strain on the PACES program (King et al., 1997: 489)[168].

[167] Project Document
[168] Project Document

***Schools faced financial difficulties due to by unsatisfactory fee collection and a lack of subsidies* (World Bank (2001): Project Document)**

Many schools in the Urban Girls' Fellowship Programme were hampered by unsatisfactory fee collection and a lack of subsidies (SCPEB, n.d)169. As the average tuition of 58 rupees per student (1999) was below the anticipated break-even level of 75 rupees per student, many schools faced financial difficulties. Although some schools were able to raise additional funds through increased enrolments (Orazem, 2000)170, only six of the original eleven schools attained self-sufficiency by the end of the three programme years (by 1999). Those schools that were able to secure in-kind transfers of facilities were the most successful at containing costs (Alderman et al., 2003)[171].

Programme appeared to have successfully targeted low-income or underserved students (Barrera-Osorio and Raju, 2011 (Impact Evaluation); Calderon, 1996 (Process Evaluation);

Positive findings on service delivery were reported for two programmes: both the PACES programme (Calderon, 1996)[172] and the Foundation Assisted Schools (FAS) Programme were reported to have successfully targeted the intended populations (Barrera-Osorio and Raju) [173.] Firstly, Calderon (1996:17) reports that the PACES programme ''has successfully targeted low-income students''. Secondly, Barrera-Osorio and Raju (2011) report that the programme was designed to target districts ranked lowest in terms of adult literacy rates. While in the first two phases of the programme this criterion was not applied, in phase 3 and 4, the period of evaluation, it was applied effectively.

Not all Upazilas followed the targeting criterion of focusing on out-of-school children (Ministry of Planning and GoB, 2014 (Qualitative research).

The Ministry of Planning and GoB (2014) report that many Upazilas involved in the ROSC school programme (Dang et al., 2013) did not follow the agreed targeting criteria (enrolment rates, gender gap in enrolment, poverty levels). It is reported that some Upazilas had less out of school children than others, depriving some deserving areas.

PPP programs well-received by stakeholders* (Angrist et al.; *Orazem,,2000 (Project Document);* Villa & Duarte (2002) (Project Document*)

Findings on the take up of the interventions among intended participants were reported for four programmes. This information suggested that these programmes were generally successful and well-received by various stakeholders. The UGF and PACES Programmes were reported to have expanded rapidly (Calderon, 1996: 9-10; Orazem, 2000: 11)[174]. In case of the former, the rapid expansion of the programme is reported to have been at least partly influenced by the enthusiasm and willingness of the government agency involved to take a very active role in shaping the project despite the initial scepticism of the government (Orazem, 2000)[175]. In case of the latter the authors report that ''since many public secondary schools in Colombia were turning away applicants due to overcrowding, PACES is likely to have opened up places in public school for other pupils by reducing public-school queuing'' (Angrist et al., 2002:1543).

[169] . Project Document
[170], Project Document
[171] Impact Evaluation.
[172] Process Evaluation.
[173] Impact Evaluation.
[174] Process Evaluation [PACES]; Project Document [UGF Programme]
[175] Project Document.

The Concessions School Programme in Colombia was also reported to have been well-received by the mayors of the receptive cities, a factor which may have had a positive influence of program implementation (Villa and Duarte, 2002)[176]

Stakeholder participation central to some programs (Irarrázaval et al., 2012 (Mixed Methods); Kim et al. 1998 (Impact Evaluation))

Firstly, In the UGF program Parent Education Committees (PECs) developed a school proposal and selected the private providers (Kim et al. 1998). A survey by the Society for Community Strengthening and Promotion of Education, Balochistan (SCSPEB) nine years after the programme all original UGF schools, the same schools studied by the included impact evaluations, still existed and were reported to have devoted and motivated staffs as well as parental and community support (SCSPEB, 2004)[177]. Secondly, through interviews with different stakeholders in SEP programme schools, Irarrázaval et al. (2012) found that in almost all schools that were part of their sample, there were opportunities for teachers, and in some cases also for students, to participate in the development of the school improvement plan Irarrázaval et al. (2012:32)[178].

***Lack of organisational capacity a barrier to effective implementation and monitoring, with reports of ghost' voucher awardees (*Ministry of Planning & GoB, 2014*;* World Bank, 2013; King et al., 1999 (Process Evaluation); SCSPEB, n.d (Project Document).**

The lack of organisational capacity by some of the actors involved posed a problem for both the PACES and UGF programmes (King et al., 1999; SCSPEB, n.d.)[179]. In case of the former, a lack of government capacity as well as bureaucratic obstacles led to issues in the monitoring of the program, which made it difficult to guard against 'ghost' voucher awardees (King et al., 1997)[180]. In case of the latter, the newly established schools lacked technical capacity and they were provided with limited guidance to fill this gap (SCSPEB, 2004)[181].

Additional documentation to the IE by Dang et al. (2011) reports some management and monitoring issues at various scales in the programme. Additional documentation by the Ministry of Planning & GoB (2014) reports that The Upazila Education Committee, who were responsible for assessing the feasibility of the establishment of a school in the first place (p.20), was largely inactive due to their preference to first assist projects with higher priorities. Its officers (Upazila Education Officers) were moreover swamped by too many responsibilities, resulting in a lack to monitor regularly.

Additional documentation to Dang et al.'s (2013) evaluation also note organisational issues that may have limited programme impact. The World Bank (2013) notes that ROSC was launched during the early stage of The Second Primary Education Development Programme **(***PEDP-II***).** As a result, the two operations competed for similar expertise ''in a fairly limited domestic market'' (p.8), resulting in delays in recruiting ROSC personnel, and in turn, in uneven attention to community mobilisation.

The Ministry of Planning & GoB (2014) furthermore reports issues with the Education Service Providers (ESPs). The awarding of 100 TK per month to these ESPs for monitoring students often resulted in the registration of fake students. In addition, since many NGOs

---

176 Project Document
177 Project Document. This information is based on a survey done with all the original UGF schools nine years after the start of the programme. No methodology was reported.
178 Mixed Methods Paper.
179 Process Evaluation [PACES]; Project Document [UGF Programme]
180 Process Evaluation
181 Project Document.

were ESPs, NGOs were established spontaneously upon hearing of the project. This resulted in the dissatisfaction of local community leaders with NGOs. It is furthermore reported that, during ROSC preparations, it was highly unfeasible to assess the capacity and provisions of NGOs and how the money they received for ESP services would be spent. Although the CMC is expected to select responsible ESPs, in practice however, parties outside of the CMC often selected service providers.

Center Management Committees (who were in charge of the day-to-day management of schools) met infrequently **(Ahmed, 2004)**

The study by Ahmed (2004) provides descriptive statistics on the Center Management Committee (CMC), who were in charge of the day-to-day management of ROSC schools together with Education Service Providers (ESPs). Very few CMCs met on a weekly or even bi-weekly basis, with the majority of CMCs in both grant and grant plus allowance areas meeting on a monthly basis (57.7 and 72.3 per cent of CMCs respectively). A smaller but substantial percentage of CMCs met at an irregular basis (no specific time interval) (28.9 per cent of CMCs in grant areas and 14.1 per cent of CMCs in grant plus allowance areas). In the majority of CMCs both the president, a school teacher and a Guardian representative attended the meeting.

High turnover of government of implementation staff (CfBT Education Trust, 2010; King et al., 1997; World Bank, 2013:8)

Three programmes were affected by a high turnover of implementation or government staff. The FAS programme reported had a high turnover of implementation staff (CfBT Education Trust, 2010). King, Orazem and Wohlgemut (1997) report that a high turnover of government staff might have eroded government support for the PACES programme. Finally, government turnover also slowed down the implementation of the ROSC School programme as it resulted in changes in community leaders, ESPs, and local officials who had been closely aligned with previous governments. Although this occurred three times over the course of the programme, it is reported that *"original concept of the project was not disturbed"* (World Bank, 2013:8).

Schools did not comply with at least one of the programme condition and the ministry of education did not attempt to enforce compliance by encouragement or sanctions (Adelman 2015[182])

The majority of schools in the Tuition Waiver Programme did not comply with at least one of conditions associated with the progamme (providing three school text books per child; a maximum of 45 children per class and a maximum of 2 classes that can participate in the programme per school). In addition, the ministry of education did not attempt to enforce compliance by encouragement or sanctions (Adelman and Holland 2015).

Intermediate outcomes

Four out of eleven included studies measured outcomes along the causal chain as part of their counterfactual analysis (Angrist et al., various; Muralidharan & Sundararaman, 2013; Barrera-Osorio & Raju, 2011; Dang et al., 2011).

Students enrolling had higher scores than the district average **(Zhang, 2009:9)**

Baseline student data revealed that "Students enrolled in magnet schools had a mean 6th grade test score that was 0.47 SD above the district average. In contrast, students enrolled

[182]Impact Evaluation

in neighborhood middle schools had a mean 6th grade test score 0.10 SD below the district average" (Zhang, 2009:9).

There was limited change in children's time use at home and household spending patterns (Muralidharan and Sundararaman, 2013 (Impact Evaluation).

Muralidharan and Sundararaman (2013) looked at intermediate outcomes related to the educational and financial behaviour of students and their families, which might have changed as a result of financial relief vouchers offered. Although, the authors find that children in the treatment group spent 12.83 minutes (p value less than 0.01) less doing household chores than children in the control group and spent 65.94 rupees less on uniforms and 69.94 rupees less on notebooks/textbooks per year, they find no significant overall change in household spending or in time spent doing homework among voucher-winning students. This leads the authors to suggest that "the impact of the program (if any) is most likely to be due to changes in school as opposed to household factors" (p.2)..

***Children in private school spent more time at school due to longer days and school year* (Muralidharan)**

Private schools have a longer school, longer school year (2 working weeks or 11 days longer per year), and have considerably lower pupil teacher ratios (around a third lower) than government schools.

Private schools spent less time on Maths and Telugu, but more on English, Science, Social Studies and Hindi (Muralidharan and Sundararaman, 2013)

Muralidharan and Sundararaman (2013) note that "after two and four years of the program, we find no difference between the test scores of lottery winners and losers on math and Telugu (native language). However, private schools spend significantly less instructional time on these subjects, and use the extra time to teach more English, Science, Social Studies, and Hindi." (Abstract). The authors report that "private schools allocate less time per week to Math (200mins) and Telugu (160mins) than government schools. More time is spent on English (90mins), Social Studies (65 mins), Science (100mins), Hindi (215mins), Computer use (45 mins) in private schools and an additional hour per week is spent on arts, craft, sports and study hall (Muralidaran 2013: 18). It is for this reason that the analysis, which is limited to Math and Telugu, might not provide a complete picture of the impact of the voucher. They therefore conducted additional tests in Science/Social Studies (EVS) and Hindi (based on time table data). They find that: "While this still does not account for all the subjects (computer use for instance), the tested subjects now account for over 80% of instructional time in both types of subjects and are also closer to being equal across school types (81% for private and 85% for public schools). The full set of test score results are presented in columns Table 7 – columns 5 to 10, and we see that voucher winning students score slightly better on EVS (though this is not significant). The most striking result though is that they do dramatically better in Hindi – scoring over 0.5 standard deviations better than students who did not win the voucher, and the impact on Hindi scores of actually attending a private school is even more pronounced with students who attend private schools scoring nearly 0.9 standard deviations (SD) better. Averaging across all subjects, students who won a voucher score 0.13 SD better than those who did not, and the causal impact of attending a private school is estimated as 0.23 SD (column 10), and both estimates are significant at the 1% level" (p.18).

Teachers in private schools less educated, younger and paid lower salaries (Muralidharan and Sundararaman, 2013 (Impact Evalution)

As reported by Muralidharan and Sundararaman (2013): ''private schools pay substantially lower teacher salaries (less than a sixth of that paid to public school teachers), and hire teachers who are younger, less educated, and much less likely to have professional teaching credentials. However, they hire more teachers and have smaller class sizes and less multi-grade teaching than public schools. Using official data as well as data collected from direct observations conducted during unannounced visits to schools, we find that private schools have a longer school day, a longer school year, lower teacher absence, higher teaching activity, and better school hygiene'' (Muralidharan and Sundararaman 2013: 2);

''Government school teachers are more likely to be male, are considerably older, have more years of teaching experience, are more likely to have completed a college degree, and are much more likely to have completed a teacher training course (Table 3 - Panel B). However, they are less likely to be from the same village as the schools that they are assigned to, and are paid six times higher salaries'' (Muralidharan and Sundararaman 2013:13). Government school teachers are more likely to be male, older, have more years of teaching experience, to have a college degree and have teacher training. They are less likely to be from the same village and are paid higher salaries'' (Muralidharan and Sundararaman 2013: 13)[183;] ''The total spending per-child spending in the government schools is over 3.5 times the mean per-child spending in the private schools in our sample (Table 3 - Panel C).21 As the discussion above makes clear, the main driver of these differences in costs is the much higher salaries paid to government school teachers'' (Muralidharan and Sundararaman, 2013:13).

***Teachers in private schools better qualified and higher per student spending (*Zhang, 2009)**

Magnet Schools were reported to have better qualified teachers and higher per pupil spending than neighborhood public schools. They were also sought after, and enrol a disproportionate number of students from families with high socioeconomic status (SES) and high primary school test scores

Private schools outperform government schools on measures of classroom practices, teacher absence and teacher performance (Muralidharan and Sundararaman, 2013)

As reported by Muralidharan and Sundararaman (2013): ''However, private schools hire more teachers per student…..Private schools significantly outperform government schools on all measures of observed classroom processes (Table 4 – Panel A). Classrooms in private schools are significantly more likely to be engaging in active teaching (51% vs. 34%), have a greater likelihood of a teacher being in the classroom (97% vs. 92%), and are much less likely to be multi-grade classrooms where more than one grade is taught simultaneously by the same teacher (24% vs. 79%). Moreover, enumerators observed teachers in private schools as being more likely to be in complete control of the class (69% vs. 41%) and as more effective in teaching and maintaining discipline (50% vs. 36%). We find from observations at the teacher level (Table 4 – Panel B) that government school teachers were considerably more likely to be absent than private school teachers (24% versus 9%) and

[183] Teacher characteristics reported are reported in table 3 including: male, age, years of teaching completed college or masters, teacher training, from same village, current gross salary per month (Muralidharan 2013: Table 3).

less likely to have been actively teaching at the point of observation (35% versus 50%)'' (pp. 13-14).

Private schools had better infrastructure, equipment and supplies than public schools* (Barrera-Osorio, 2006; *(Muralidharan and Sundararaman, 2013)

As reported previously, concession schools were handed over to private schools with the highest standardised test scores. Both summary statistics and descriptive evidence has shown that these providers did indeed manage to provide relatively high-quality resources. Based on field visits, Barrera-Osorio (2006) notes that concessions schools had very good infrastructure in contrast to public schools. They were built on better lots than the average public schools, with better equipment and a complete set of supplies for classrooms, laboratories and libraries. Since concession schools were allowed to implement their own pedagogic model, there was a difference in certain school features. Some schools were reported to encourage parent-teacher meetings on a regular basis, others were reported to work with the community, or provide psychological and nutritional support in addition to regular education.

''Private schools more likely to have drinking water, functional toilets and separate toilets for girls, functional electricity, and to have a computer, with the differences being quite stark on some of these measures. Government schools are more likely to have a library and radio'' (Muralidharan and Sundararaman 2013: 13).

''Finally, enumerators also coded measures of school hygiene based on their observations when they entered the schools and we find that private schools are less likely to have indicators of poor hygiene such as having garbage dumped on the school premises, having stagnant water (breeding ground for mosquitos), or having a heavy presence of flies on the school premises (the most common carrier of pathogens from open human and animal waste)'' (Muralidharan and Sundararaman 2013: 14).

***Increase in availability of teachers, classrooms and blackboards, but not in student-teacher and student-classroom ratios* (Barrera-Osorio and Raju, 2011)**

The FAS programme was aimed at increasing educational outcomes through the improvement of *inter alia* the schools' physical infrastructure and, student-teacher and student-classroom ratios schools (Barrera-Osorio and Raju, 2011)[184]. The authors evaluate whether the programme succeeded to do so in the first place: the authors find evidence of large positive impacts on the number of teachers, classrooms, and blackboards. . Having said this, the authors find no evidence of impacts at the cut-off on student-teacher and student-classroom ratios (p. 33). The authors also find no effect on the number of toilets. This is concerning, they note, ''given the large expansion in enrolment in marginal passers and its potential negative bearing on the use and maintenance of the facility'' (Barrera-Osorio and Raju, 2011:33)

No significant changes in availability of school inputs after programme (Barrera-Osorio et al., 2015 (Impact Evaluation).

Although the programme did not yield significant changes in school governance, it did have a significant impact on school inputs. Barrera-Osorio et al. (2015) note that:

''Treatment schools had a similar number of teachers, with no systematic changes in teacher

[184] Impact Evaluation

composition in terms of percentage of female teachers and permanent teachers….In terms of school infrastructure, the only discernible impact of the program was on the presence of a science laboratory in schools; with approximately 20% more treatment schools reporting having a science laboratory. There do not appear to be significant differences in other conditions, such as working toilets, class cleanliness, or number of furniture for students. These results seem logical. Given the limited transfer amount, it seems unlikely that schools would be able to invest significantly in school infrastructure. However, it appears that participating private schools are using at least part of the government transfers to adapt existing infrastructure and purchase equipment for school laboratories….These, admittedly marginal, improvements in school stability and availability of school inputs, in terms of teacher effort and science laboratories, could explain part of the observed improvements in student performance.'' (p.11).

Significant changes in teacher presence after programme *(Barrera-Osorio et al., 2015 (Impact Evaluation).*

Barrera-Osorio et al. (2015) note that, although there were no significant changes in school inputs after the programme. There were significant changes in the ''utilization of the teacher input'' (p.11). The authors notes that ''a higher proportion of treatment teachers was present and more likely to be in class at the time of the unannounced school visits'' (p.11).

Lack of and low quality of school facilities* Ahmed, 2006 (*Impact Evaluation Baseline Survey)

Some programmes led to the establishment of new schools. The ROSC school program was one such programme and aimed to not only improve access to but also the quality of education. Several additional documents, however, report a lack of and low quality of school facilities. A survey by of Ahmed (2006), who provides an in-depth overview of the baseline survey used by Dang et al. (2013), sheds light on the school facilities available in the ROSC schools. It shows that many of the schools still lack basic facilities including a blackboard, desk and chairs for the teachers, desks and chairs for students, toilet facilities for students, and drinking water. Almost none of the schools have electricity (table 6.3). In addition The Ministry of Planning & GoB (2014) report that facilities provided by the schools are not satisfactory; about 84 percent of students sit on the floor and all students use the water source of the nearest neighbour for drinking water and use the neighbour's toilet for defecation (34-35). The document additionally reports that teachers reported a lack of seats during the primary education completion exam.

Voucher winners more likely to access private education (Angrist et al., 2011)

Angrist et al. (2010) show that voucher winners were 51 percentage points more likely than losers to have been using some kind of scholarship (including non-PACES scholarships that lottery winners were 6-7 percentage points more likely to have begun sixth grade in a private school and that 15-16 percentage points more likely to be in private school at the time their survey (Angrist et al., 2010:10) [185] [186].

[185] Angrist et al. (2001:10) note: ''just as not all winners were using a scholarship, some losers obtained scholarships from programs other than PACES and one loser was awarded a PACES voucher after re-applying the following year''.
[186] *Angrist et al. (2001:10) also note: ''the effect of winning the PACES lottery on the probability of the private-school attendance was even larger in seventh grade, probably because losers were more likely to have left private school by then''*

Participating schools have significantly higher proportion of students from educationally-favourite backgrounds (Barrera-Osorio et al., 2015 (Impact Evaluation).

Barrera-Osorio et al. (2015) report that: ‘‘students in treatment schools are younger on average. They also appear more likely to be coming from households that are: (i) more invested in children’s schooling (parents reported to be more likely to visit the school), (ii) are financially more secure (have a higher index of assets); and (iii) are more educated (students in treatment schools report a higher education level for their father). More directly capturing differences in unobservables as shown in figure 2, students in the treatment group perform better on the primary leaving exam than students in the control group. … Overall, it seems that students in treatment schools are different from their peers in control schools.

Specifically, they come from backgrounds that are positively associated with student achievement. These results strongly suggest student selection on the part of low-cost private schools associated with the PPP program. Such selection presents a very plausible explanation for the observed gains in student performance in PPP schools.’’ (p. 12)

Contextual Barriers and Facilitators

Inflation reduced the value of the voucher, increasing co-payment (Bettinger et al. 2008 (Impact Evaluation))

The majority of the programmes targeted at low-income or underserved populations do not charge students and their parents co-payment of education. The voucher offered by PACES programme, which targeted low-income students, was initially set to cover most tuition fees. Inflation, however, led to a significant devaluation of the voucher. By the time of the study evaluating the programme (1998) the voucher covered about 56 per cent of tuition, meaning that voucher recipients had to supplement it with private funds (Bettinger et al., 2008)

There was opposition from teacher unions to privatisation of education (Orazem 2000 (Project Document)

The PACES implementation saw an enormous opposition from the teachers’ union, who were against the privatisation of education. In response, the government committed to increasing enrolment in public schools at the same time as the voucher program (Orazem, 2000)

Requirement that community contributed land made it difficult to assign land for schools (Orazem 2000 (Project Document)

In the UGF program in Pakistan the government financed the establishment of the new schools, but it was required that the community donate the land necessary to build the schools. The schools, however, were established in the urban slums of Quetta, where its inhabitants have no legal entitlement to land. It was reported that this made it difficult for the community to assign land to program purposes, even if this land was not in use (Orazem, 2000).

Poor weather conditions were a common reason for not going to school **(Barrera-Osorio et al., 2015 (Impact Evaluation)**

‘‘Some common reasons for not reaching schools were weather conditions leading to lack of access or school closure’’ (Barrera-Osorio et al., 2015: 6).

Multilevel interventions

In this section, we report the results of the synthesis of qualitative findings from the 12 studies of 10 unique programmes. We identified several additional documents including donor audits and implementation documents (aside from the Impact Evaluations) that present qualitative, process and project information for these programmes.

Process and implementation outcomes

Opinions about the programme were generally positive (Cerdan-Infantes and Vermeersch, 2007)

The authors of the Impact Evaluation assessing the Full-Time School (FTS) programme in Uruguay cite evidence from a qualitative study of the programme carried out by Equipos/MORI in 2001, which found that opinions of the programme were generally positive.

The implementation of the extension of the school day, school building and teacher training was practically universal in participating schools (Cerdan-Infantes and Vermeersch, 2007)

The authors of the Impact Evaluation assessing the Full-Time School (FTS) programme in Uruguay cite evidence from a qualitative study of the programme carried out by Equipos/MORI in 2001, which found that some sub-components of the programme were implemented unequally across schools. The IE authors, nonetheless, find that *"the implementation of the extension of the school day, school building and teacher training was practically universal in participating schools"* (Cerdan-Infantes & Vermeersch, 2007).

The programme faced financial issues, including delays in budget approvals (Paqueo and Lopez-Acevedo., 2003)

Paqueo and Lopez-Acevedo (2003) report that the PARE programme in Mexico faced financial issues, including delays in budget approvals. The authors note:

"Some of the problems faced by the compensatory programs were: (a) annual delays in budget approval for project expenses; (b) persistent complicated internal procedures and controls for approving budgets for specific activities; and (c) prolonged postponement of decisions regarding important studies for evaluation'' (Paqueo and Lopez-Acevedo., 2003)

The programme was not well implemented in the urban schools (Paqueo and Lopez-Acevedo, 2003)

''A surprising finding is that in urban schools improvement in the learning achievement of students appears lower for the experimental vis-à-vis control group. It is difficult to explain this result. It is possible that the selection of the control group was not properly done. Having said this, we note that the PARE program was not well implemented among the urban schools. Delivery of planned interventions were either not delivered or delayed.'' (p.7). The authors additionally note that the programme was implemented in urban schools only in its last year.

Headmasters increased school fees of CSP schools (Kremer et al., 2003)

Kremer et al. (2003) report that there were increase in schools fees in CSP schools that were not intended by the programme: "We also have anecdotal reports that some headmasters increased the collection of school fees since they knew that parents would be willing to pay increased fees to attend CSP school" (Kremer et al., 2003)

Teacher training was difficult to transfer to the classroom (Sotomayor, 2005)

In some cases the teacher training element of the intervention successfully introduced a group-focussed, active and participative classes. In others, teachers found it difficult to transfer the new pedagogical principles to classes and struggled to organise class sequencing, time-management and student discipline. (Sotomayor, 2005: 258).

There was a lack of supervisors, resulting in complaints from schools about inadequate support (Carlson 2000: 52, Undurraga 1994: 23-5 and Guttman 1993: 22).

Supervisors that worked with schools on behalf of the ministry of education to implement the P900 programme in Chile came under criticism and faced various challenges. They visited schools in an infrequent fashion (one reported not having seen a supervisor in years) (Carlson 2000: 52). High staff turnover of supervisors was reported as a problem for schools, as well as high workloads for supervisors resulting in limited time available for each school (Carlson 2000: 52, Undurraga 1994: 23-5 and Guttman 1993: 22).

Some schools were withdrawn early from the programme (Undurraga, 1994)

Undurraga, (1994: 23-25) reports that schools were withdrawn from the P900 programme too early. It is clear from MINEDUC (2001: 20) that some schools that had left the programme later re-entered. This supports the idea from other documents that schools were withdrawn too early.

The programme changed year on year (Tokman 2002; Chay, 2005; Lockheed, 1999)

The P900 programme design was continually modified to better meet the needs of target schools (Tokman 2002: 2 & 6). It is not clear what this change entailed. Fewer programme elements were available early on in the programme; it took some time for the key components to be rolled out, a factor that Chay speculates may have limited effectiveness in year 1 (see Chay, 2005, 1253).

Similarly, Lockheed (1999) notes that the NHP program was not implemented uniformly across all treatment schools. This was an intended feature of the implementation to ensure that schools received only the intervention components that they needed (Lockheed et al., 2006).

Teacher workshops were often substituted with other activities when schools left the programme (Ugarte, 2011)

Ugarte (2011) cites findings from a critical analysis of the P900 Programme by Santiago Consultores (2000). This study finds that the teachers' workshops were highly appreciated by the school principals and the teachers themselves. However, once the school would exit the program, these workshops were interrupted and substituted with other activities.

While student support was well developed, the educational management component was not as effective (Ugarte, 2011)

Ugarte (2011) cites findings from a critical analysis of the P900 Programme by Santiago Consultores (2000). This study finds that student support was the best developed area of the P900. Differently, the area of education management, was not very appreciated nor effective as it did not seem to have an effect on the quality of the relationships between teachers and school principals (p.4)

There was resistance from schools to participate in the TSFS programme in Chile, causing delays in implementation (Sotomayor, 2006; Sotomayor and Dupriez, 2007)

The negative connotation associated with being labelled an underperforming school resulted in resistance from schools to participate in the TSFS Programme in Chile. There was also resistance to the programme design that placed emphasis on student results over education in general. It was a time consuming process for implementing agencies to overcome this resulting in delay in effective implementation (Sotomayor, 2006 and Sotomayor and Dupriez, 2007).

The BRIGHT programme was mainly implemented as intended however some villages which were selected to receive a BRIGHT school did not because of poor infrastructure (Kazianga et al, 2013; De Hoop et al., 2012)

Of the villages that were selected, most received a BRIGHT school as intended. However, some of the selected villages did not have a suitable water source and so did not receive their school (De Hoop et al., 2013; Kazianga et al., 2013). A shortage in program funds prevented a further six villages from receiving their BRIGHT schools (De Hoop et al., 2013).

In addition, there were around four villages in the programme which should not have been selected but did receive a BRIGHT school. Both Levy et al (2009) and Kazianga et al (2013) note that even though these villages did not rank highest in terms of their eligibility for the programme they were the next highest ranking in their area. This was at least partially consistent with the assignment rules for the programme.

Despite the largely unproblematic implementation and generally better quality of schools, the long-term progress of BRIGHT schools may have been negatively affected by lack of maintenance. (Kazianga et al, 2013)

An MCC evaluation found that though responsibilities for maintenance were established at the beginning of the project though this does not seem to have been acted upon. Further, the government did not provide the funders with evidence that they had available funds for maintenance of schools after the programme had finished. This lead to concerns that the longer-term impact of BRIGHT schools would lessen due to concerns by parents that schools were unsafe or unproductive places for their children (MCC, 2010).

Use of school facilities was varied. In some cased schools were underused whereas in others schools were oversubscribed. (Kazianga et al, 2013)

BRIGHT schools were designed to accommodate 50 students per class but it was found that often the classes had over 80 students while in other areas the classes only had 3 students. The MCC audit (2010) suggested that may have been due to the nomadic nature of many of the people in the programme area.

Transfer of students was not a factor influencing the impact of the programme (Rodriquez & Sanchez, 2010).

It was suggested that part of the PER impact may be due to transfer of students across school. However, Rodriquez & Sanchez (2010) reject this possibility on the basis that there are many barriers to school transfers in rural Colombia including expense. The authors also carried out robustness checks and found that there results were indeed robust (Rodriquez & Sanchez, 2010).

Comparison schools were exposed to similar programmes (Lockheed, 1999).

Lockheed et al (1999) comments that it is rare to find a school in Jamaica that has not received some kind of schooling intervention which may account for the lack of impact in the evaluation of NHP.

There reportedly was a successful change in behaviour within schools (Sotomayor and Dupriez, 2007)

A qualitative study noted that there had been successful change in behaviour within the classroom after the TSFS programme in Chile (Bellei 2013: 237). This was reported as partially caused by the responsiveness of the programme to adapt in response to the initial resistance (Sotomayor and Dupriez, 2007).

While the number of classes offered at programme schools increased only modestly, the programme led to substantial increases in class size (Kremer et al., 2003)

Kremer et al. (2003) report that while the number of classes offered at programme schools increased only modestly, the programme led to substantial increases in class size after the Child Sponsorship Programme (CSP) in Kenya:

''Class size in grades 1-7 increased by 8.9 students despite an average increase of .27 classes offered per grade in each school. Table 2 shows the average class size for the program and comparison schools before and after the intervention. School register data for Years 1-3 suggest an increase in class size of 4.1 students. Years 4 and 5 show an increase in class size of 8.9 students on a base of around 29 students. Since class size results from both student enrolment and teacher postings, it can fluctuate from one year to the next as each group responds to the trend of the other. In grades 3 to 8, the ones for which we have data on test scores, the program increased class size by 11.2 pupils.'' (Kremer et al., 2003).

There is some evidence to suggest that the BRIGHT programme changed parents' attitudes towards education, making them more willing to accept children spend time in school (Levy et al., 2009).

Parents of BRIGHT school students were more likely to say that there should be no limit to the amount of time spent on their child's schooling. They were also less likely to say that their children should spend no time on education at all suggesting that the BRIGHT programme had a positive effect on parent's attitude towards education. Further, this effect was more pronounced for girls than boys (Levy et al., 2009).

The programme should not affect children's participation in other work. Though the effect of this is unclear (De Hoop et al., 2013; Levy et al., 2009).

BRIGHT schools may or may not have affected child work. While De Hoop et al (2013) found that child work (including household chores as well as economic activity) was not affected despite increased enrolment rates, Levy et al (2009) found that child work decreased.

International Initiative for Impact Evaluation
London International Development Centre
36 Gordon Square
London WC1H 0PD
United Kingdom

3ieuk@3ieimpact.org
Tel: +44 207 958 8351/8350

www.3ieimpact.org