Improving early grade reading in South Africa

Stephen Taylor, Department of Basic Education

Jacobus Cilliers, Georgetown University

Cas Prinsloo, Human Sciences Research Council

Brahm Fleisch, University of the Witwatersand

Vijay Reddy, Human Sciences Research Council

Grantee Final Report

Accepted by 3ie: March 2019

Note to readers

This impact evaluation has been submitted in partial fulfilment of the requirements of grant PW2.10 issued under Policy Window 2. This version is being published online as it was received. A copy-edited and formatted version will be available in the 3ie Impact Evaluation Report Series in the near future. 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 and omissions are the sole responsibility of the authors. All affiliations of the authors listed in the title page are those that were in effect at the time the report was accepted. Any comments or queries should be directed to the corresponding author, Stephen Taylor at tayor.s@dbe.gov.za.

The 3ie technical quality assurance team comprises Beryl Leach, Deo-Gracias Houndolo, an anonymous external impact evaluation design expert reviewer and an anonymous external sector expert reviewer, with overall technical supervision by Marie Gaarder.

Suggested citation: Taylor, S, Cilliers, J, Prinsloo, C, Fleisch, B and Reddy, V, 2019. *Improving early grade reading in South Africa, 3ie Grantee Final Report.* New Delhi: International Initiative for Impact Evaluation (3ie).

Funding for this impact evaluation was provided by 3ie's donors, which include UK aid through the Department for International Development, the Bill & Melinda Gates Foundation and the Hewlett Foundation. A complete listing of all of 3ie's donors is available on the 3ie website.

Contents

Abbreviations, acronyms and key terms

ANA	Annual National Assessments
CRC	Community Reading Coach
Class Act	EGRS service provider
DBE	Department of Basic Education
Dr Kenneth Kaunda	One of the two districts where EGRS was implemented
EGRA	Early Grade Reading Assessment
EGRS	Early Grade Reading Study
GPLMS	Gauteng Primary Language and Mathematics Strategy
HSRC	Human Sciences Research Council
Ngaka Modiri Molema	One of the two districts where EGRS was implemented
PIRLS	Progress in International Reading Literacy Study
RCT	Randomized Controlled Trial
SGB	School Governing Body
RDD	Regression Discontinuity Design
ToC	Theory of Change

List of tables

List of figures

Acknowledgements

The Department of Basic Education would like to acknowledge the work of the EGRS Research Team who provided oversight to the project and to the production of this report. The Research Team includes Stephen Taylor (DBE), Brahm Fleisch (University of the Witwatersrand), Jacobus Cilliers (Georgetown University), Cas Prinsloo (HSRC) and Vijay Reddy (HSRC). Carol Nuga-Deliwe, Nompumelelo Mohohlwane and Janeli Kotzé have worked closely with the Research Team and have played a critical role at an intellectual and direction-setting level.

This project would not have been possible without generous financial support from a number of partners. In this regard we would like to thank the International Initiative for Impact Evaluation (3ie) for funding the evaluation side of the project and for providing peer reviewed feedback and support throughout the project. From 3ie, there was significant advice and support provided by Howard White, Beryl Leach, Deo-Gracias Houndolo and Heather Lanthorne. We would also like to thank the ZENEX Foundation, UNICEF and Anglo American Chairman's Fund for funding the implementation of the three EGRS interventions. In particular, we thank Gail Campbell, Fatima Adam and the rest of the ZENEX Foundation staff team for engaging on the content of the project and providing critical feedback and support. Similarly, at UNICEF we would like to thank Saadhna Panday-Soobrayan, Nadi Albino and Wycliffe Otieno for robust discussions and support on the evaluation findings at various stages throughout the project.

The DBE would like to thank Class Act Educational Services for doing an excellent job of delivering the three EGRS interventions. The quality of training provided and responsiveness to additional requests went above and beyond what was formally required. In particular, we would like to acknowledge the work of Penny Groome, Lorraine Marneweck, Maryna du Plooy, Nonhlanhla Hlomuka, Thili Thela and Kaitlin Alsofrom at Class Act, as well as the Coaches and Trainers – Dorothy Molapo, Sabi Mlambo, Helen Kgobane and Kgomotso Phaletse.

The Human Sciences Research Council (HSRC) have done a tremendous job of managing the evaluation side of the project - data collection, data capturing and working with the research team on instrument design and research interpretation. Led by Cas Prinsloo, we must thank the HSRC team members: Lolita Winnaar, Charlotte Motha, Matthews Makgamatha, Jacqueline Harvey, Shawn Rogers, Elmi de Koning and Matselane Maja. For Setswana language expertise on the development and revision of learner tests, we acknowledge the excellent work done by Badiri Moila. Social Data Collectors, led by Rosina Langa, were subcontracted by the HSRC to conduct fieldwork in Waves 2 and 3 and were a pleasure to work with.

We would also like to thank the North West Department of Education for partnering with us as the host province. The hard work by Mpho Motlhabane, Mahlodi Teu, Thabo Phorabatho, Godi Poti and Boitumelo Moeng is much appreciated. We also recognize the support from officials in the districts of Ngaka Modiri Molema (led by the Director Benny Monale) and Dr Kenneth Kaunda (led by the Director Hasimbhay Motara) together with all the curriculum coordinators and area office staff who supported the project and served on the curriculum reference group.

The Department of Planning, Monitoring and Evaluation has been a hugely supportive partner, both financially and through including EGRS on the National Evaluation Plan. The personal

involvement of Thabo Mabogoane, Ian Goldman, Jabu Mathe and Christel Jacob has been much appreciated.

We also thank the University of the Witwatersrand for undertaking the financial management and contracting regarding the interventions side of the EGRS project as well as for academic collaboration, led by Brahm Fleisch.

Members of the DBE Directorate for Research Coordination, Monitoring and Evaluation have served tirelessly to coordinate and manage this project. We thank Justice Libago, Sasabona Maseko, Tshirado Tshirado, Neo Makgato, Ernest Rasekgwalo, Tariro Chiweza, Mikateko Mbambo and Jesal Kika for their hard work.

Within the DBE, certain individuals across various parts of the organization have supported the project and provided advice along the way and deserve special mention: Mathanzima Mweli, Waheeda Carvello, Haroon Mahomed, Chintha Maharaj, and Martin Gustafsson.

A number of other people have advised and supported the EGRS from inception and continue to do so. We thank Servaas van der Berg, Luis Crouch, Kamilla Gumede, Linda Zuze, Cheryl Reeves, Ursula Hoadley, Nicholas Spaull, Nick Taylor, Kerryn Dixon, Bobby Soobrayan, Benjamin Piper, Sonja van der Vyfer and Saguna Gordhan.

We are grateful to the teachers, the parents and ultimately the children who participated in this project in various ways. We hope this project has benefited them but also recognize that their participation is intended to lead to the benefit of children throughout the system going forward.

Executive summary

Project Overview

One of the biggest developmental challenges facing South Africa is the high number of children who do no learn to read for meaning in the early years of school. This is the foundational skill upon which all others build and as such this has become a leading priority for the Department of Basic Education (DBE). In order to address this challenge, the DBE initiated the Early Grade Reading Study (EGRS) in collaboration with academics at the University of the Witwatersrand, the Human Sciences Research Council (HSRC) and Georgetown University (USA). This is a large-scale educational impact evaluation – the biggest in South Africa - and aims to build evidence about what works to improve the teaching and learning of early grade reading in African languages in the country.

The core of the project is a comparison of the cost-effectiveness of three promising intervention models to improve reading outcomes in learners' home language (Setswana). The project commenced in 2015 by working in 230 quintile 1-3 schools in the North West province. Each intervention has been implemented in a separate group of 50 schools with a further 80 control schools where ordinary schooling is continuing. The project uses a formal impact evaluation methodology known as a Randomised Control Trial (RCT) complemented with a 60-classroom observation study and eight detailed case-studies. The study design enables the researchers to estimate the impact of each intervention model on measures of reading, as well as understand where, how and why different elements of the intervention models are working.

The evaluation assessed three interventions:

1. **A structured learning programme & centralised training**: The first intervention provides teachers with lesson plans aligned to the National Curriculum Statement Grades R-12 (NCS) including the Curriculum and Assessment Policy Statements (CAPS), as well as additional quality reading materials and training at centralized workshops twice a year.
2. **A structured learning programme & specialist on-site coaching**: The second intervention (implemented in a different group of 50 schools) provides teachers with the same set of lesson plans and reading materials but provides ongoing support to teachers through on-site coaching and small cluster training sessions.
3. **Parental intervention:** The third intervention (implemented in a further 50 schools) holds weekly meetings with parents to discuss the importance of learning to read in the early grades and to empower them with the knowledge and tools to become more involved in their child's literacy development.

The three interventions were implemented in the grade 1 class of 2015 and at the grade 2 level in 2016, thus following the same cohort of learners. This year (2017) the two structured pedagogic interventions have continued at the grade 3 level, thus ensuring that this cohort of learners were exposed to the interventions for the entire Foundation Phase.

Three waves of data have been collected to date. A baseline data collection ("Wave 1") was collected at the start of 2015 when learners had just begun grade 1. A midline data collection

("Wave 2") was collected at the end of 2015. A third wave of data was collected at the end of 2016, when most learners were in grade 2. Data collected towards the end of 2016 when the learners had received two years of the interventions forms the basis for the evaluation findings presented in this report.

Year 2 evaluation findings: What works?

Of the three intervention models we have been evaluating, the Coaching intervention is showing a substantial positive impact after two years of intervention (end of grade 2). This intervention included lesson plans, reading materials and on-site coaching by reading experts. **Learners who received two years of this Coaching intervention were approximately 40% of a year of learning ahead of the students in the schools that received no intervention ('business-as-usual' schools).** This is a truly significant improvement by international standards. The other two interventions (centralised Training; and the parent involvement intervention) appeared to have a small positive impact, less than half the size of the coaching intervention.

We measure the impact of the Home Language Literacy interventions on children letter recognition, word recognition, non-word recognition, paragraph reading (oral reading fluency), phonological awareness, comprehension, writing and two additional school subjects, English and mathematics – in case there were spillover effects.

Although the Training intervention had moderate positive effects on some of the sub-tests, the Coaching intervention registered statistically significant positive effects on all home language literacy measures, with similar effect sizes across the sub-tests. There was no significant effect of the coaching intervention on the short mathematics test that was administered. This means that we have no evidence of a negative effect through crowding out of teaching time for mathematics. Interestingly, we observe a significant positive effect on English. This might be attributable to an improved underlying language ability (obtained through the home language intervention) or simply due to improved classroom management and transferable instructional methods acquired by the teacher through the coaching intervention. Either way, this is an encouraging finding for the Coaching intervention.

Although the overall impact of the parent intervention was small, it does appear to have had a significant positive impact on phonological awareness. This was probably the specific reading skill that was most directly targeted through the parent meetings. Sound games were a key method taught to parents to use at homes in the development of their child's phonological awareness.

Who benefits most from the interventions?

Boys catch up somewhat: The effective Coaching intervention is helping boys catch up some of the way to girls. Although girls still perform better than boys in the “Coaching” group, the gap is smaller than it is in the control group.

Impact concentrated in urban schools: For all three interventions, the observed impacts are larger in urban township settings, but there is no measurable impact in deep rural settings. This means that we may need to approach interventions in rural schools differently.

Middle-to-top learners benefited most: The impact of Coaching is largest for children in the middle and upper part of the achievement distribution with small or negligible impacts for the weakest children. Importantly, there is no evidence of a negative effect for any part of the performance distribution. One implication of this finding is that structured pedagogic programmes that make use of lesson plans may benefit certain groups of children more, depending on the level at which the lessons are set.

Large-classes benefited most: Both the teacher support interventions (“Training” and “Coaching”) had the largest impacts in relatively large classes (38 to 45 learners). In smaller classes, it may be that teachers in the control schools are already able to effectively manage classrooms, provide structured learning and differentiated attention to a variety of learners. However, in larger classes the EGRS interventions helped teachers to provide better instruction in a challenging environment. Both of the pedagogic interventions emphasized good classroom management practices such as how to reorganize classrooms, work in small groups while keeping the larger classroom occupied and bring routines and predictability to the classroom. However, in the very largest classes (50 plus learners) the impact of EGRS interventions was smaller, possibly indicating that beyond a certain threshold it remains difficult to conduct effective teaching. This emphasizes the need to eliminate excessive class sizes (50+) in the Foundation Phase.

How much did teaching practice and parent behaviour shift in response to EGRS interventions?

Through the use of mixed methods research (teacher questionnaires in all 230 schools, lesson observations in 60 schools and a set of detailed case studies), we investigate underlying change mechanisms by observing how the learning environment, teaching practice, and classroom activities changed as a result of the programmes.

If any teacher support programme is to be effective, teachers themselves need to feel positively inclined towards the programme. Based on questionnaires administered to teachers in all groups of schools, teachers in the Coaching schools were considerably more likely to report feeling a high level of professional support than those in the control schools, with teachers in the Training group also somewhat more likely to experience high professional support.

Two other results are worth emphasizing. First, even though there is no large difference in access to graded readers, the lesson observations reveal that **far more learners are actually reading graded readers in the Coaching and Training schools**. **This increase is**

substantially larger for teachers who received Coaching relative to teachers who received Training. Second, even though we find no change in the probability *that* learners practice reading in the classroom, there is a noticeable difference in *how* they practice reading: **Teachers in both Training and Coaching groups are more likely to do group-guided reading, resulting in more opportunities for learners to receive individual attention. The impact is, again, larger for teachers who received Coaching relative to Training**. These results suggest that there are some teaching practices such as group-guided reading that are difficult to enact and require additional development to be effective. They also reveal an important interaction between resources and teaching practice: graded readers are only useful if teachers have developed the skills to use them effectively in the classroom.

Low attendance was a major limitation in the Parent intervention. In 2015, just over a third of parents attended at least three sessions while in 2016 just under a third attended at least three sessions. Nevertheless, compared to the control group, parents in this intervention group reported attending a significantly higher number of parent meetings at their school. However, no other indicators of parental involvement in home reading or educational activities shifted substantially, confirming that there was no large change in parental behavior in response to the intervention.

Recommendations and DBE Plans for moving forward

1. **The DBE should take steps to provide Foundation Phase teachers with a curriculum-aligned structured learning programme using lesson plans and integrated reading support materials.**
    a. Although the learning programme implemented in the EGRS was clearly more effective when supported by on-site coaching, the fact that intervention 1 had some impact suggests that the learning programme itself (lesson plans and integrated materials) can offer benefits to curriculum delivery.
    b. In order for lesson plans and additional integrated reading materials to be provided to all language groups and into the future, the DBE should set up processes to ensure that lesson plans and materials are revised and approved based on effective quality assurance activities.
2. **As far as possible, schools using the structured learning programme should be supported by on-site specialised reading coaching.**
    a. The main finding of the EGRS is that a structured learning programme aligned to the NCS, together with additional high quality reading support materials (graded reading books, flash cards, posters), can make a significant difference to learning outcomes, if accompanied by effective and carefully monitored support to teachers in the form of on-site coaching.
    b. Coaching is the best alternative: Whereas previously very little evidence existed about effective large-scale teacher support modalities in South Africa, we now have evidence that on-site coaching to Foundation Phase teachers can shift learning outcomes, and that this is a cost-effective strategy. Modelling of lessons, in a safe space, as they navigate the lesson plans for teaching learners to read is critical.

c. Direct in-service training better than 'train-the-trainer' models: Direct in-service training of teachers (4 two-day workshops over the course of 2 years), while less effective than on-site coaching, is in turn likely to have more impact than "cascade" models where specialists "train the trainers" who then interact with teachers.
d. Existing subject advisers cannot fulfil the role of a coach: The low ratio of subject advisors to schools (especially in the Foundation Phase) makes it impossible for subject advisors to fulfil the role of reading coaches, as implemented in EGRS; nor do we recommend increasing the number of subject advisors to allow this since the recruitment process, oversight structures and modus operandi of the coaches is different to that of subject advisors.
e. In order to realize effective on-site coaching, the DBE should develop guidelines for on-site coaching as well as institutional support for the learning programme provided by subject advisors and curriculum leaders within schools.
f. Prioritizing schools is a viable option: On-site coaching interventions could be implemented in priority schools (e.g. 100 or 500 schools in a province) on a temporary basis (e.g. 2 years at a time) and through independent contracting and oversight structures. The cost for 100 schools would be about R6 million at current prices.

3. **Provinces should ring-fence finances for the implementation of the structured learning programme and on-site coaches**
   a. A finance review should be commissioned of the cost implications of the implementation of the structured learning programme and the on-site coaching. The review should provide the detailed costs involved of implementing the programme in a province, as well as investigate potential ways in which provinces could fund the implementation of the programme.
4. **Large-scale implementation should be immediately pursued in the districts of Ngaka Modiri Molema and Dr Kenneth Kaunda.**
   a. While systems are being set up to facilitate widespread implementation of the structured learning programme and on-site coaching, further implementation should build on the momentum created by the EGRS in Ngaka Modiri Molema and Dr Kenneth Kaunda.
   b. All schools in these districts should receive the lesson plans and integrated reading materials together with a short 1-day orientation training workshop. Subject advisors should receive training on the programme so that they are equipped to provide instructional support to teachers.
   c. A select group of schools should be identified to receive on-site coaching.
   d. The expanded implementation in the North West should continue to be evaluated to establish whether the same level of impact can be achieved on a larger scale.
5. **Further research is needed to build on the initial findings of the EGRS.**
   a. Develop reading norms in the African languages: Reading norms cannot simply be adapted from one language into another due to differences in language structures. It is a complex exercise requiring longitudinal data. Therefore, the EGRS data could be used towards the development of reading norms in the African languages.

b. Parental involvement needs further research and may be promising: Whilst parental involvement is a hugely deterministic factor in a child's learning outcomes, the biggest challenge from a policy perspective is how to shift parent involvement at scale. Given the potential cost-effectiveness of such interventions, researchers and policy-makers should continue to investigate mechanisms to do so.
c. Learning what works in deep rural settings: Formative research and subsequent impact evaluation is required to figure out what kinds of school support programmes make a meaningful difference in deep rural settings.
d. Measuring long-run EGRS impacts: The DBE is planning to administer subsequent data collections on the same sample of learners to measure the long run impacts of these reading interventions. It remains an open question whether a home language reading intervention that impacted on reading proficiency in the early grades will yield a persistent benefit to those learners who were impacted. It is conceivable both that the effect may dissipate over time or that it may compound if a solid reading foundation enables more learning later on and in different subject areas. It may be that this will depend on the quality of teaching received in subsequent years.
e. EGRS for EFAL in Mpumalanga: A second phase of the Early Grade Reading Study (EGRS 2) is underway in the Mpumalanga province, since the start of 2017. This project aims to investigate the effectiveness of two alternative interventions on English as First Additional Language in the Foundation Phase.
f. Early Grade Mathematics Study: The DBE is embarking on an Early Grade Mathematics Study over the next 2 to 5 years, with the first activity being a detailed scoping study to identify and design promising interventions with strong theories of change as well as cost-structures that would be sustainable on a large scale.

1. Introduction

The acquisition of reading is foundational to all subsequent learning; yet the majority of children in the developing world are being left behind in this regard. In India, for example, less than half of grade 5 pupils can read a grade 2-level story (Banerji, Bhattacharjea & Wadhwa, 2013: 392). Similarly flat learning trajectories have been documented across sub-Saharan Africa and south Asia (World Bank, 2018). South Africa is not unique in this regard: results from international assessments of reading and literacy have revealed that the majority of children in grades 4, 5 and 6 have not yet learned to read with comprehension.[1] These children, who have not learned to read, can thus not read to learn in subsequent grades and in all their subjects. Since reading is a gateway to future learning, addressing these shortcomings should be a policy priority.

A large body of evidence suggests that teachers play a critical role---perhaps the most critical--- in shaping how much a child learns in a year and his/her future productivity.[2] Yet, across the world, teacher quality is highly variable. In recognition of this, governments and donors invest billions of dollars annually with the hope of improving teaching practices of the existing pool of teachers - by some estimates the United States spends 18 billion annually on teacher professional development (Fryer, 2017); and Popova, Evans and Arancibia (2016) calculate that nearly two thirds of World Bank-funded education programs include a professional development component - but with disappointing results. For example, many studies in the United States have found no impact of professional development programs on student learning, especially when conducted by government at scale;[3] and a recent meta-analysis of evaluations of in-service teacher training programs in developing countries concluded that “teacher training programs vary enormously, both in their form and in their effectiveness" (Popova, Evans & Arancibia, 2016: 4).

A plausible reason for the failure of many training programs is that they commonly focus on changing knowledge, yet teaching is a skill that needs to be developed through ongoing practice (Kennedy, 2016). New teaching techniques are often difficult to enact at first, but can become easier over time. Resistance to adopting new techniques is amplified by the fact that it requires abandonment of an old approach that teachers were comfortable in doing.

A potentially cost-effective way to encourage enactment of difficult teaching techniques is to combine training with structured lesson plans. Lesson plans reduce the cost of transition since they require no additional lesson preparation. They prompt certain behavior, which serves as useful reminders. And the structure also provides a regular routine, which facilitates practice and fosters habits. In the United States, for example, “off the-shelf' lesson plans for mathematics teachers were found to improve students' math achievement (Jackson & Makarin, 2016).

---

[1] The Progress in International Reading Literacy Studies (PIRLS) of 2006, 2011 and 2016 indicate that nearly 80% of grade 4 children in South Africa have not yet learned to read with meaning. The SACMEQ study of 2007 indicated that a large proportion (even by the standards of the region) of grade 6 South African children were still “functionally illiterate”.

[2] E.g. Rivkin, Hanushek & Kain (2005), Bau and Das, (2017); Araujo, Carneiro, Cruz-Aguayo & Schady (2014).

[3] Harris & Sass 2011; Jacob & Lefgren 2004.

But training with lesson plans might not be sufficient to change behavior without ongoing observation and feedback from someone who has already mastered the skills - a coach. Some techniques might remain too difficult or arduous to implement. Worse, the new technique might be incorrectly applied and bad habits could form. An expert coach, who observes the teacher in the classroom, can provide feedback to rectify mistakes and improve on techniques. The coach can also demonstrate correct teaching techniques, and provide additional motivation for the teacher to change behavior.

Recently there has been promising evidence that a bundled intervention – of providing teachers lesson plans and learning materials, combined with training and reading coaches who visit teachers on a regular basis to observe teaching and provide feedback – can lead to large improvements early-grade reading. Piper, Zuilkowski & Mugenda (2014), for example, find that the Kenyan PRIMR programme consisting of daily lesson plans, instructional materials, 10 days of teacher training as well as on-site support led to substantial gains in both English and Kiswahili at the grade 1 and 2 level. Piper & Korda (2011) report on the EGRA plus programme that was implemented in Liberia. This programme included lesson plans, reading materials, on-site coaching, as well as the provision of school performance information to parents. Huge effect sizes were reported in the evaluation, but it remains difficult to isolate the specific programme components driving impact. Lucas, McEwan, Ngware & Oketch (2014) report on the “Reading To Learn” intervention which was rolled out through a randomized experiment in both Uganda and Kenya with effect sizes of 0.2 SD and 0.08 SD respectively. This intervention used a structured learning approach supported by additional reading materials, teacher training and coaching as well as encouraging School Management Committees to prioiritize the early grades.

But there is much we do not know. Many of the before-mentioned programmes were bundled sets of interventions, so we do not know which components were responsible for the success of the program. In particular, we do not know if a more cost-effective version without the reading coaches can be more cost-effective. We also know very little about the mechanisms underlying the success of the programs: i.e. how did teaching practice change? Whilst the design of the EGRS does respond to the need to isolate the specific mechanisms driving change, we also acknowledge that educational programme components may have important complementarities, as argued by Kerwin and Thornton (2018). Therefore, the two structured learning interventions evaluated in EGRS do contain complementary inputs but also vary in terms of the modalities for professional development to teachers.

In South Africa, although there have been and are various initiatives underway to support early grade reading in South Africa, there is little or no sense of what is working and why. Intervention programmes lead by the Department of Basic Education (DBE), provincial departments or other partners such as the National Education Collaboration Trust (NECT), are typically not set up with an impact evaluation design in mind. Moreover, there are competing models of support in the system. Some initiatives use so-called cascade training models where district officials are orientated to a particular programme or new set of resources, and they in turn train teachers on a decentralized basis. Sometimes, conventional teacher training workshops are held. Since 2010, the Gauteng Primary Literacy and Mathematics Strategy (GPLMS) provided additional graded reading booklets and scripted lesson plans to teachers. The programme also featured regular on-site visits from specialist reading coaches to observe classroom practice and offer assistance.

Beyond teachers, a key ingredient to acquisition of reading proficiency is clearly the contributions made by the parents of a child. The challenge in many developing countries is that parents themselves struggle with weak literacy, so how can one expect an illiterate or semi-literate parent to contribute to the reading acquisition of their child? Rigorous evaluations have demonstrated the challenge in promoting parental involvement (e.g. Lieberman, Posner, and Tsai, 2014). Yet, the role of parents cannot be ignored; and there are many foundational components to reading acquisition (such as vocabulary development) which does not require a literate instructor.

Is centralized training combined with structured lesson plans sufficient to change teaching practice and ultimately improve pupils' early-grade learning? Or is it more cost-effective to have reading coach who visits schools, observe teaching practice, and provide feedback. Which method – training or coaches - is more cost-effective, and for which types of learning outcomes? Can training and motivating of parents be a more cost-effective alternative to improving early grade-reading? Does this depend on the characteristics of the pupil, school, community or teacher?

To answer these questions we conducted a field experiment in 230 poor public primary schools in one province in South Africa, comparing three different approaches to improving early-grade reading. The first two programs (50 schools per treatment) compare two different forms of teacher professional development: training of teachers at a central venue, or ongoing observation and feedback from a reading coach. The third program trains and motivates parents to contribute towards their child's learning.

Our pre-analysis plan was written prior to the analysis of the endline data (28 October 2016) and registered at the American Economic Association RCT registry. The results discussed in this report do not deviate from the pre-analysis plan.

The report has the following structure: section 2 describes the program, theory of change and research hypotheses; section 3 provides more detail on the context; section 4 provides a timeline; section 5 discusses the evaluation design; section 6 discusses the program or policy; section 7 reports our main results; section 8 provides a discussion; section 9 describes ongoing and planned policy engagement to ensure that lessons from this evaluation are implemented; and section 10 ends with policy recommendations.

2. Intervention, theory of change, and research hypotheses

Intervention

This study evaluates three different interventions, all aimed at improving reading and literacy in the home language, which in the case of the North West province is Setswana. The beneficiaries of the interventions were a cohort children entering grade 1 at the start of 2015 over a two-year period (thus working with grade 2 learners in 2016). The project has now been extended for a third year of interventions at the grade 3 level in 2017. However, this extension does not apply to the parent involvement intervention.

Intervention 1: Structured lesson plans, additional reading materials + central training

Intervention 1 provides teachers with daily lesson plans, which are aligned to the curriculum as specified in the Curriculum and Assessment Policy Statements (CAPS) for home language literacy in the Foundation Phase. The four learning areas in the curriculum for grades 1 to 3 are Home Language literacy, First Additional Language (which is usually English), Mathematics, and Life Skills. The lesson plans are thus intended to strengthen the enactment of the curriculum and should not be seen as an alternative to current policy. They provide detailed specification for each lesson including information on methodology and content to be taught for each instructional day. The lesson plans incorporate the use of learning support materials including the government-provided "DBE workbooks" as well as certain additional materials (graded reading booklets, flash cards and posters), which are provided through the EGRS. Standard practice in schools not receiving daily lesson plans would be to conduct curriculum planning with other teachers in the school and then to devise their own daily lesson plans. The intention of these planning processes would be to ensure that daily practice facilitates the delivery of the official curriculum, but to the extent that this is not done effectively the provision of lesson plans can be expected to improve curriculum coverage, sequencing and pacing.

The graded reading booklets provide a key resource for the teacher to use in group-guided reading and individual work so as to facilitate reading practice at an appropriate pace and sequence of progression. EGRS provided the Setswana "Vula Bula" graded reading book series developed by the Molteno Institute for Language and Literacy. These books were developed in the relevant African languages as opposed to being translated, and progress in accordance with the natural phonic progression of each language. The group-guided reading activity is prescribed in the official curriculum but is difficult to enact well, and is therefore often completely absent in classroom practice. It is designed to create opportunities for individualized attention and reading practice. Additionally, it may provide opportunities for peer learning amongst children, but this is not an explicit goal of the activity.

Intervention 1 trains the teachers on how to use the lesson plans and accompanying materials through central training sessions, each lasting 2 days, and occurring twice yearly. These sessions were conducted for grade 1 teachers in February and July of 2015 and for grade 2 teachers in January and July of 2016. Similar sessions occurred during 2017 as the project was extended into grade 3. The choice to conduct two-day residential training was informed by what was typically being implemented in South Africa. Some international studies showing evidence of impact used more training days (e.g. 10 days in the Liberian EGRA Plus project), but this would have been unaffordable and impractical in the South African context. In one sense 4 days of training per year is therefore a relatively "light dose", but on the other hand the fact that it was residential training as opposed to afternoon sessions, meant that there was more dedicated time available and perhaps better attendance.

Intervention 2: Reading Coaches, scripted lessons, graded readers.

Exactly the same set of instructional materials (structured lesson plans, graded reading booklets and other materials) is provided to Intervention 2 schools. Therefore, if the lesson plans are implemented with the same level of fidelity across Interventions 1 and 2, classroom practice and hence learning outcomes should be identical across the two groups. However, the modality of supporting teachers differs. Instead of bi-annual central training sessions, ongoing support to teachers consisting of regular (monthly) on-site coaching from specialist "reading coaches" is provided. In addition to these on-site visits, there are occasional meetings with the coach and a small cluster of nearby Intervention 2 schools. The ratio of coaches to schools was roughly 1:16, which is similar to what has been found to have been cost-effective in a Kenyan experiment (Piper, Zuilkowski, & Mugenda, 2014). The evaluation of Interventions 1 and 2 should thus shed light on a) whether this structured lesson programme can improve the enactment of the curriculum and thus improve reading acquisition, and b) whether the mode of teacher support is important in determining effective enactment.

There are various components that should take place during the on-site coaching visits. Firstly, the coach is to observe various protocols upon arrival at the schools including reporting to the reception area, signing into the school's administration book and greeting the principal. Secondly, the coach should observe the teacher's classroom – taking note of whether there is a reading corner, how the classroom is set up, whether there is learning material on the walls and the general condition of the classroom. While the teacher is presenting their lesson, the coach should take note of classroom management and discipline, the teacher's preparation level, the teacher's pacing during the lesson, learners engagement throughout the lesson, and the teachers use of strategies and methodologies set out by the EGRS programme. The coach will also monitor the learner's books, ensuring that their written work is at the correct standard and that assessment has been done accurately. After the lesson the coach facilitates a reflection discussion with the teacher about the lesson. Before the coach gives his/her comments, he/she will probe the teacher by asking how they felt the lesson went. Thereafter the coach will provide his/her observations, focusing both on positive aspects and identifying areas for improvement. The coach and the teacher will then set targets for the teacher to achieve by the next lesson observation.

The afternoon cluster sessions are important because they allow for teachers to share their successes with one another – giving teachers the opportunity to take note of good reading practices to implement in their own classrooms. During the afternoon workshops, coaches are also able to focus on improvement strategies regarding issues that many teachers may be struggling with in the lessons that were observed.

Intervention 3: Parental involvement

Intervention 3 is designed to promote parental involvement to support their children's reading progress. At each of the 50 schools in this Intervention group a Community Reading Coach (CRC) was recruited. The CRC was identified through communication with the school principal who recommended a suitably qualified and available person in the community. The CRCs attend a 1-day training session facilitated by the service provider (Class Act) at the start of each school term (quarterly). The CRCs are trained to deliver weekly training sessions for parents at their respective schools. For their services, CRCs are paid a stipend of R400 per month (about $35). Under this arrangement, CRCs are essentially volunteers receiving a small stipend, rather than employees receiving a salary.

A total of 30 sessions is scheduled for each year covering a total of 10 topics per year. Each topic has 3 sessions where the topic is the same but the activities of the session differ. Thus a parent can attend roughly 1 in 3 sessions and still be exposed to all topics, while parents who attend more regularly can still enjoy fresh activities. The topics covered in these sessions include the importance of learning to read for later educational and labour market success, training on how to support their child's reading at home and the provision of low-cost materials and reading games to use at home. As with Interventions 1 and 2, grade 1 parents were invited in 2015 and grade 2 parents in 2016. It should be noted that the design of this parent involvement programme errs on the side of being high dosage (weekly meetings) and comprehensive (many topics are covered). During the formative research phase it was debated whether a more specific change mechanism should be targeted through this intervention so as to build on similar international research. However, the comprehensive approach was favoured in light of the absence of any evidence of effective parent involvement programmes in South Africa. In retrospect, holding weekly meetings may have had the unintended effect of lowering attendance due to it such frequent attendance being perceived as un unrealistic goal.

Theory of change

Reading acquisition

All three interventions relate to the educational theory of how reading acquisition occurs. An effective reader is one who reads with rich comprehension and engagement with the substance of the text. Reading comprehension is the product of two components: vocabulary and decoding.

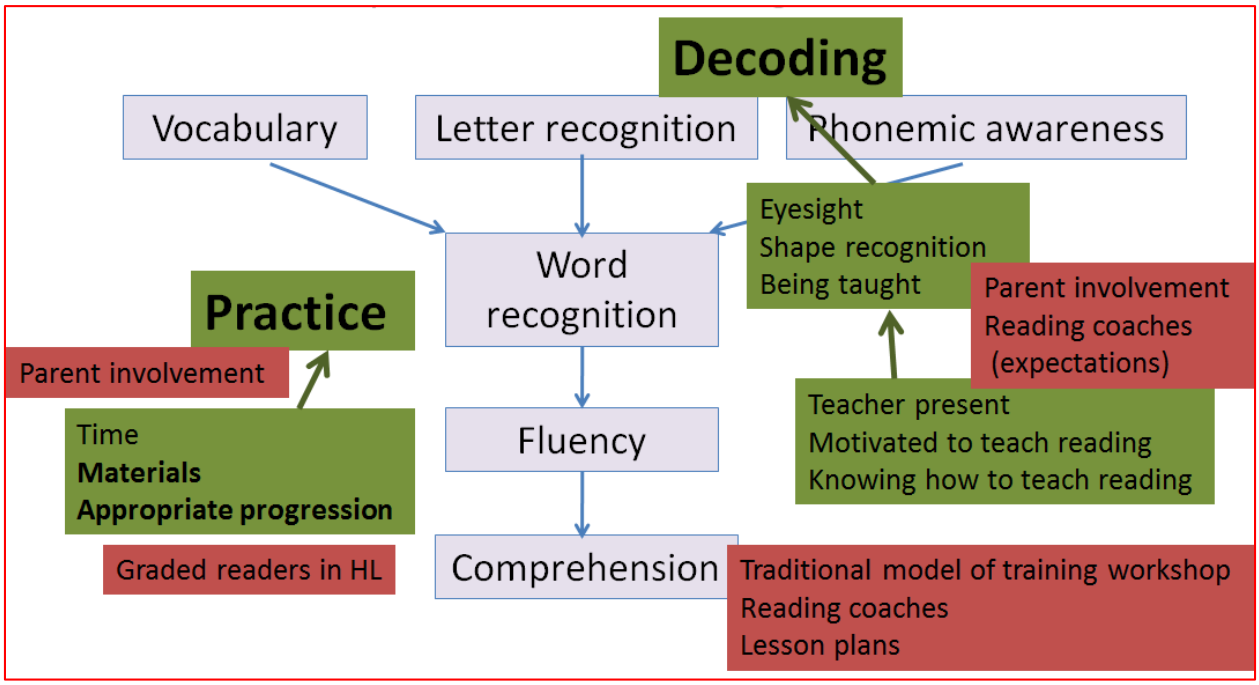
To a great extent vocabulary (and language acquisition in general) comes naturally through hearing others speaking and then emulating this. Through speaking and hearing others speaking, phonological awareness also develops - this involves sound segmentation and recall of sound patterns. This phonological awareness is important for children to learn to decode since written symbols are associated with particular sounds. Decoding thus consists of letter recognition and phonemic awareness.

Unlike learning to speak, decoding does not come naturally; it is a method that must be taught systematically. It is important to emphasize that reading is produced by the product of vocabulary and decoding: If one has a perfect vocabulary but has not been taught the method of decoding one will not be able to read at all. Letter recognition and phonemic awareness are mastered through systematic teaching and consistent practice. This leads to the next stage of reading acquisition: word recognition. Through practice and appropriate progression from simpler sounds and words to more complex ones, word recognition becomes established leading to the next phase of reading acquisition: fluency. It is only once decoding and word recognition have become fluent, even to the point where it becomes automatic[4] and unconscious, that it is possible to reach the ultimate goal of reading comprehension. The strong empirical relationship between oral reading fluency and comprehension demonstrates this point.

[4] For example, do not read this word: "exactly".

In order to learn the basics of decoding, a child requires a teacher who is present, capable and motivated to deliver systematic reading instruction. In order for decoding to become fluent a child requires suitable graded materials and the discipline (perhaps imposed) to practice a lot. The interventions to be tested in this study address these needs in various ways. Figure 1 presents a theoretical diagram illustrating how reading acquisition occurs, what supportive conditions need to be in place and how each of the interventions being evaluated in the EGRS address key stages in the development of reading acquisition.

Figure 1: Theoretical diagram of how reading acquisition occurs



There is a growing body of evidence from developing countries that early grade reading interventions can have a significant impact. The "EGRA Plus" programme administered in Liberia, for example, produced substantial gains in reading achievement relative to comparison children who did not receive the programme. Key aspects of this programme included a cascading model of reading coaches, the distribution of scripted lesson plans and reading assessment tools, and the dissemination of report cards to parents (Gove and Wetterberg, 2011).

A number of international studies using structured learning programmes and coaching were briefly described in Section 1. However, these studies often do not tell us which component of the intervention was responsible for the success of the program. This is important for policy purposes, because we want to find the most cost-effective intervention which could be scaled up by government. For example, the "EGRA plus" programme in Liberia was clearly highly resource-intensive because it required ongoing monitoring from qualified reading coaches, but we do not know if one might be able to reach the same results with a sub-component of the program. Moreover, there is uncertainty about the transferability of the findings given different language and social contexts.

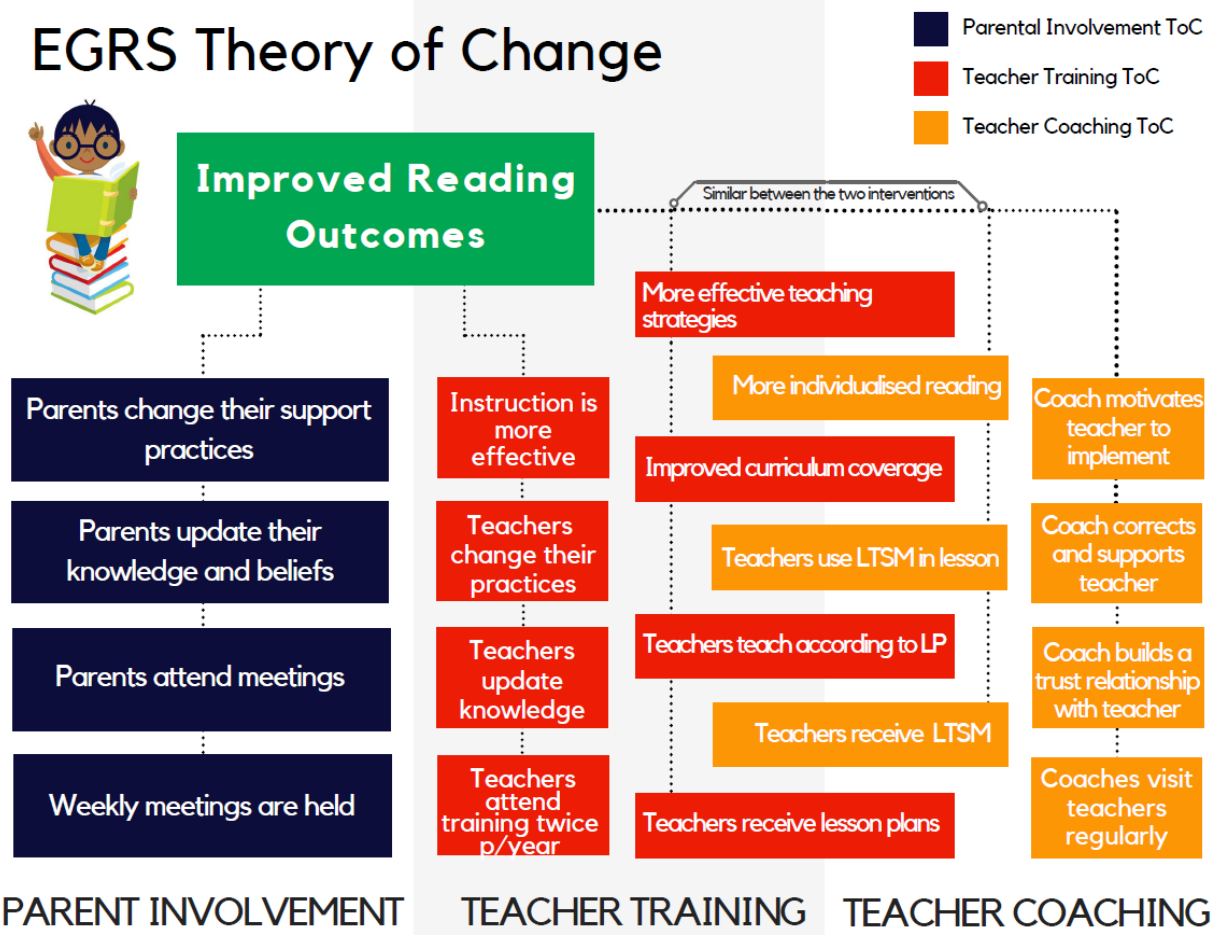
Similar programs have been implemented in South Africa, but since they were not credibly evaluated, we do not know if they truly improved pupils' reading acquisition. The Department of Basic Education typically holds training programs similar to our intervention 1; and Gauteng has implemented a model of reading coaches, similar to intervention 2. Since it has not been possible to produce a robust empirical impact evaluation of these programmes, we have little knowledge about their effectiveness. Fleisch and Schöer (2014) attempted a Regression Discontinuity Design (RDD) to evaluate the impact of the Gauteng Primary Language and Mathematics Strategy (GPLMS) and findings pointed to a positive impact, though the findings were tentatively made given significant data constraints. Sailors et al (2010) evaluated a reading intervention in South Africa, which followed a similar model to intervention 2, but there are considerable methodological limitations to this study.

There have also been several experiments around the world that have provided information to parents with the goal of fostering parental involvement and thus improving learning outcomes. But there is much we still do not know. In Pakistan, pupils who came from villages where the community was provided with information of school performance performed better in independently administered tests, compared to pupils from villages where no such information was administered. The improvement was particularly large for schools with low initial learning outcomes (Andrabi, Das, & Khwaja, 2015). In a different programme in India, school communities were informed of their school performance and also educated on their rights, roles and responsibilities in school governance through 8 public meetings. Education performance improved as a result (Pandey, Goyal, & Sundararaman, 2009). However, in a recent impact evaluation in Kenya, informing parents on their child's reading progress had zero impact (Lieberman, Posner and Tsai, 2013). The authors hypothesize necessary conditions for an information-intervention to work, all of which we address in our study: (i) information is new; (ii) it highlights under-performance and potential to improve; (iii) it is combined with measures which enable parents to act on this information.

All interventions aim to improve reading acquisition in the home language. Strictly speaking, the targeted outcome is home language literacy more broadly, since this is the Foundation Phase curriculum area being given support through our programmes. The choice to address home language literacy is motivated by research showing long-term benefits to strong home language skills prior to switching to a second language. Taylor and Von Fintel (2016), for instance, show that in South Africa using home language as the language of instruction during grades 1, 2 and 3 has been associated with better English acquisition in grades 4, 5 and 6.

Figure 2 presents a summary of the theories of change for each of the three EGRS interventions in a diagram. A detailed discussion of each theory of change follows.

Figure 2: EGRS Theory of Change



Intervention 1: Teacher Training

There are a number of reasons why one might expect a structured learning programme using scripted lesson plans to improve instructional practice and learning. Firstly, we know that curriculum coverage, pacing and sequencing is currently inadequate in the majority of South African classrooms. The National School Effectiveness Study revealed this through a learner exercise book review conducted in a large sample survey (Taylor, Van der Berg, & Mabogoane, 2013) while classroom observation studies have unpacked this in greater detail (Hoadley, 2010). A structured learning programme clearly has the potential to improve curriculum coverage, pacing and sequencing.

Secondly, the use of lesson plans can facilitate the adoption of new methods by teachers and thus expand their own repertoire of instructional practices. Most teacher training interventions implicitly assume that changed knowledge will lead to changed practice in the classroom. However, the use of lesson plans allows one not to rely completely on this assumption. When a lesson plan prescribes the use of a certain instructional method, the teacher may implement that method even though she may not yet possess a deep understanding of the rationale behind the method. Through the regular practice of that method, however, the teacher's knowledge may be enriched as they begin to see the method's effectiveness. In this way there is an iterative relationship between "knowing" and "doing" in which improved classroom practice emerges.

A third reason to expect lesson plans to improve classroom practice is that they integrate the effective use of reading materials. Van der Berg (2008) made the case that additional school resources often make no impact because they are not well managed by schools. An HSRC study of grades 1 – 4 classrooms in 20 Limpopo schools found that little reading activity occurred, that the use of texts was limited and that when reading was taught the predominant activity was the teacher reading to the class (Prinsloo, 2008). One of the most important national interventions over the last few years has been the provision of the so-called "DBE Workbooks". These colourful books are a type of hybrid between a textbook and an exercise book, with lots of exercises for learners to complete in the books themselves. The lesson plans provided through EGRS incorporate the DBE Workbooks into the daily lessons referring to specific page numbers for exercises to complete. The additional reading materials provided through EGRS (posters, flash cards and graded reading books) are similarly integrated into the structured learning programme through the daily lesson plans. Several reports bemoan a lack of African language reading materials in Foundation Phase classrooms (NEEDU Report, 2012; Ministerial Audit of provincial reading programmes, 2012). Learning to read requires practicing and gradually moving from simple language structures to more complex letter blends and words. The language structures differ across languages and translations of reading booklets is therefore inappropriate. Interventions 1 and 2 in the EGRS fill this gap by providing sets of Vula Bula graded reading booklets developed in Setswana, and by promoting their effective use through the lesson plans.

Yet there are potential negative (or perhaps ambiguous) consequences of following a prescriptive set of lesson plans. As is the case in many developing countries, South African classrooms often comprise a wide range of learner proficiency levels. For example, in our control group sample of 80 schools, nearly 40% of grade 2 children could not read a single word in the paragraph reading test; yet about 25% of children could read at least 50 words in a minute. And this sample already excludes the two most affluent "quintiles" of schools, making it a *relatively* homogeneous sub-set of South African schools. Some might argue that scripted lesson plans could reduce teacher autonomy to differentiate the level of instruction to meet the variety of needs present within the classroom. If this is the case, the structured learning programme may benefit a certain range of the learner proficiency distribution depending on where the lessons are pitched. This is definitely something that we are aware of and will test for. However, there is one aspect of the EGRS learning programme that should in fact promote differentiated instruction, namely that the lessons routinely use "group-guided reading" sessions. This activity, which is prescribed in the CAPS, involves a set of between 6 and 10 learners sitting with the teacher to read selected reading material. This activity promotes individualized attention to learners and thus promotes the opportunity for individual decoding as opposed to whole class reading or "chorusing" after the teacher. Also, the programme encourages teachers to group learners according to their level of proficiency, thus promoting a degree of differentiated learning.

Intervention 2: Teacher Coaching

The reading coach intervention provides basic orientation to the lesson plans and additional reading materials at the start of each term followed up by on-site coaching visits approximately once a month. The fact that the coach actually observes classroom practice makes it more likely that teachers will in

fact implement the new practices as prescribed in the lesson plans. Moreover, it promotes the correct implementation of instructional methods since the coach is able to indicate things could be done differently. In this way, the coach can guide a teacher through the process of trying a particular method, reflecting on that activity, and doing it again in an improved form. Thus, the iterative relationship between “knowing” and “doing” is strengthened in Intervention 2 by the presence of a coach who acts as a mentor along the way.

The assumption is that, just like learning to read, the ability to teach is a skill that needs to be developed over time and might not be accomplished in one-off training. Furthermore, the reading coaches could also improve teacher motivation as they are frequently monitored, provided with much-needed additional support, and can also find inspiration from watching an excellent example provided occasionally by coaches. This programme thus addresses both teacher capacity and teacher motivation. Another way to describe the difference between Interventions 1 and 2 is that while they share an underlying pedagogical theory of change (centered around instructional alignment and coherence using prescriptiveness as a vehicle), they differ in their theory of action (where Intervention 2 has a stronger component focused on changing behavior using accountability and motivation).

Intervention 3: Parental Involvement

Parents pay a critical component to learning to read, as it requires continuous practice, both at school and at home. For parents to be *willing* to play this role they need to appreciate (i) the importance of reading; and (ii) that their child is most likely not learning enough at school and requires additional support. This is the purpose of the information. For parents to be *able* to play this role, they need to understand the necessary steps in learning to read and also have appropriate material to practice reading with their child. This is the purpose of the training and additional practice material.

However, as Figure 2 above indicates, there are a number of steps along the causal chain that need to be in place before a parental involvement intervention is likely to impact on childrens’ reading outcomes. Firstly, weekly meetings have to actually be held and the content of what is presented needs to adhere to the intended content and quality. Secondly, parents have to attend the meetings that are held. If large proportions of parents do not attend or do not attend regularly the average impact across all targeted learners will be substantially diminished. Thirdly, parents who attend the meetings need to change their attitudes, beliefs and knowledge as a result. Fourthly, such changes in thinking need to translate into changes in parent behavior and practices in the home in the way they relate to their children. For example, parents need to play the literacy games that are promoted at the parent meetings with their children. Finally, these new practices need to actually have a positive effect on their children’s literacy development.

It is worth considering what schooling as usual looks like in the control group, at least with respect to the types of support targeted through the three treatments. A variety of in-service teacher training activities are conducted throughout the country but little is known about the dosage or quality of such training. According to a 2011 national School Monitoring Survey conducted by the Department of Basic Education (2015: 37), the average teacher spent 39 hours a year on professional development activities, though about half of teachers spent 12 or fewer hours. In the endline survey administered

to teachers through this project 64% of control group teachers in grade 2 reported that they had received some form of professional development support in the area of teaching Setswana home language literacy. Therefore, the schooling as usual scenario usually involves some sort of support but it is hard to know what the exact nature of this support is.

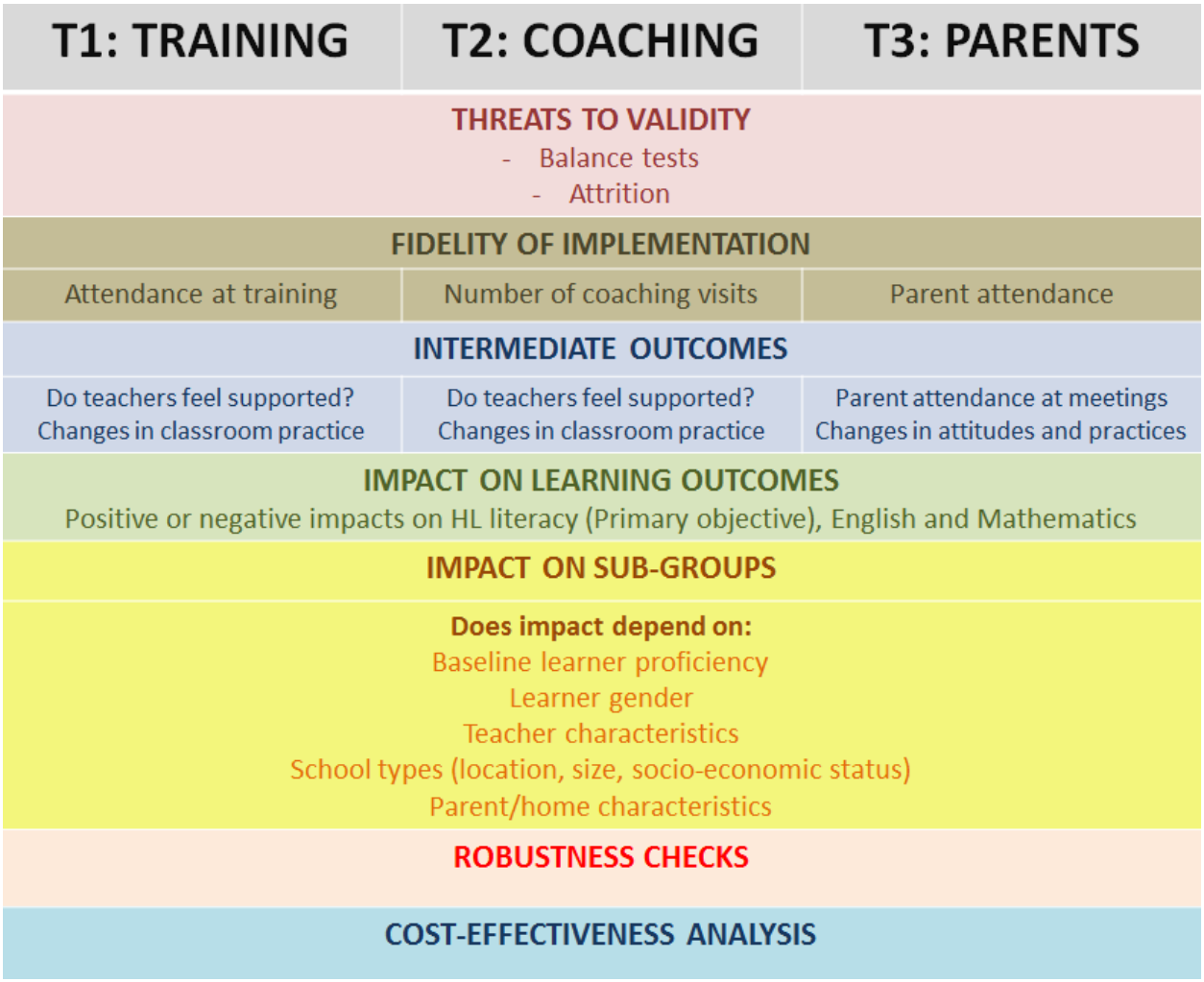
Even schools not receiving the on-site coaching intervention are supposed to receive support visits by district level subject advisors. However, the ratio of subject advisors to schools is usually far too low for these officials to play a similar role to that of the coaches who visit schools roughly once a month. As a result some teachers do not even receive a single visit in a year of teaching. According to the same 2011 School Monitoring Survey, almost half of teachers in grades 1 – 3 in the North West Province did not even receive a single visit from a subject advisor that year.

The main form of parent involvement in South African schools occurs through the School Governing Bodies, through which parents should be jointly responsible for key aspects of school governance by law. According to the DBE (2015: 43), 81% of schools had SGBs meeting minimum standards of functionality. However, there is no standard programme to facilitate greater involvement of parents or guardians in their children's actual learning.

Strategy for analysis

The Theory of Change for each intervention described above, together with the overall experimental design, informs the plan for data analysis to be presented in this report. Figure 3 graphically depicts the plan for the analysis to be presented. The logic of the sequence in the diagram is informed by the theories of change, although the order of the analysis presented in the report does not always exactly follow this logic. For example, intermediate outcomes happen in order to cause reading outcomes to change, but it made sense to first examine overall impact before exploring how intermediate outcomes may have changed.

Figure 3: Strategy for analysis



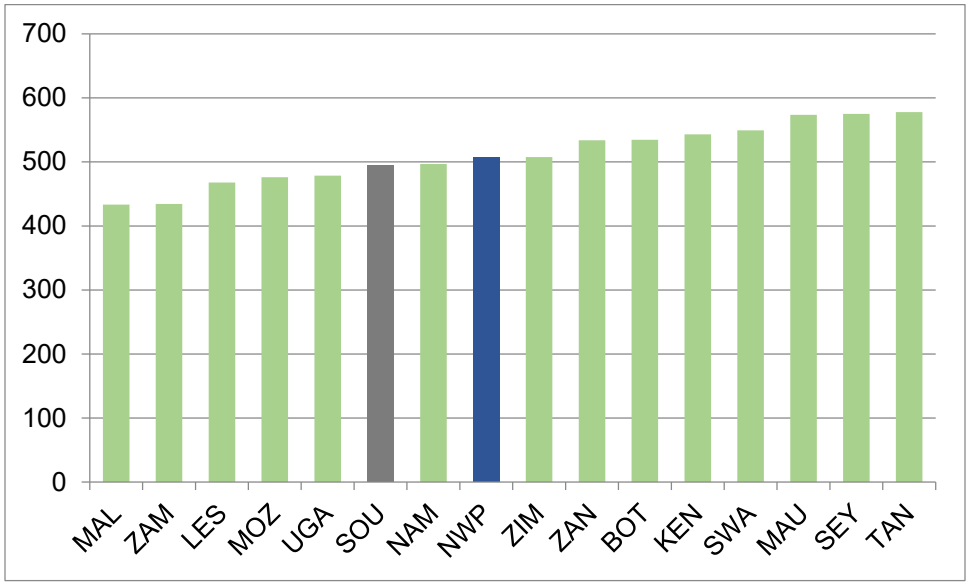
3. Context

South African children have high rates of access to school. Approximately 99% of 7-15 year-olds attend educational institutions of some kind. However, local and international surveys of educational achievement, such as TIMSS, PIRLS and SACMEQ, have shown that the quality of learning achieved by those in school is low and highly unequal across socio-economic divides. The PIRLS study of 2016 showed that a striking 78% of South African children were not yet reading with comprehension by the end of grade 4. The problem is particularly severe amongst poor children. Consequently, massive inequalities in educational achievement are established early in primary school and research shows no evidence of these inequalities being reduced in later years. These educational inequalities mirror and perpetuate the historically unjust patterns of economic inequality in South Africa. Therefore, early interventions, such as improving the acquisition of reading amongst poor children, can be expected to have important effects not only on educational outcomes but also on future economic inequality.

The EGRS has been implemented in the North West province, in the districts of Dr Kenneth Kaunda and Ngaka Modiri Molema, within schools where the main home language is Setswana. According to the PIRLS 2016 study, over 90% of Setswana-speaking children had not learned to read with comprehension by grade 4. According to the SACMEQ study of 2007, the average reading proficiency of grade 6 children in South Africa was in fact slightly below average for the Southern and East African region. Figure 4 below illustrates this as well as the fact that the North West Province registered reading achievement in SACMEQ that was only slightly above the South African average. At least as far as average reading outcomes are concerned, therefore, this research setting is therefore comparable to other countries in the region like Mozambique, Uganda, Namibia, Zimbabwe and Botswana.

The North West province was therefore chosen on the basis of 1) it being a relatively poor province, thus making it relevant to the majority of the underperforming South African school system; 2) it is relatively homogenous in terms of home language (Setswana) making it more affordable to develop learning support materials in a single language; 3) it is within driving distance from the Gauteng province where the national DBE is located; and 4) the senior management of the North West provincial education department were eager to partner with the DBE on this project. The district of Bojanala was excluded because another special targeted intervention was taking place in that district at the same time. The district of Dr Ruth Segomotsi Mompati was excluded since it is particularly far West of Gauteng and since enough schools existed in the districts of Dr Kenneth Kaunda and Ngaka Modiri Molema.

Figure 4: Average grade 6 reading scores in SACMEQ

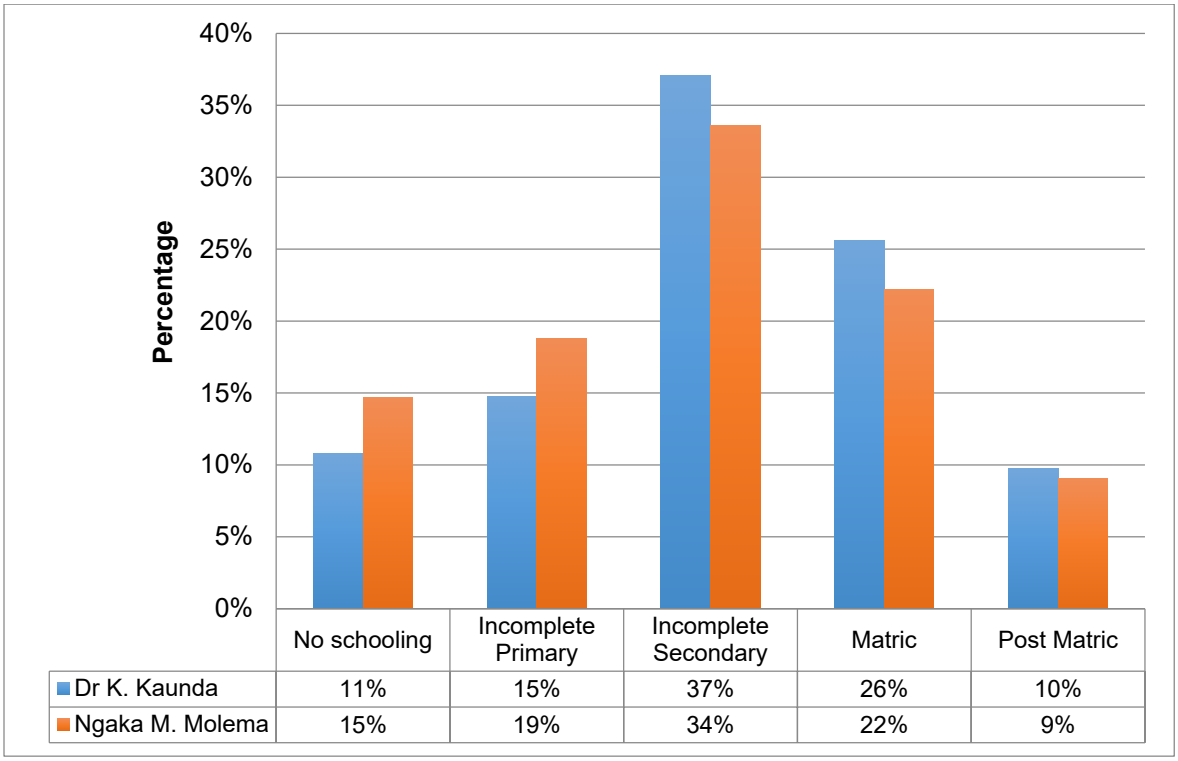


Source: Own calculations using SACMEQ III data

In the 2011 Census, people were asked to indicate the highest level of education that they had completed. It referred to the highest level completed, not the level currently in, if the person was still studying. Figure 5 shows the education levels of adults aged 20 and older by district. The category 'Matric' refers to the secondary school leaving examination. This figure shows that Dr Kenneth Kaunda

district had slightly higher proportions of people with matric and post matric qualifications compared to those in Ngaka Modiri Molema district. Overall, this figure implies that the majority of people who would be parents to grade 1 and 2 pupils would have relatively low levels of education.

Figure 5: Highest Education level for adults aged 20 and older by district



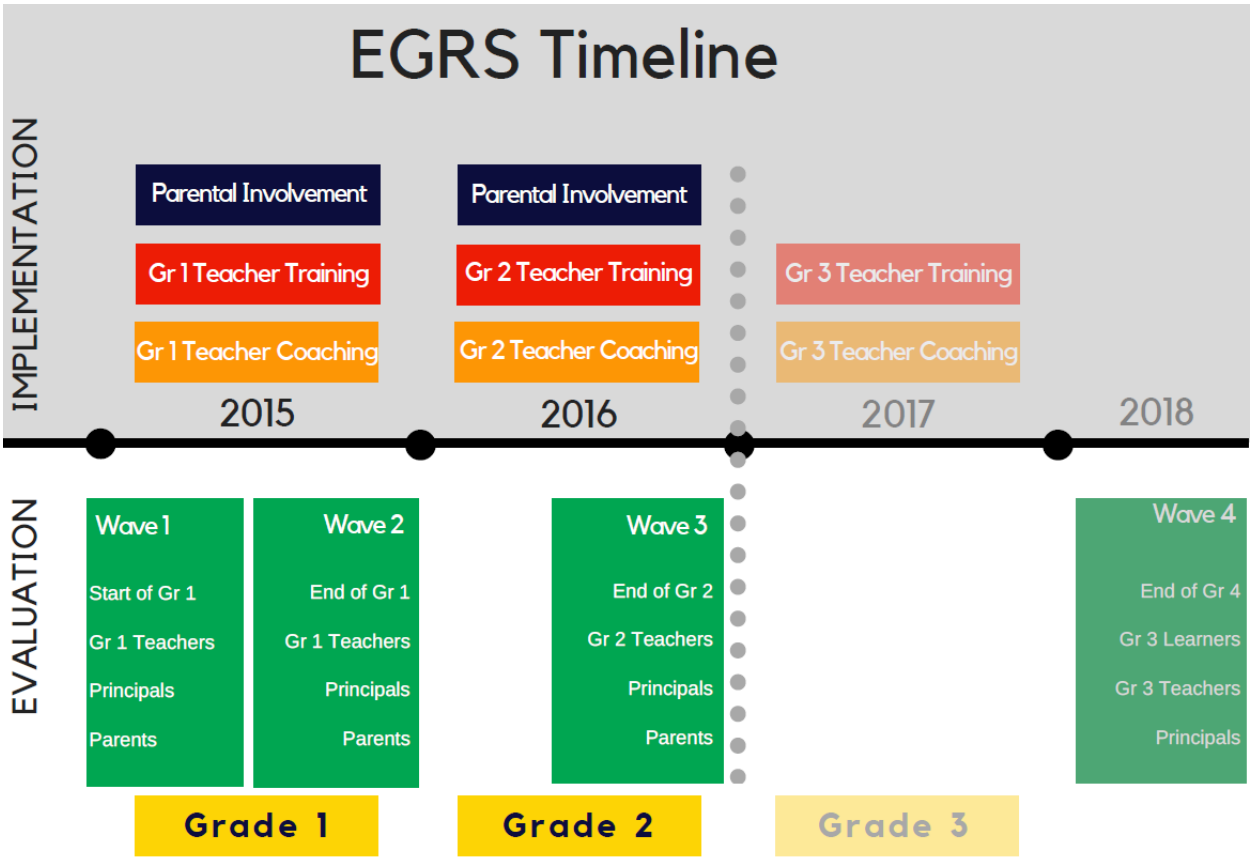
	No schooling	Incomplete Primary	Incomplete Secondary	Matric	Post Matric
Dr K. Kaunda	11%	15%	37%	26%	10%
Ngaka M. Molema	15%	19%	34%	22%	9%

Source: Own calculations using Census 2011 data

4. Timeline

Figure 6 provides a graphical summary of the timeline for the study. A period of planning, fund raising and piloting of interventions took place during 2013 and 2014. Figure 6 splits the timeline into the two main sides of the project, namely the implementation of interventions and the evaluation activities.

Figure 6: Timeline of EGRS



All three interventions began at the start of 2015, though with a slight delay in order to allow the baseline data collection to occur prior to the start of interventions. In 2015, the interventions were targeted to grade 1 learners (and their teachers and parents). As described elsewhere in this report, the parent intervention consisted of weekly meetings with parents, the training intervention consisted of 2-day centralized training sessions at the start of the year and mid-year, and the coaching intervention involved on-site visits to each teachers roughly once a month as well as occasional afternoon cluster workshops. In 2016, all three interventions continued according to the same design, only now they were targeted to the grade 2 learners. Therefore, the same cohort of learners received a second year of treatment (except for those who repeated grade 1) but a new cohort of teachers participated in the programmes. The original plan was for the project only to last two years. However, based on feedback from various stakeholders it was decided to extend the two structured learning interventions for a third year, targeting grade 3 in 2017. The midline evaluation had indicated that these two interventions were having a positive effect, whilst the evidence was suggesting no significant impact of the parent intervention. Moreover, the school curriculum is structured so that the first three grades constitute the Foundation Phase. Therefore, it made sense to extend the learning programme for the full three years of the Foundation Phase. At the time of writing, there are plans to extend the structured learning programme to all schools in the districts in which EGRS was conducted during 2019 and 2020. On-site coaching will be provided in about half of these schools.

The evaluation side of the programme commenced with a baseline data collection at the start of 2015 as children entered grade 1. A midline evaluation was then conducted at the end of 2015 (end of grade 1). An endline data collection was then administered at the end of 2016 (end of grade 2). The impact evaluation reported on in this report is mainly based on this data collection. The initial plans and funding for the project had only covered these three waves of data collection. However, at the time of writing a fourth wave of data collection is planned for August/September 2018. This will evaluate whether the third year of intervention had a further impact and whether the initial impact of the interventions persisted beyond the duration of the project into grade 4. In October 2016 two additional evaluation activities were conducted. Firstly, a set of 4 detailed school case studies was conducted in order to assess the ways in which teachers experienced and implemented the two structured learning programmes. Secondly, a classroom and lesson observation study was conducted in a sub-set of 60 schools – 20 control, 20 “Training” and 20 “Coaching” schools. This was aimed at understanding how classroom practice had been changed through the teacher-based interventions.

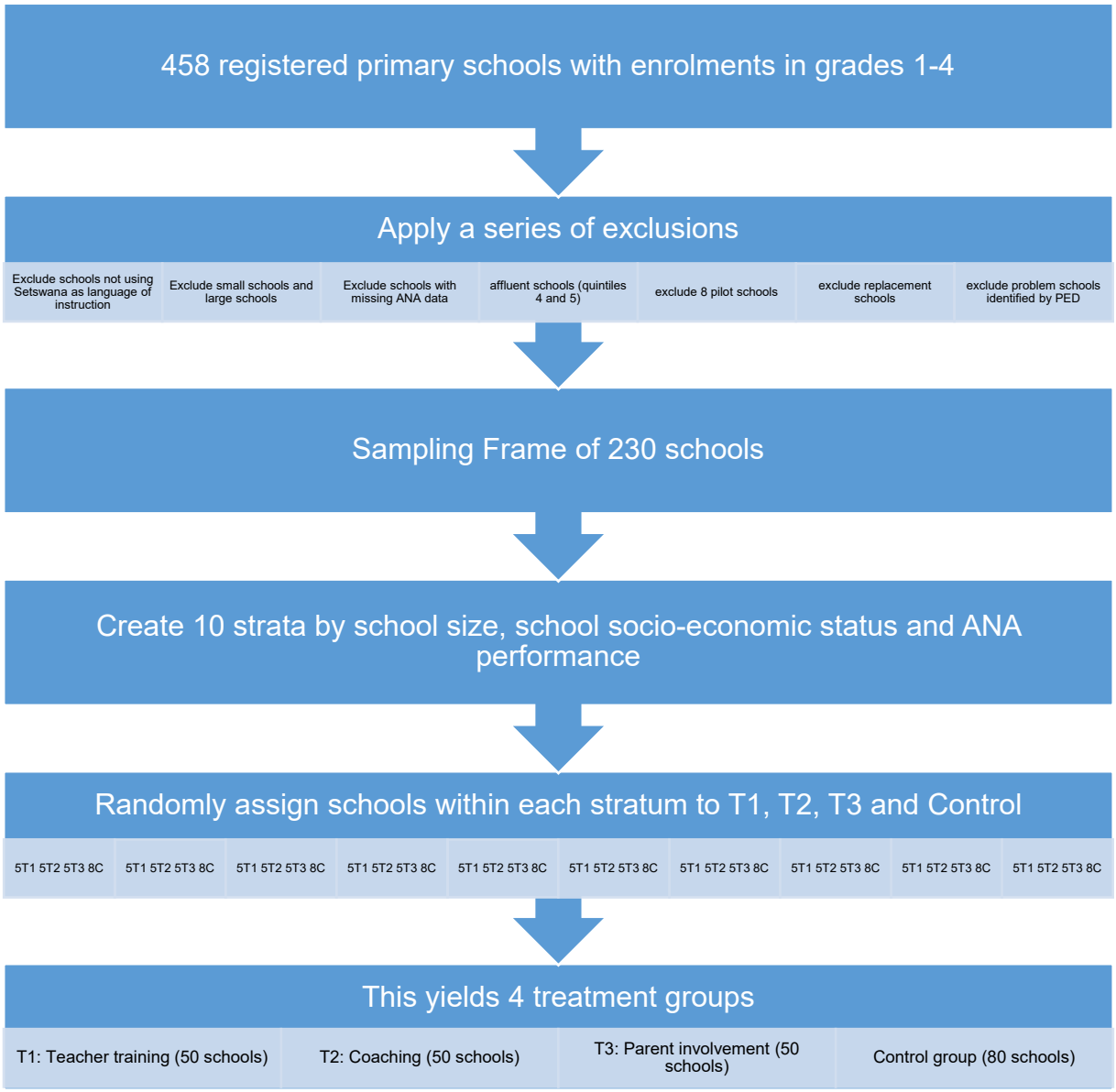
5. Evaluation: Design, methods and implementation

Sample selection and assignment to intervention group

Through a process of elimination we developed a sampling frame of 230 eligible schools. Beginning with 458 primary schools in the districts of Dr Kenneth Kaunda and Ngaka Modiri Molema registered according to 2014 administrative data, we started by excluding relatively affluent schools (those in quintiles 4 and 5). Next, we excluded schools in which the language of instruction in the Foundation Phase was not Setswana. We then excluded 1 school which was missing in the 2014 ANA dataset. We also excluded 8 schools that had already been selected for the purposes of piloting of instruments through the course of this project. We further excluded particularly small schools (fewer than 20 grade 1 enrolments) since many of these schools would practice multi-grade teaching rendering the grade-specific lesson plans less appropriate. We also excluded particularly large schools (more than 180 grade 1 enrolments) to limit intervention costs. Three more schools were excluded after the North West PED checked our list of schools and found specific problems with these schools (e.g. the school had been closed down, or a particular conflict around school management was occurring). After all of these exclusions 235 eligible schools remained. Using a random number generator, we then excluded 5 schools, which we retained as possible replacement schools. Thus we obtained the sampling frame of 230 schools.

To increase power and balance between Intervention arms, we performed stratified randomization. We created 10 strata of 23 similar schools based on school size, socio-economic status, and performance in the Annual National Assessments. Within each stratum, we then randomly assigned 5 schools to each Intervention group and 8 to the control group. Thus we randomly assigned 50 schools to each Intervention and 80 to the control. Given that we collect data on 20 grade 1 learners per school, this sample should be sufficient to identify a minimum detectable effect size of 0.21 standard deviations when comparing an Intervention group with the control group and a minimum detectable effect size of 0.23 standard deviations when comparing two Intervention groups. These calculations assume a 95% confidence interval, an alpha value of 0.8, an intra-class correlation coefficient (rho) of 0.3 and a correlation between pre- and post-test scores of 0.7. Figure 7 presents a schematic diagram to describe the sampling procedure that was followed.

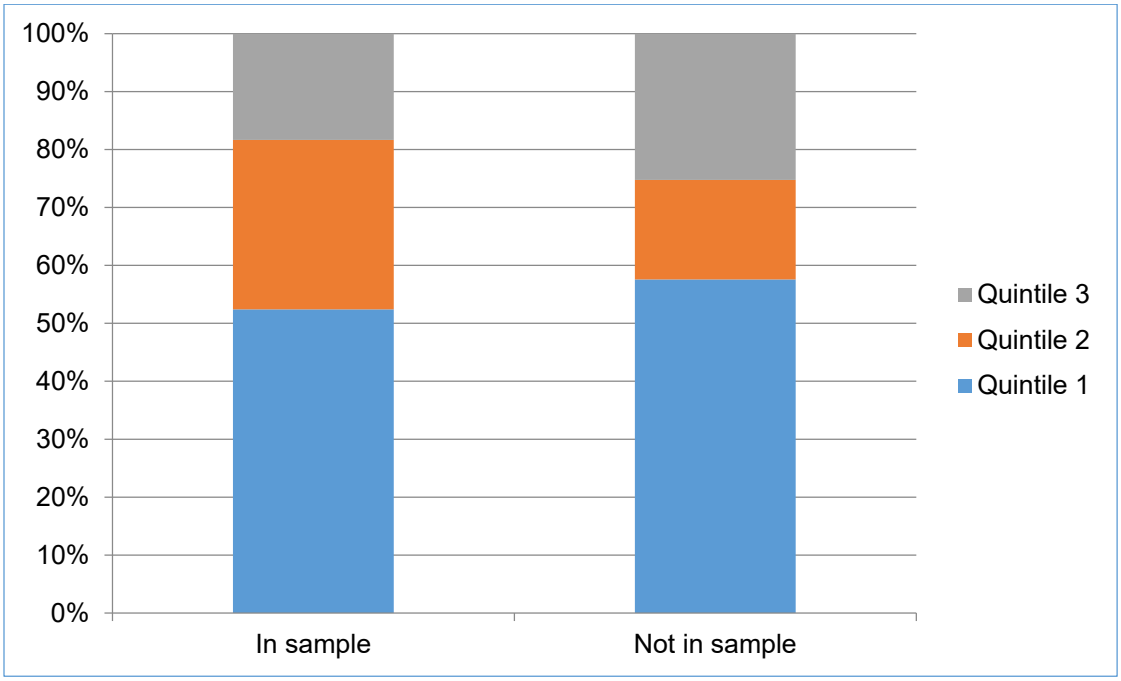
Figure 7: Diagram showing sampling procedure



The next three figures compare certain characteristics of schools finally included in the sample and those schools excluded from the study sample on the basis of the above procedures. Quintile 4 and 5 schools (fee-charging schools) and schools not using Setswana as the language of instruction are not included in the analysis, since the sample is not intended to be representative of such schools, since this is not the policy-relevant target group for the programmes. Strictly speaking, the exclusion of especially small schools is also unproblematic because these are likely to use multi-grade teaching which would render grade-specific lesson plans inappropriate. However, there is a question of the extent to which the study sample differs systematically from the large schools and those few schools excluded due to specific problems or due to being in the pilot or due to being replacement schools.

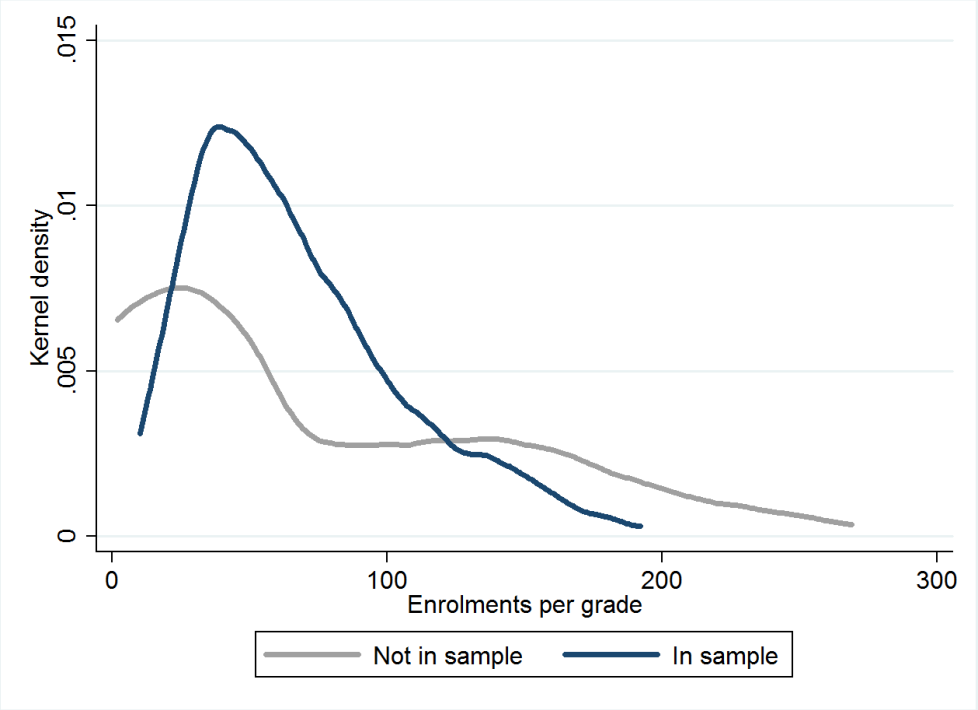
Figure 8 shows that the study sample includes slightly fewer quintile 1 schools as well as slightly fewer quintile 3 schools. This slight difference emerges due to the exclusion of small schools (which are more commonly quintile 1, deep rural and multi-grade schools) as well as the exclusion of large schools (which are more commonly urban township quintile 3 schools). Figure 9 shows the distribution of enrolments per grade and indicates that the study sample contains fewer very small schools as well as fewer very large schools, which was by design. Figure 10 shows that the distributions of learner achievement in the 2014 ANA were not significantly different between the schools in the study sample and the excluded schools.

Figure 8: School poverty quintile classification for schools included and excluded from the study sample



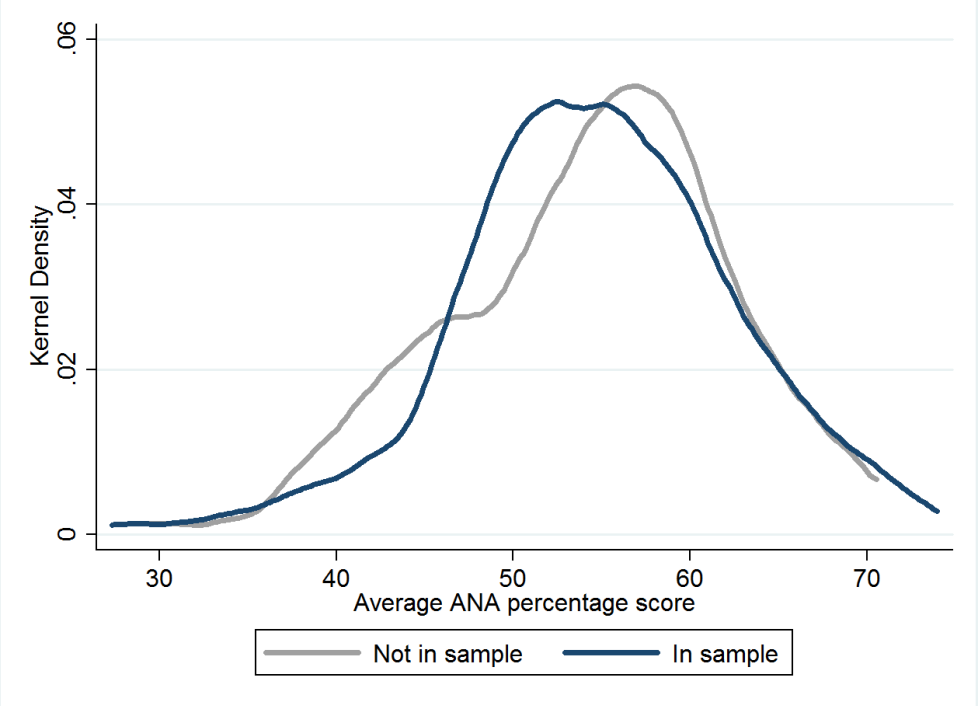
Source: DBE SNAP Survey data

Figure 9: Grade enrolments for schools included and excluded from the study sample



Source: DBE SNAP Survey data

Figure 10: School average ANA scores in 2014 for schools included and excluded from the study sample



Source: Own calculations using DBE Annual National Assessments data for 2014

Data

We visited each school three times: once prior to the start of the interventions (February 2015), again after the first year of implementation (November 2015), and finally at the end of the second year (November 2016). During these school visits we administered four different survey instruments: A pupil test on reading proficiency and aptitude conducted on a random sample of 20 pupils who entered grade one at the start of the study, a school principal questionnaire, a teacher questionnaire, and a parent/guardian questionnaire. We assessed the same pupils in every round of data collection, but surveyed a different set of teachers between midline and endline, because pupils generally have different teachers in different grades. Finally, we also conducted lesson observations on a stratified random sub-set of 60 teachers in September 2016. The data-collection and data-capture organizations are independent from the implementing organization and research team, and were blind to the treatment assignment.

Learner assessment

The learner assessment administered at the end of grade 2 (October/November 2016) was designed primarily to measure home language (Setswana) literacy outcomes, as this was the primary goal of each of our interventions. However, we included two grade-appropriate mathematics items and a few English reading items since these reflect the other two main learning areas within the Foundation Phase, namely mathematics and English as First Additional Language. Several teachers in the project had spoken about how the home language learning programme we were running was time-consuming and difficult to fit into the school timetable. Therefore, we suspected that it was possible that the EGRS home language literacy programme could have a negative effect on either mathematics or English through crowding out instructional time in those learning areas. Alternatively, we hypothesized that an effective home language programme could have positive spillover effects on mathematics or English through the acquisition of transferable skills or through the teacher's practice in one subject area improving due to training applied to another area. Therefore, we decided to include the mathematics and English items in the learner assessment.

The learner assessment administered at the end of grade 2 (October/November 2016) was designed primarily to measure home language (Setswana) literacy outcomes, as this was the primary goal of each of our interventions. However, we included two grade-appropriate mathematics items and a few English reading items since these reflect the other two main learning areas within the Foundation Phase, namely mathematics and English as First Additional Language. Several teachers in the project had spoken about how the home language learning programme we were running was time-consuming and difficult to fit into the school timetable. Therefore, we suspected that it was possible that the EGRS home language literacy programme could have a negative effect on either mathematics or English through crowding out instructional time in those learning areas. Alternatively, we hypothesized that an effective home language programme could have positive spillover effects on mathematics or English through the acquisition of transferable skills or through the teacher's practice in one subject area improving due to training applied to another area. Therefore, we decided to include the mathematics and English items in the learner assessment.

The summary statistics for all the sub-tests are reported in Table 1. The first item, common to all three waves of data collection, was letter recognition. This item was only slightly adapted from the EGRA instrument developed in Setswana. Figure 11 shows learner performance on letter recognition at baseline, Wave 2 and Wave 3. At baseline, 42% of children had scored zero, as one might expect at the start of school. By the end of grade 1 there were about 13% of children who scored zero. Although there is clear improvement over time, it is concerning that there were still over 8% of learners who could not read a single letter after two years of schooling. If we exclude those repeating grade 1, then we can say that about 6% of the sample could not read a single letter at the end of grade 2. It is, however, encouraging that about 36% of learners could read at least 50 letters correctly in a minute by the end of grade 2.

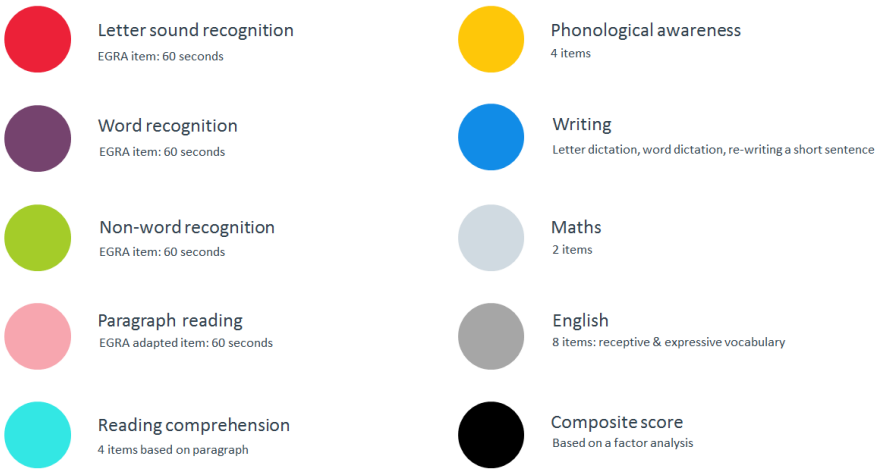
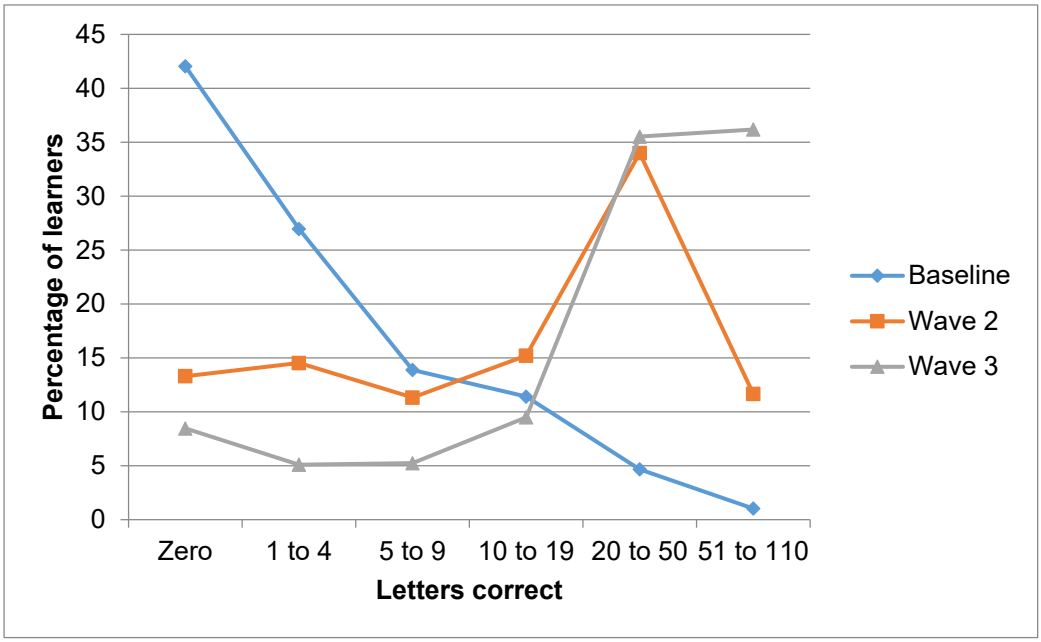


Table 1: Summary statistics for each sub-test in Wave 3 learner assessment

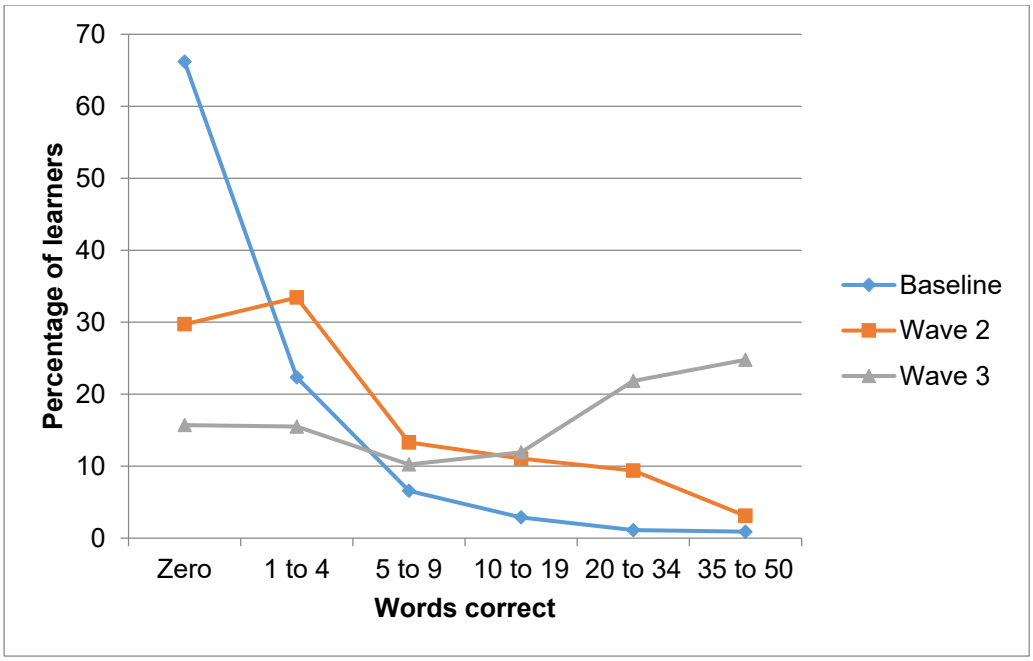
	count	mean	min	p10	p25	p50	p75	p90	max
Letter recognition	3781	39.5	0	2	16	41	60	74	110
Word recognition	3781	19.4	0	0	3	17	34	45	50
Non-word recognition	3781	14.4	0	0	0	13	26	34	50
Oral Reading Fluency	3781	25.6	0	0	0	23	50	64	66
Reading comprehension	3781	1.27	0	0	0	1	2	3	4
Phonological awareness	3781	1.82	0	0	1	2	3	3	3
Writing	3781	5.97	0	3	4	6	8	9	9
Mathematics	3781	0.6	0	0	0	1	1	1	2
English	3781	3.14	0	0	0	2	6	8	8
Composite score (SD)	3781	0.05	-1.53	-1.17	-0.92	0.04	0.93	1.40	2.18

Figure 11: Letter recognition at Baseline, Wave 2 and Wave 3



The second sub-test, word recognition, was also a standard EGRA item and was common across all three waves of data collection. As one would expect, there was a big floor effect at baseline as well as at the end of grade 1. By the end of grade 2, about 16% of children could not read a single word, but there was also a substantial proportion (25%) who could read at least 335 words correctly in a minute.

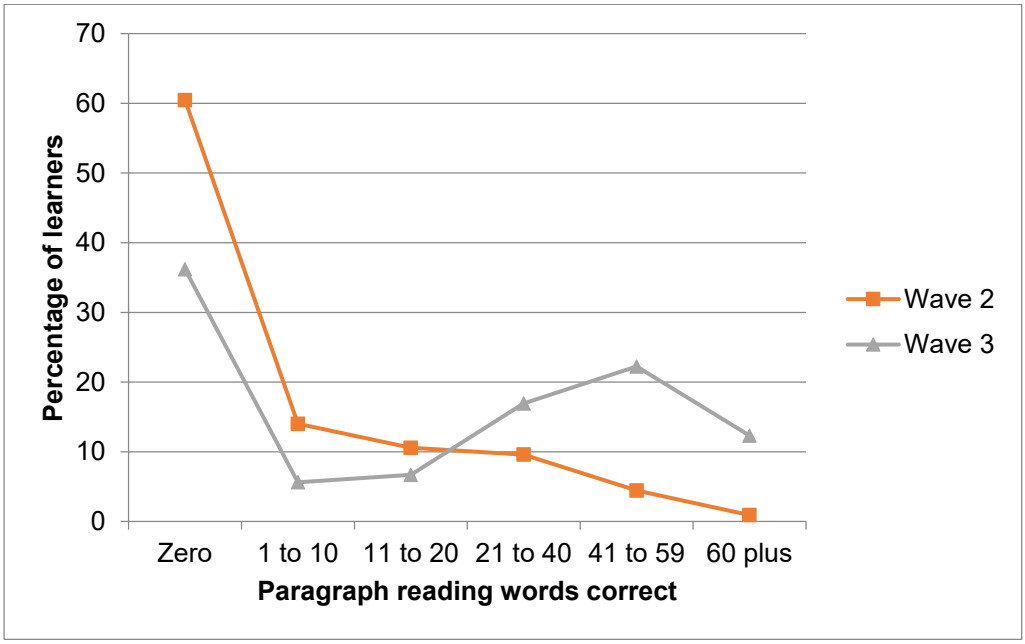
Figure 12. Word recognition at Baseline, Wave 2 and Wave 3



The third sub-test was non-word recognition. These are nonsense words which comprise letter combinations which are common to the Setswana language, but are not actual words. This tests decoding rather than whole word recognition. The scores tended to be somewhat lower than for word recognition.

The fourth sub-test was paragraph reading, which is really a measure of Oral Reading Fluency. Interestingly, for this item there were many learners who could not read a single word (36%) as well as many learners who scored highly, but few learners who read between 1 and 20 words correctly in a minute, as can be seen in Figure 13. This may suggest that once learners master the skill of decoding, reading fluency quickly follows. After reading the paragraph, learners were asked four comprehension questions about the passage. As one would expect, there was a strong correlation between oral reading fluency and comprehension scores.

Figure 13: Paragraph reading (ORF) in Wave 2 and Wave 3



Three items were included to tests phonological awareness. Two of these items asked learners to identify where in a spoken word a particular sound was located – at the beginning, in the middle or at the end. The third items asked learners to join together two sounds spoken by the fieldworker to form a single word. Although phonological awareness is a critical ingredient in the development of reading, one challenge associated with testing this skill is that fieldworkers (and hence learners) sometimes struggle to understand what is being asked of them, or to understand the difference between syllables and distinct sounds.

Several writing items were included in the Wave 3 test. One advantage of this sub-test is that there was no floor effect, with the score at the 10th percentile being 3 out of a possible 9 marks. First, learners had to write a single letter that was dictated to them, followed by a single word dictated to them. Then, learners had to fill in a missing word in a short sentence corresponding to a picture. Finally, learners

were asked to rewrite a 3-word sentence provided to them but with the correct punctuation. In this last item marks were awarded for each word, for placing spaces between the words, for using a capital letter at the start of the sentence and for placing a question mark at the end of the sentence.

As discussed earlier, two mathematics items (taken from a previous Annual National Assessment paper for grade 2) were included in order to assess whether the interventions had any positive or negative spillover effects. The first item was “8 + 3 = ___”, and the second item was “halofo ya 28” which means “half of 28”. The average score was 0.6 out of 2.

Similarly, a few English items were included. The first four items were single English words for the learner to read. Thereafter, the learner was asked to read a short English sentence consisting of four words. The average score was 3.14 out of 8.

We also constructed a composite score from all the home language literacy items using Principal Components Analysis (PCA). This is a type of factor analysis, which identifies the common underlying variation amongst a set of variables and regards the first principal component in this variation as being reflective of the underlying construct, which in our case would be Setswana reading literacy. For ease of interpretation we then standardized the composite index to have a mean value of 0 and a standard deviation of 1 (within the control group). In much of the literature reporting on Early Grade Reading Assessments it is not common practice to derive a composite score, but rather to interpret changes in each of the sub-tests separately. To some extent this is motivated by a theoretically driven approach to analyzing the development of reading and to some extent by the idea that the various sub-tests do not easily fit into a single underlying construct. However, we were also concerned about the risks of cherry-picking results if we present impact analysis on each of seven outcomes and then for all of these outcomes also go on to present heterogeneous treatment effects. Therefore, we have opted to derive the composite score and use this in all our heterogeneous treatment effect estimations and robustness checks. However, for the main impact evaluation model we also present the results on each of the sub-tasks separately.

Overall, the learner test information from Waves 2 and 3 appears to have been of a better quality than that obtained through the baseline. This is evident in the Cronbach’s alpha values which were 0.83 for both the Wave 2 and Wave 3 learner tests, but only 0.65 for the baseline assessment. Cronbach’s alpha provides a measure of how well the various sub-tests fit together as measures of a single underlying construct, where a value closer to 1 is better. Table 2 below also provides evidence of how the baseline test provided a weak signal of learner proficiency. While the Wave 2 and Wave 3 total scores are highly correlated (0.72) each of them is only weakly correlated with the baseline total score.

Table 2: Correlation coefficients between Wave 1, 2 and 3 scores

	Baseline Total Score	Wave 2 Total Score	Wave 3 Total Score
Baseline Total Score	1		
Wave 2 Total Score	0.25	1	
Wave 3 Total Score	0.22	0.72	1

Empirical Strategy

For our main regressions we employ the following empirical strategy:

$$y_{icsb1} = \beta_0 + \beta_1(Coach)_s + \beta_2(Train)_s + \beta_3(Parents)_s + X'_{is0}\Gamma + \rho_b + \varepsilon_{icsb1}$$

Where y_{icsb1} is endline aggregate learning score for pupil i in classroom c, school s, and strata b; ρ_b refers to strata fixed effects; X'_{is0} is a series of individual and school-level controls collected at baseline; and the error term, ε_{icsb1}, is clustered at the school level.

In order to increase statistical power, we control separately for each domain of reading proficiency collected at baseline: vocabulary, letter recognition, working memory, phonological awareness, word recognition, words read, and sentence comprehension. To further increase statistical power and account for any incidental differences that may exist between treatment groups, we control for individual and community-level characteristics which are highly correlated with y_{icsb1} m or were imbalanced at baseline. The additional controls for our preferred specification include: pupil gender, pupils' parents' education, district dummy (schools were randomly spread across two districts), performance in the most recent standardized Annual National Assessments (ANA), a community-level wealth index, and average secondary school attendance rate in the community surrounding the school. Where data is missing for some observations for the control variables, we imputed missing values and added a dummy indicating missingness as a control.[5]

When we examine heterogeneous treatment effects we estimate the following equation:

$$y_{icsb1} = \beta_0 + \beta_1(Coach)_s + \beta_2(Train)_s + \beta_3(Parents)_s + \beta_4(Coach \times g)_s + \beta_5(Train \times g)_s + \beta_6(Parents \times g)_s + \beta_7 g + X'_{is0}\Gamma + \rho_b + \varepsilon_{icsb1}$$

Where g refers to the sub-group of interest, which could be at either the individual, class/teacher, or school level.

When we look at intermediate outcomes at a teacher/classroom level, the estimating equation becomes:

$$y_{csb1} = \beta_0 + \beta_1(Coach)_s + \beta_2(Train)_s + \beta_3(Parents)_s + X'_{s0}\Gamma + \rho_b + \varepsilon_{csb1}$$

[5] For categorical variables, we assigned missing values to zero; for continuous variables we assigned missing observations to equal the sample mean

6. Programme design and implementation

Intervention one (Training)

Although 50 schools were randomly selected for this programme, one school fell out of the programme because it was merged with another school and thus no longer existed, leaving 49 schools actively participating. All 50 schools were, however, tested as part of the evaluation side of the project. In 2015, grade 1 teachers received the intervention, while in 2016 grade 2 teachers in the same schools were targeted. Teachers were provided with teaching and learning materials and were trained on their use twice each year for two days at a time in a conventional model common to large-scale training. The teachers received the official support of the North West Department of Education, and Class Act was in contact with them during the year through social media platforms.

The table below details the teaching and learning materials that teachers were provided with during the project.

Table 3: Intervention 1 materials

ITEM	DESCRIPTION
Vula Bula Reading Books	Commercially produced grade One and grade Two Setswana graded reading books. These were used in group guided reading lessons
Book register	An exercise book set up as an accession register for the Vula Bula reading books
Teacher file	A management file to keep teaching and learning materials
Setswana HL scripted lesson plans	This document contains the individual lesson plans that teachers followed in 2015 and in 2016
Flashcard words	Printed sets of the words teachers needed to teach sight words in reading lessons
Reading words	A learner resource that listed the sight words taken from the Vula Bula books. These word lists were taken home so that the learners could practice reading
Assessment records	CAPS and SA-SAMS compliant assessment record tables. Teachers used this resource to record formal assessments per learner
Assessment rubrics	Criteria for teachers to use to award objective assessment ratings for learner tasks
Curriculum tracker	A tool for teachers to manage curriculum coverage
Weekly routine	A tool for teachers to manage curriculum pacing
Core methodologies	Detailed pedagogical support that helped teachers learn how to use tried and tested methodologies for different language components
Handwriting poster	A poster that demonstrated the form and directionality of lower and upper case letters
Theme posters	Posters that detailed interesting scenes that were used for vocabulary development
Facilitators' Guides	Detailed handbooks for trainers to follow when they trained teachers.

Source: Class Act implementation monitoring data

The 2-day training events occurred four times over the course of 2015 and 2016 and were well attended, as can be seen in Table 4. The first training event in February 2015 covered the lesson plans for Term 2 only (and not Term 1) since the learning programme only began in Term 2 due to the

other preparatory activities taking place in Term 1. This would not have meant any disruption in the learning programme in intervention schools since the lesson plans are aligned to the official curriculum, only specifying a greater level of detail and with particular activities, instructional methods and resources integrated into the lessons. Only on one occasion was a school not represented at all at a training session. The attendance rate for teachers was also high (between 85% and 100%) and was sustained throughout the two years. Teachers who did not attend the residential training sessions were provided with catch-up training. Attendance rates were a little lower for school leaders (principals or HODs) but this is not a major concern since they were not the primary recipients of training.

Table 4: Attendance rates at training events

	GRADE ONE (2015)		GRADE TWO (2016)	
	ACTUAL TERM 2 FEB 2015	ACTUAL TERM 3 & 4 JULY 2015	ACTUAL TERM 1 & 2 JAN 2016	ACTUAL TERM 3 & 4 JULY 2016
% Schools attended	100	98	100	100
% Teachers attended	100	85	98	93
% School leaders attended	74	78	90	80

Source: Class Act implementation monitoring data

Intervention Two (Coaching)

Intervention Two also targeted teachers (grade 1 in 2015; grade 2 in 2016), but in a different randomly selected group of schools across the same districts as Intervention One. One school fell out of the programme because it was a multi-grade school and the principal therefore requested not to be part of the project since we were using grade-specific lesson plans. This left 49 schools actively having participated. All 50 randomly selected schools were still included in the data collection for evaluation purposes. [6] These teachers were provided with the same teaching and learning materials as Intervention One (see Table 3 in the previous section). But they received more intensive cluster-based training four times a year and had the support of instructional coaches in their schools and in their classrooms. The teachers also received the official support of the North West Department of Education and were in ongoing contact with their coaches between support visits throughout the year

Table 5 summarizes the attendance of teachers at the various training engagements as well as the dosage of on-site coaching visits. High attendance levels were noted throughout the project demonstrating ongoing commitment. Teachers were supported throughout the project in their classrooms between 2 and 3 times per term. Fewer coaching visits were possible per teacher in the last term of 2016 due to a combination of social unrest in one district and learner assessments and other outside disruptions in both districts. In addition to classroom-based support, teachers received additional support during needs-driven afternoon workshops amongst nearby clusters of schools, which were facilitated by coaches. Although these workshops did happen to some extent in Year 1, these support initiatives became more structured in Year 2 and were reported on in Year 2. Due to

[6] Excluding them from the sample, or replacing them with other schools could introduce bias.

the shortened length of Term Four and due to disruptions to schooling in the area no afternoon workshops were run by coaches.

Table 5: Summary of attendance and dosage of Intervention Two

	GRADE ONE (2015)			GRADE TWO (2016)			
	TERM 2 FEB 2015	TERM 3 JUL 2015	TERM 4 SEP 2015	TERM 1 JAN 2016	TERM 2 APR 2016	TERM 3 JUL 2016	TERM 4 SEP 2016
% Schools attended 1-day training	100	92	100	100	100	96	100
% Teachers attended 1-day training	100	89	100	99	100	92	99
Average number of on-site coaching visits	3	2	2	2	3	3	1
% Teachers attended cluster-based afternoon workshops				48	59	61	0

Source: Class Act implementation monitoring data

Intervention Three

Intervention Three was conducted in a third randomly selected group of 50 schools. One school fell out of the programme as it was a boarding school and the parents stayed too far away to attend weekly meetings. Therefore, 49 schools actively participated in the programme, although some of these schools experienced disruptions to the programme due to delays in finding a Community Reading Coach (CRC) to run the weekly parent meetings. This intervention focused on parents/guardians (grade 1 in 2015 and grade 2 in 2016) through weekly meetings facilitated by CRCs. The weekly sessions were aimed at helping parents understand how their children were learning to read and to provide parents with strategies to use at home to stimulate a culture of reading. CRCs were recruited with the help of school principals and received regular training sessions from Class Act (6 sessions in 2015 and 7 in 2016). The CRCs were in contact with Class Act during the project through social media platforms. The table below details the materials that parents were provided with during the project. For each module, three meetings were held, each covering the same topic but using a different set of activities. This configuration meant that parents could attend once every three weeks and still have "full coverage", but that if parents attended every week they would not repeat the exact same session.

Table 6: Intervention 3 materials

ITEM	DESCRIPTION
Module One	'Small things can make a difference' plus a set of family reading cards
Module Two	'Playing with sounds to support reading' plus a set of family reading cards
Module Three	'Reading pictures' plus a set of family reading cards
Module Four	'Letter sounds' plus a set of family reading cards
Module Five	'Incidental reading' plus a set of family reading cards
Module Six	'Preparing to read a story – Part One' plus a set of family reading cards
Module Seven	'Reading a story' plus a set of family reading cards
Module Eight	'Preparing to read a story – Part Two' plus a set of family reading cards
Module Nine	'Reading Remediation'
Module Ten	'Reading stories 3 and 4'
Module Eleven	'Reading story 5'
Module Twelve	'Reading stories 6, 7 and 8'
Module Thirteen	'Reading stories 9, 10 and 11'
Module Fourteen	'Reading stories 12, 13 and 14'
Module Fifteen	Consolidation
Grade One reader	Platinum Series Le Re Tlhabetse Readers published by Maskew Miller Longman / Pearson: Book 1
Grade Two reader	Platinum Series Le Re Tlhabetse Readers published by Maskew Miller Longman / Pearson: Book 2
Facilitators' Guides	Detailed handbooks for trainers to follow when they trained CRCs. These were also used for parent training.

Source: Class Act implementation monitoring data

CRC attendance at training sessions was mostly fairly high, as Table 7 shows, and where CRCs did not attend catch-up training was provided as far as possible.

Table 7: CRC attendance at training

	APR 2015 MOD 1	MAY 2015 MOD 2&3	JULY 2015 MOD 4	SEPT 2015 MOD 5&6	OCT 2015 MOD 7	NOV 2015 MOD 8	
% ATTENDANCE	100	90	71	86	90	82	
	JAN 2016 MOD 8 REVISION	MAR 2016 MOD 9&10	APRIL 2016 MOD 11	MAY / JUNE MOD 12	JULY 2016 MOD 13	SEP 2016 MOD 14	OCT 2016 MOD 15
	98	90	90	86	73	75	78

Source: Class Act implementation monitoring data

CRCs kept attendance registers of parents who attended the weekly meetings. Parents signed against the learner's name so as to be able to track attendance relative to the learner. This was done with a view to linking the attendance data to the learner test data collected independently. Table 8 shows the percentage of learners who were represented at the meetings. The percentage who attended at least one session per module is shown. The average parental involvement throughout the duration of the project was approximately 31%. This low level of parental involvement remained a concern across both years despite attempts to address this during Year 2.

Another way to present the level of parent attendance is shown in Figure 14. This indicates that in Year 1 about 50% of parents did not even attend once, while in Year 2 not even 40% of parents

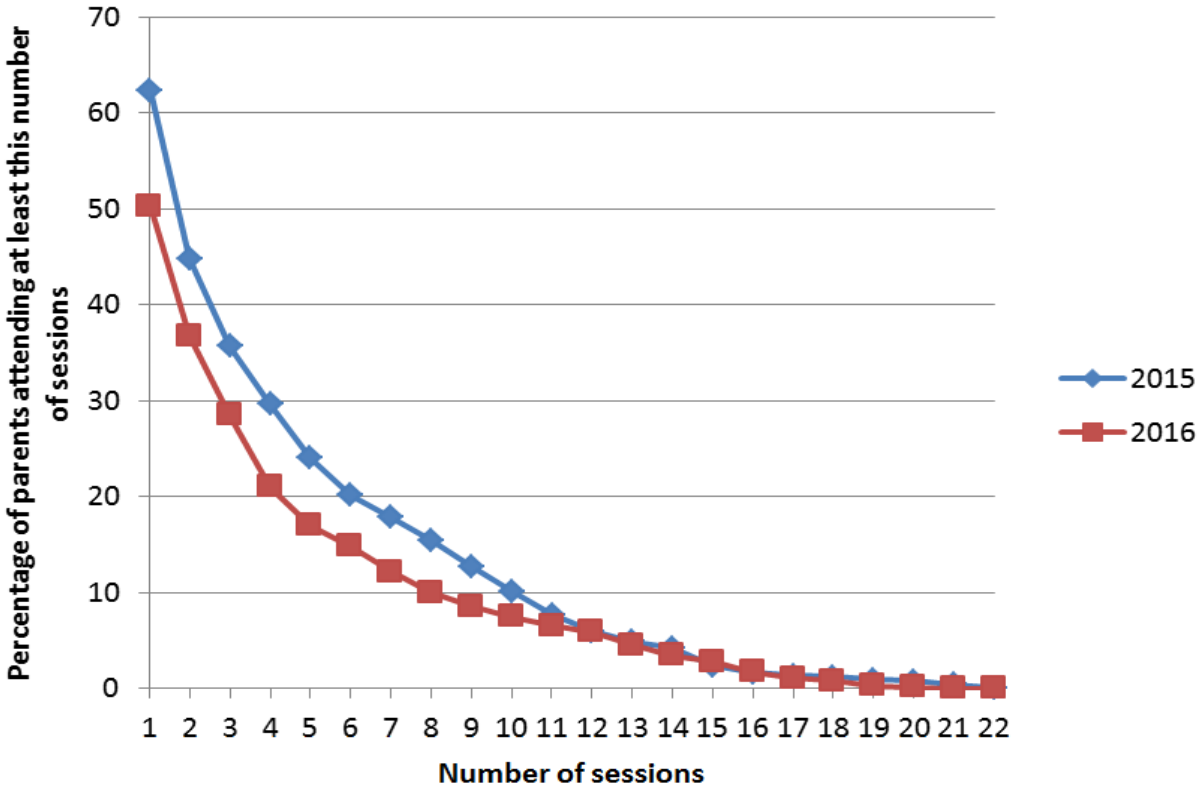
attended any sessions. Approximately one-third of parents attended at least three sessions per year. This low level of attendance obviously diminishes the likelihood of an overall impact on the average learner.

Table 8: Percentage of learners whose parent/guardian attended meetings

	MOD 1	MOD 2&3	MOD 4	MOD 5&6	MOD 7	MOD 8	
% ATTENDANCE	42	36	26	38	31	22	
	MOD 9	MOD 10	MOD 11	MOD 12	MOD 13	MOD 14	MOD 15
	36	35	34	26	25	19	37

Source: Class Act implementation monitoring data

Figure 14: Parent attendance at Intervention 3 meetings



Source: Class Act implementation monitoring data

7. Impact analysis and results of the key evaluation questions

Balance

Table 9 shows balance and basic descriptive statistics of our evaluation sample. Each row represents a separate regression of the baseline variable on treatment assignments and strata dummies, clustering standard errors at the school level. Columns (1) and (2) show the control mean and standard deviation; columns (3), (5), and (7) the coefficient on the two treatment dummies. Standard errors are indicated in columns (4) and (6) and (8).

Our sample of schools comes predominantly from poor and remote communities. 85% of schools are from rural areas. In only 44% of schools do the majority of parents have a high school degree or higher, and 52% of schools do the head teacher estimate that 20% or less of the parents are unemployed. A sizable fraction of classrooms ended up being multi-grade classrooms (6.2 of grade two classrooms). We were thus not perfectly able to identify and exclude *ex ante* all schools that do multi-grade teaching. Since the program was not intended to work in multi-grade settings, we will also report impacts on a sample that excludes these classrooms as well. Teachers are mostly female and are educated: 95% of the grade 2 teachers have a degree or diploma. Nonetheless, reading comprehension levels are low: The average score for the simple comprehension test is 66% for the grade 2 teachers.

Table 9 shows that the sample is balanced – we do not observe more statistically significant relationships than would be expected by chance. However, importantly, we observe imbalance on baseline pupil reading proficiency for the Training treatment arm. We control for this in the main regression specification.

Table 9. Balance and descriptive statistics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Control mean		*Training*		*Coaching*		*Parents*		*Obs*
	Mean	Std Dev	Coef.	Std. error	Coef.	Std. error	Coef.	Std. error	
Pupil Characteristics									
Age	6.481	0.725	0.0778	(0.0524)	-0.0249	(0.0523)	0.0596	(0.0606)	4,512
Female	0.479	0.500	-0.0156	(0.0221)	-0.0118	(0.0209)	-0.0405**	(0.0181)	4,514
Reading proficiency	0.0404	0.990	-0.209*	(0.119)	0.0670	(0.147)	-0.0389	(0.153)	4,538
Grd 2 Teacher Characteristics									
Diploma or degree	0.947	0.224	0.0138	(0.0310)	0.0412	(0.0251)	0.00110	(0.0321)	348
Age	48.92	9.087	-1.505	(1.359)	-0.256	(1.219)	-0.585	(1.419)	353
Female	1	0	-0.0145	(0.0133)	-0.000861	(0.00402)	-0.0498**	(0.0242)	351
Class size	42.17	10.42	-2.024	(1.463)	-3.184**	(1.608)	1.569	(1.880)	350
Multi-grade	0.0619	0.242	0.00549	(0.0339)	0.000907	(0.0295)	0.00554	(0.0381)	350
Comprehension test	0.663	0.216	-0.0415	(0.0305)	-0.00284	(0.0335)	-0.0418	(0.0319)	344
School characteristics									
Rural	0.850	0.359	-0.0700	(0.0681)	-0.110	(0.0681)	-0.170**	(0.0681)	230
Pass rate (ANA)	55.35	7.324	-1.184	(0.882)	-0.981	(0.882)	-1.061	(0.882)	230
Parents education - At least second.	0.443	0.500	-0.105	(0.0859)	0.0348	(0.0859)	-0.0531	(0.0864)	228
Kenneth district	0.212	0.412	-0.0125	(0.0741)	0.0875	(0.0741)	0.0275	(0.0741)	230
Female principal	0.519	0.503	-0.0186	(0.0912)	-0.0786	(0.0912)	-0.0517	(0.0918)	228
Parents employed < 20%	0.519	0.503	0.0125	(0.0900)	-0.201**	(0.0894)	-0.0890	(0.0900)	225
School - good condition	0.266	0.445	0.0523	(0.0801)	0.0523	(0.0801)	-0.0606	(0.0806)	228

Notes: Each row indicates a separate regression on treatment dummies controlling for strata indicators. Columns (1) and (2) show the control mean and standard deviation; columns (3), (5), and (7) the coefficient on the two treatment dummies. Standard errors are indicated in columns (4) and (6) and (8) and are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1

Attrition

At baseline we assessed 4 538 learners at the start of their grade 1 year in 2015. At the midline assessment (end of grade 1) we successfully re-assessed 4 143 learners, meaning that about 9% of the original sample were either absent from school on the day of the survey, or had moved to another school, or had returned to Grade R, or had stopped attending school. However, at the end of grade 2, just over half of those of absent for the midline were successfully re-assessed, implying that they must have been absent from school on the day of the midline survey. Therefore, we can say that at most 4% of the original sample appeared to have left the school before the end of grade 1.

By the time of the Wave 3 data collection (end of grade 2, 2016), we successfully assessed 3 781 learners (83.3% of the original sample). This means that, over and above the 4% of learners who left the school before the end of grade 1 in 2015, a further 13% were either absent on the day of the Wave 3 survey or had left the school since grade 1. Unfortunately, we are not able to distinguish between these two reasons.

Amongst those learners who were successfully identified at the Wave 3 survey, 591 were found to be repeating grade 1. This amounts to about 13% of the original sample, or about 16% of those successfully identified at Wave 3. The latter estimate is probably the more relevant estimate of the grade repetition rate for grade 1, though of course it is not based on a nationally or even provincially representative sample. Nevertheless, this estimate is in line with internal DBE analysis of LURITS data, and confirms that grade repetition is higher than what certain other sources, such as the General Household Survey, would suggest.

Figure 15 shows the proportions of the original sample that attrited, that were repeating grade 1, and that were found to be in grade 2. Importantly, there were no significant differences in attrition or grade repetition across the three intervention groups. Column (1) in Table 10 regresses treatment assignments on attrition status, after controlling for stratification. It shows there is no statistically significant difference in attrition rates across treatment arms. This means that the tested samples of learners will not be selectively stronger or weaker in any intervention group, something that could have introduced bias into the impact analysis.

Coach attrition was not an issue since three coaches covered the 50 schools receiving the intervention and none of them attrited. Teacher attrition (in the form of non-compliance with Interventions 1 and 2) would have weakened programme implementation. This is unlikely to have been a significant factor since coaches worked with all teachers in schools that were visited. R the training intervention, we know that teacher attendance at the training sessions was high (ranging between 85% and 100%). When it comes to teacher attrition from the evaluation dataset, we cannot test for teacher-level attrition between baseline and endline, because different teachers were surveyed in the different rounds. However, we can test whether teacher non-response at endline was balanced. We have teacher survey data for 275 teachers at endline, covering 81 percent of the pupils assessed.[7] In column (2) in

[7] We cannot tell what proportion of teachers did not respond, because children are randomly drawn at a school level, so we do not know how many teachers pupils with missing teacher data would have matched with

Table 10 we regress treatment assignment dummies on an indicator for whether a pupil's teacher also completed the teacher survey. We see that teacher non-response was random across treatment arms.

Figure 15: Attrition and grade repetition in the sample

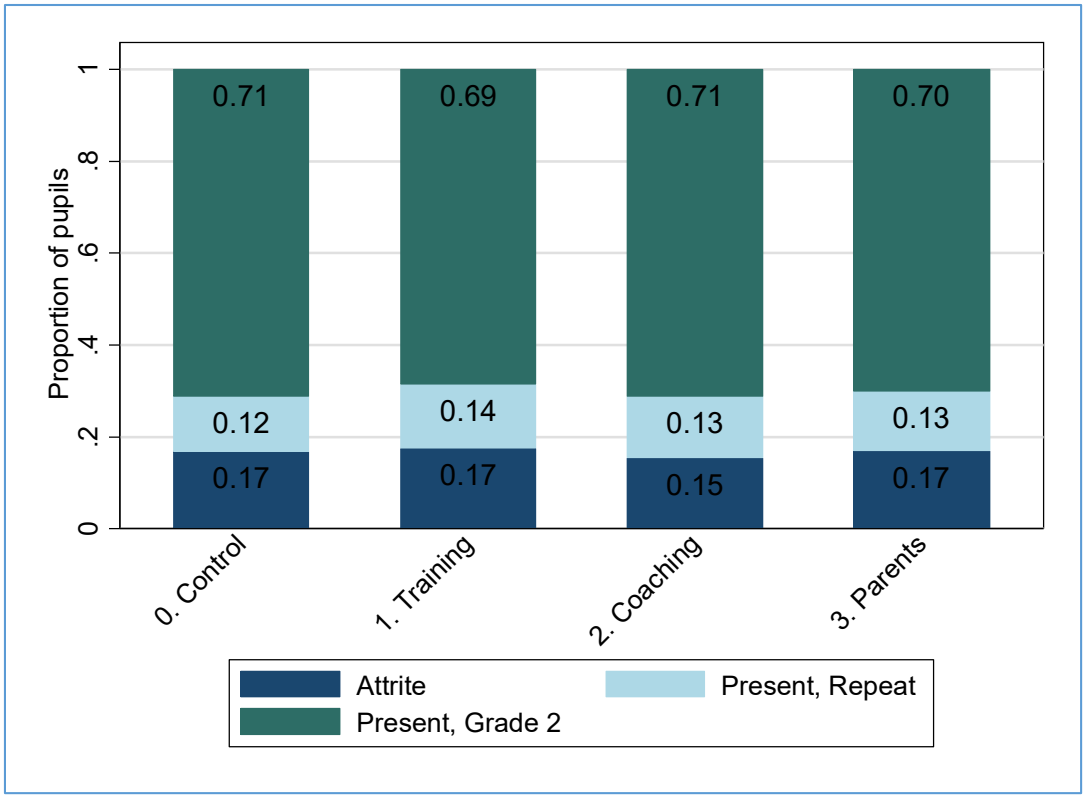


Table 10: Pupil and teacher attrition

	(1) Pupil attrition	(2) Teacher attrition
Training	0.00605	-0.0239
	(0.0222)	(0.0363)
Coaching	-0.0136	-0.0234
	(0.0183)	(0.0378)
Observations	3,539	2,951
R-squared	0.010	0.015
Control mean	0.168	0.208

Notes: Each column represents a separate regression. Column headings indicate the dependent variable. "Pupil attrition" is a dummy variable equal to one if the pupil was not surveyed at endline. "Teacher attrition" is a dummy variable equal to one if the pupil's teacher was not surveyed at endline. Observations in Treatment 3 (“Parent involvement”) are excluded since the focus in this table is on teacher attrition. Only controls included are stratification dummies. Standard errors are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1.

Unconditional comparison of means

A simple comparison of means provides an initial perspective on the possible impact of the three interventions, as reported in Table 11. Glancing over the table indicates that children in the Intervention 2 group had consistently higher average scores than the other groups. The scores for Interventions 1 and 3 do not appear noticeably higher than the control group (although one should keep in mind that Intervention 1 had lower baseline average achievement, as the balance tests revealed).

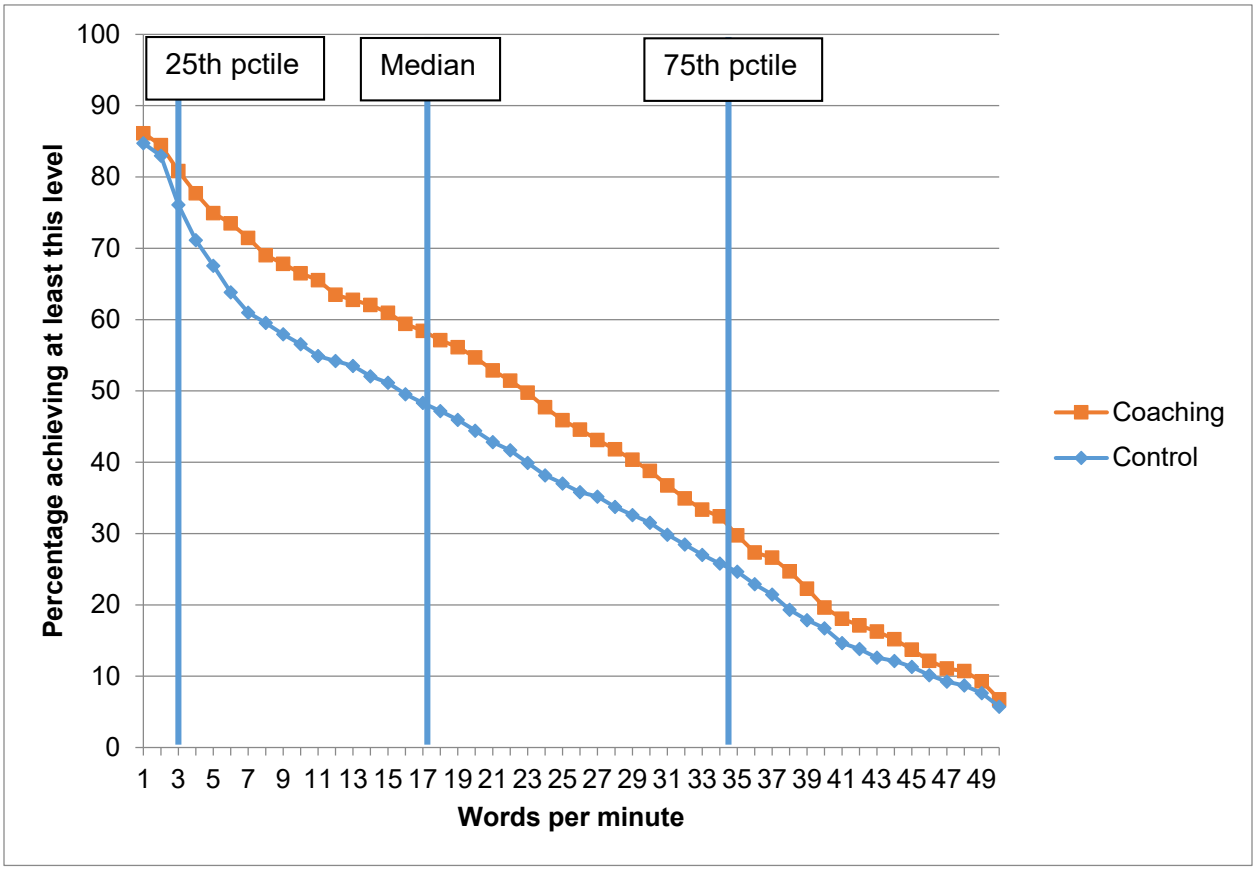
Table 11: Mean scores for all sub-tests by intervention group

	Control	Intervention 1 (Training)	Intervention 2 (Coaching)	Intervention 3 (Parents)
Letter recognition	39.04	37.52	43.01	38.70
Word recognition	18.91	18.77	22.28	18.02
Non-word recognition	13.69	13.87	16.99	13.40
Oral Reading Fluency	24.48	24.95	29.90	23.67
Reading comprehension	1.234	1.185	1.523	1.171
Phonological awareness	1.738	1.813	1.883	1.914
Writing	5.898	5.894	6.225	5.905
Mathematics	0.588	0.575	0.634	0.607
English	3.024	3.006	3.649	2.936
Composite score (SD)	0.000	-0.003	0.218	-0.0101

Note: Data from EGRS Wave 3

The next three figures present further descriptive evidence of the differences in achievement between the Intervention 2 group ("Coaching") and the control group. Figure 16 shows the percentage of children achieving above particular thresholds of words correct per minute, separately for the two groups of children. The scores at the 25th percentile, median (50th percentile) and 75th percentile of the distribution for the full sample of 230 schools are also indicated on the graph. In both groups there were roughly 85% of children who could read at least one word correctly. There were also similar percentages of children who managed to read all 50 words correctly within a minute (about 6%). However, throughout the range between zero and 100% there were consistently more Intervention 2 children able to surpass particular thresholds. Between the thresholds of about 10 words per minute and 25 words per minute there were consistently about 10% more children in the Intervention 2 group able to read at least that number of words than in the control group. The pattern in this graph points to the possibility that the impact of the coaching intervention was largest for children in the mid-range of the performance distribution.

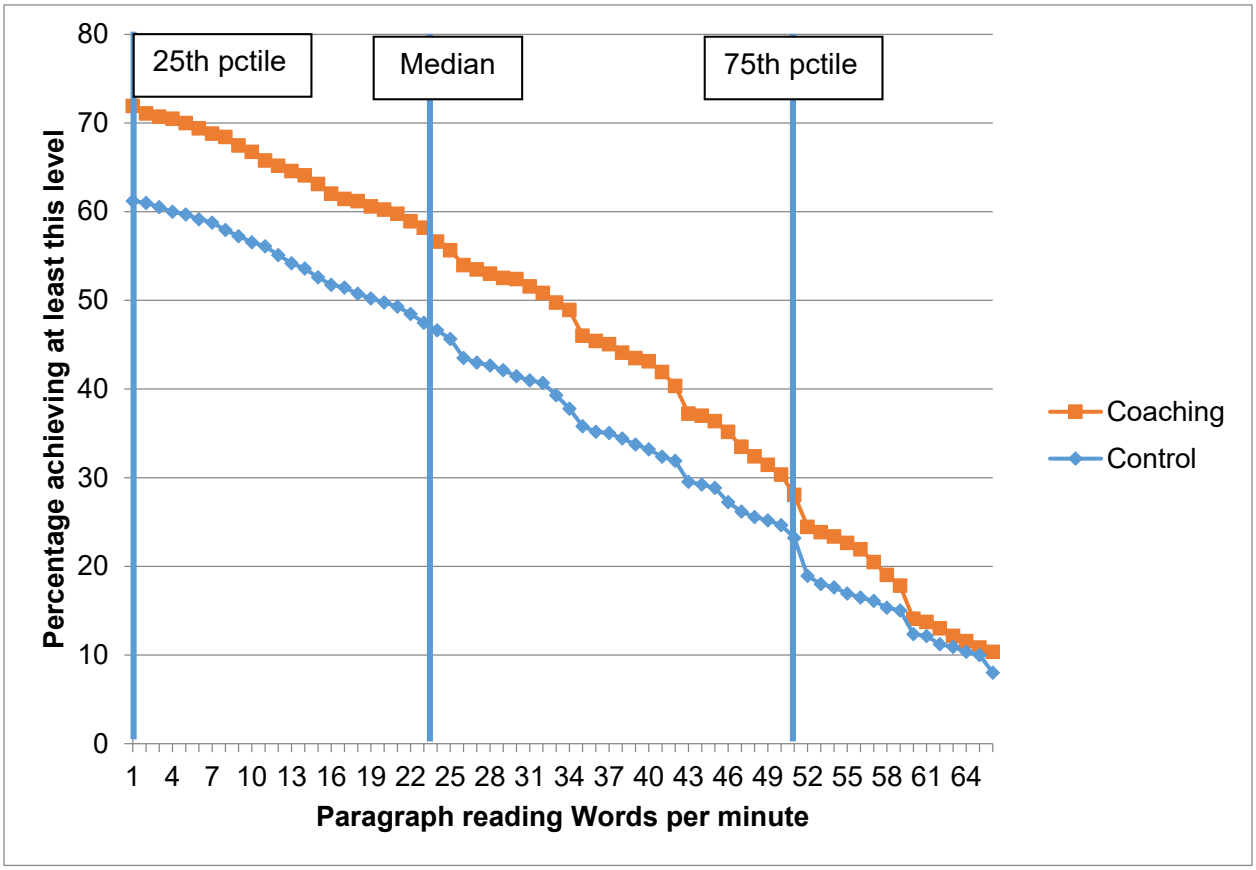
Figure 16: Word recognition for Intervention 2 and Control



Note: Data from EGRS Wave 3

Figure 17 shows the same type of graph as above, now applied to the Paragraph reading test, which provides a measure of oral reading fluency. As before, similar proportions of children could complete reading the entire paragraph within a minute (about 10%). A significant floor effect on this item (36% of children scoring zero) means that it is difficult to say anything about the impact on the bottom end of the distribution, but we can say that whereas only about 61% of children in the control group could read at least one word correctly, there were about 72% of children in the "coaching" group who could do so. Similarly, about 10% more children surpassed the median level of achievement (23 words per minute) in the coaching group compared to the control group. Only for the top 30% of the distribution did the magnitude of this impact drop off.

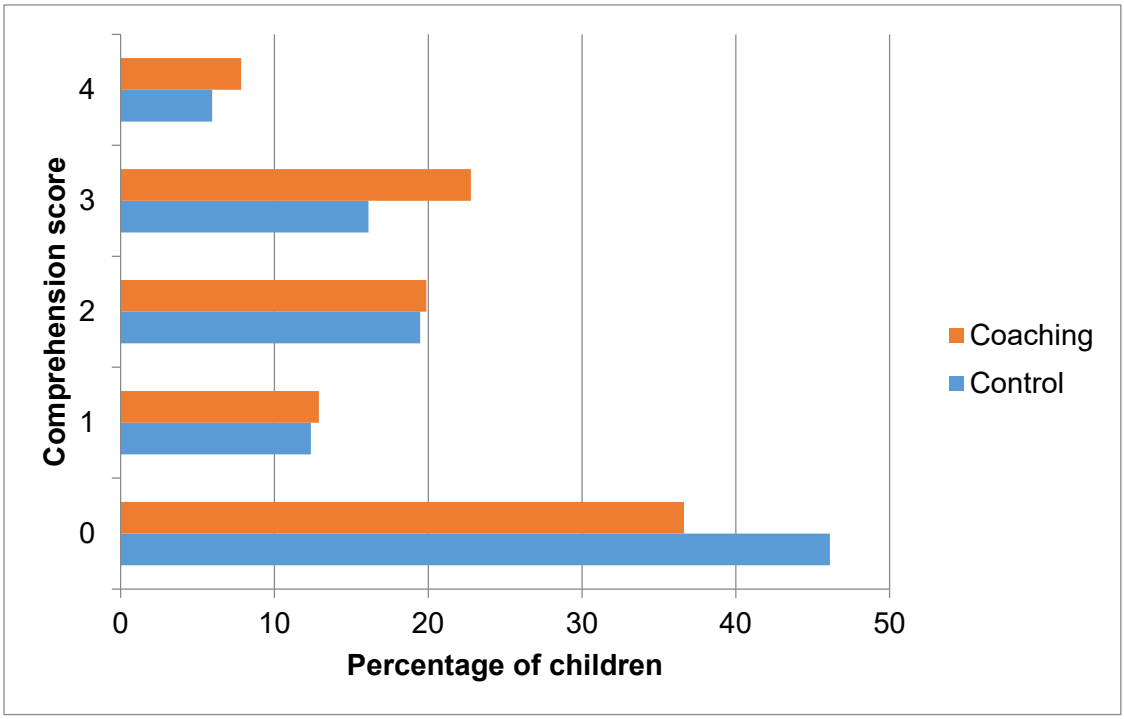
Figure 17: Oral Reading Fluency for Intervention 2 and Control



Note: Data from EGRS Wave 3

Figure 18 shows the percentages of children achieving each possible score out of 4 on the comprehension test, which was administered after the paragraph reading. Whereas 46% of children in the control group scored zero, only 37% of those in the coaching group scored zero. The entire distribution shifted upwards for the intervention group, with about 10% more children scoring either 3 or 4 out of 4 than in the control group.

Figure 18: Comprehension scores for Intervention 2 and Control



The descriptive analysis of the shifts in performance for the "Coaching" group is useful to get a sense of the magnitudes of the effects that will be described in the forthcoming regression analysis. In light of the fact that this intervention is intended to shift teaching practice and learning outcomes at a large scale, the magnitudes of the shifts in learning outcomes seen in the graphs above, though not miraculous, do appear substantial enough to warrant consideration for policy scale-up.

Regression analysis of main intervention impacts

Table 12 shows the results of an Ordinary Least Squares regression, where we regress treatment dummies, controlling for stratification, baseline reading performance and a set of controls on aggregate reading proficiency.[8] Columns (2) to (5) includes community-level controls constructed from the 2011 census;[9] column (4) and (5) exclude grade repeaters; columns (3) and (5) exclude multi-grade schools. The motivation for including controls is to account for any incidental differences that may exist between the treatment groups as well as to improve the precision of the estimates by increasing the explanatory power of the model.

[8] We control for district (schools are spread randomly across two districts), school mean score in the Annual National Assessments of 2014 (the most recent standardized school assessment), learner gender, parent education (according to the parent/guardian questionnaire).

[9] This is a community wealth index derived from several questions in the census about household possessions and the proportion of 13 to 18 year-olds in the community that are attending an educational institution.

Figure 19 shows graphically the results from columns (2) and (4). The green bars show the treatment impact and 90% and 95% confidence intervals for the mains regression on the full sample including the main set of controls. The yellow bars show the same results, but excluding grade repeaters.

We see from column (1) that Coaching had a statistically significant impact on aggregate reading proficiency of 0.232 standard deviations. The impacts are smaller and statistically insignificant for Training and Parents – 0.095 and 0.104 standard deviations – and these impacts slightly increase in magnitude (column (2)) when including the community-level controls.

Column (3) excludes multi-grade classrooms. These are settings where the Training and Coaching arms were never expected to have a positive impact, because the intervention was grade-specific. (We aimed to exclude these schools from our original sample, but were unable to perfectly do so.) Indeed, we see substantial increases in effect size when these schools are excluded. The impact on Training and Parents now become statistically significant, with impacts of 0.176 and 9.163 standard deviations respectively.

It is relevant to consider the differences in outcomes when excluding all learners who were repeating grade 1 in 2016, since these learners effectively received a lower dosage of the Training and Coaching treatments. This is because in Year 2 the interventions were only conducted amongst grade 2 teachers. Although some of the grade 1 teachers in 2016 would have benefited from the interventions in 2015, it is quite possible that the effect on their teaching would have been somewhat diminished due to no longer receiving support and not receiving new lesson plans and materials in 2016.

Column (4) reports a regression excluding grade repeaters. The motivation for excluding repeaters is that they were only exposed to half the treatment – because in Year 2 only grade 2 teachers and parents received the intervention support. Since treatment did not significantly predict differences in grade repetition rates, estimating the treatment impact using only the sample of non-repeaters seems defensible. One caveat to note is that repeaters are a systematically lower-performing group and thus any differential treatment effect based on initial ability would also affect the estimates here, over and above the influence of dosage. As expected, the estimated treatment effects for both Training and Coaching are larger and statistically significant: 0.318 and 0.171 standard deviations.

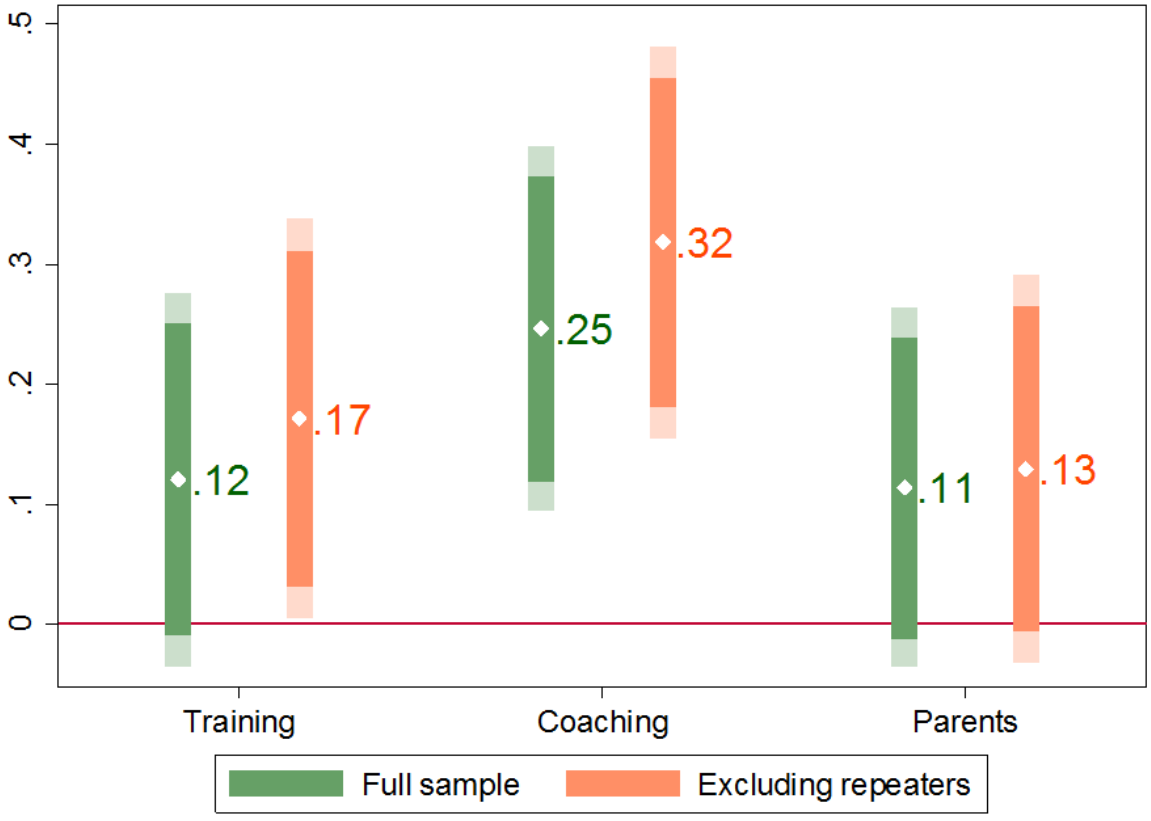
For completeness, column (5) reports a regression where both repeaters and multi-grade classroom are excluded from the sample.

Table 12: Main regression results

	(1)	(2)	(3)	(4)	(5)
Training	0.095	0.120	0.176**	0.171**	0.235***
	(0.078)	(0.079)	(0.080)	(0.085)	(0.087)
Coaching	0.232***	0.246***	0.289***	0.318***	0.359***
	(0.076)	(0.077)	(0.080)	(0.083)	(0.087)
Parents	0.104	0.114	0.163**	0.130	0.188**
	(0.076)	(0.076)	(0.078)	(0.082)	(0.085)
Community controls?	No	Yes	Yes	Yes	Yes
Exclude repeaters?	No	No	No	Yes	Yes
Exclude multi-grade schools?	No	No	Yes	No	Yes
Observations	3,781	3,781	3,540	3,192	2,978
R-squared	0.170	0.173	0.175	0.173	0.174

Note. Each column represents a separate regression, of aggregate reading proficiency on treatment dummies, controlling for stratification, baseline reading performance, female, district, ANA exam performance, and parents' education. Columns (2) to (5) includes community-level controls constructed from the 2011 census; column (4) and (5) exclude grade repeaters; columns (3) and (5) exclude multi-grade schools. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 19: Graphical representation of estimated treatment effects



Dynamic impacts

For comparison, Table 13 examines impacts at midline- i.e. impacts roughly nine months after implementation started. Column (1) shows results for the restricted set of controls, column (2) shows our preferred estimate which also includes community controls; and column (3) excludes multi-grade classes.

The impact of Training on aggregate learning is slightly larger at midline: 0.134 versus 0.12 standard deviations, although the difference is small, and not statistically significant. It therefore does not seem as if Training had any cumulative impact in the second year. We cannot distinguish whether this is because (i) the second year had no impact, or (ii) some of the impacts in the second year were transitory, so the second year of training was required to maintain the gains.

In contrast, we see a steady improvement over time for the Coaching arm. The impact of Coaching increased from 0.14 to 0.246 standard deviation from the first to the second year. The impact of the Parent programme increased from 0.069 at midline to 0.114 standard deviations at endline, although in neither case was the estimated impact statistically significant.

Table 13: Midline regression results

	(1)	(2)	(3)
Training	0.113	0.134*	0.162**
	(0.079)	(0.079)	(0.080)
Coaching	0.124	0.140*	0.177**
	(0.081)	(0.080)	(0.082)
Parents	0.060	0.069	0.078
	(0.075)	(0.074)	(0.073)
Community controls?	No	Yes	Yes
Exclude multi-grade classes?	No	No	Yes
Observations	4,143	4,143	3,576
R-squared	0.196	0.198	0.196

Note. Each column represents a separate regression, of aggregate reading proficiency at midline on treatment dummies, controlling for stratification, baseline reading performance, female, district, ANA exam performance, and parents' education. Columns (2) and (3) includes community-level controls constructed from the 2011 census; column (3) excludes multi-grade schools. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Intervention effects on sub-tests

In much of the literature on measuring early grade reading, composite scores are not derived or reported. This is partly due to limited conceptual meaning of a composite score and partly because the various components or stages of learning to read are considered in relation to each other. In particular, Oral Reading Fluency and Reading Comprehension are viewed as key outcomes, while

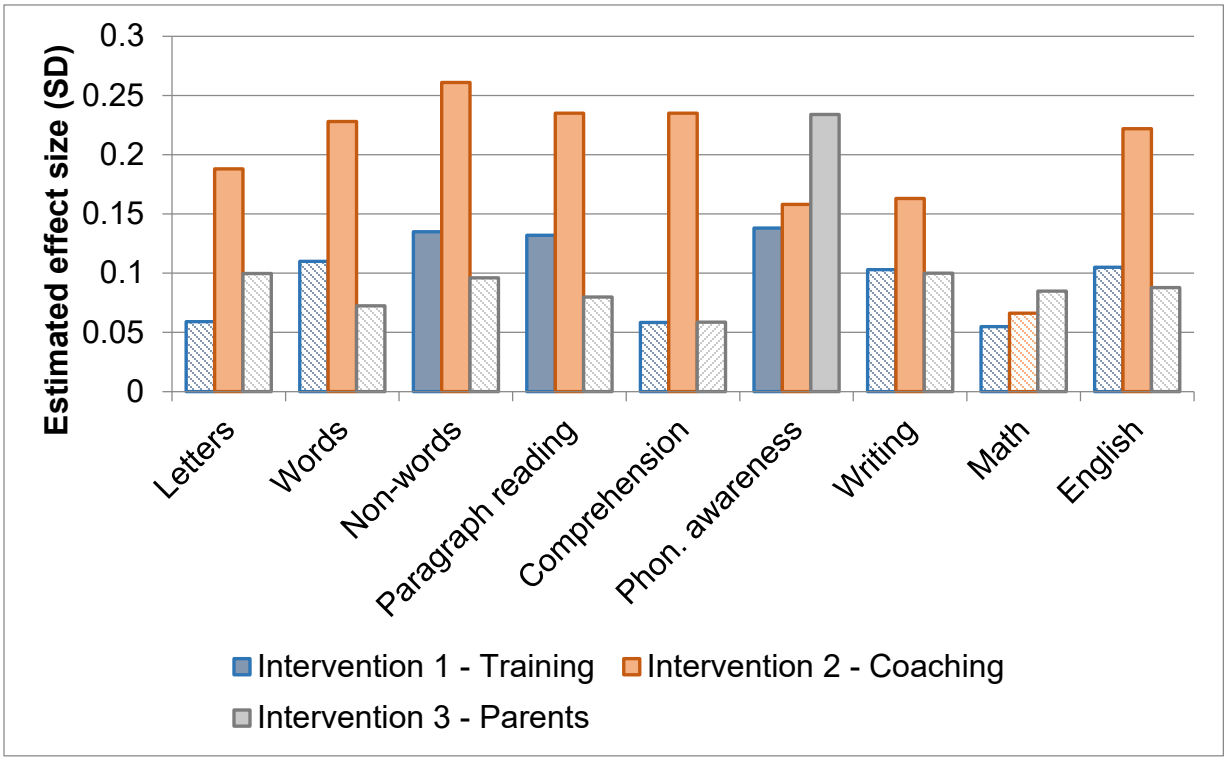
skills such as letter recognition and phonological awareness are viewed as necessary components in learning to read. We chose to derive a composite reading score for the sake of avoiding multiple outcomes for every type of analysis reported in our analysis. However, we also present the main regression results for each of the sub-tests, including the English and mathematics items.

The same set of controls as used in the main specification above was included in these models. Figure 20 graphically presents the results of regressions for each of the sub-tests, with effect sizes expressed in terms of standard deviations. Solid bars represent statistical significance at the 90% level. Several results are worth highlighting. Firstly, there were no negative effects of any intervention on any sub-test. The training intervention appears to have had moderate positive effects on most sub-tests with statistically significant impacts on non-word recognition, paragraph reading (Oral Reading Fluency) and phonological awareness. However, the effect sizes on word recognition and writing were not much smaller. Therefore, it seems unwarranted to make strong conclusions about Intervention 1 being particularly effective or ineffective in specific dimensions of reading literacy.

Intervention 2 registered statistically significant positive effects on all home language sub-tests, with similar effect sizes across the sub-tests. Therefore, it is not as if there is any one or two dimensions of learning that are driving the positive results for Intervention 2. There was no statistically significant effect of Intervention 2 on the rudimentary mathematics test. This means that we have no evidence of a negative effect through crowding out of teaching time for mathematics. Interestingly, we observe a positive effect on English (significant at the 95% level). Here too, there could have been a negative crowding out effect especially since the national curriculum gives teachers in the Foundation Phase the choice to either spend 3 hours a week on English and 7 hours on home language literacy or 2 hours on English and 8 hours on home language literacy. In our intervention, lesson plans were designed on the assumption that teachers would opt for the full 8 hours dedicated to home language literacy. The positive effect on English could be attributable to improved underlying language ability (as obtained through the home language intervention) or simply due to improved classroom management and transferable instructional methods acquired through the coaching intervention. Either way, this is a highly encouraging finding for the intervention.

Although the overall impact of the parent intervention was small and statistically insignificant, it does appear to have had a significant (at the 95% level) positive impact on phonological awareness. It is possible that this is a false positive – 1 out of 20 zero effects could be expected to come out as a false positive at the 95% level, and here we have one out of 9 outcomes for the parent intervention being positive. However, phonological awareness was certainly the component of the learner test that was most directly targeted through the parent meetings. Sound games were a key method taught to parents to use at homes in the development of their child’s phonological awareness.

Figure 20: Effects on sub-tests



Note: Each regression model includes the same set of controls as in Column 2 of Table 12.

Intervention effects on sub-groups of interest

The question of whether an intervention had a differential effect on various sub-groups is important for policy and for understanding when and how these interventions are effective. Therefore, we collected a considerable amount of contextual information about learners, their teachers and the schools they are in. When an intervention has a differential effect on various subgroups this is often referred to in the literature as a heterogeneous treatment effect. However, there is a risk when investigating numerous possible heterogeneities of so-called data mining – that sooner or later a statistically significant result is bound to occur. The existence of a midline assessment as well as an endline assessment reduces this risk somewhat. We were particularly cautious in interpreting observed heterogeneous treatment effects in the Midline report. However, to the extent that we observe similar heterogeneities in the midline data and again after the second year of interventions we can be more confident that a genuine effect is occurring.

Differential intervention effects based on learner characteristics

As laid out in our Pre-Analysis Plan, we investigate learner-level treatment heterogeneity based on learner gender, learner age, and the initial performance of the learner at the start of grade 1. In general,

we investigated these for all three interventions. However, when the results prompted us to dig a little deeper, the focus was mainly on Intervention 2, since this is where the clearest impact was observed.

Table 14 presents the regression results for models which measure the interaction effect for different pupil-level characteristics: gender, age and base aggregate reading proficiency. We include the same set of controls as in columns (2) to (5) in Table 12 above. Column (1) indicates interaction with fender. The main effect for each intervention is here interpreted as the effect of the intervention for boys, which the coefficient on the interaction term represents the additional effect (whether positive or negative) for females. The results indicate that each of the interventions had a positive statistically significant impact for boys, but a lesser effect for girls. For Interventions 1 and 2, this is consistent with what was observed after one year of interventions. Therefore, it seems fair to conclude that the two pedagogical interventions, especially the "coaching" intervention, are helping boys catch up some of the way to girls.

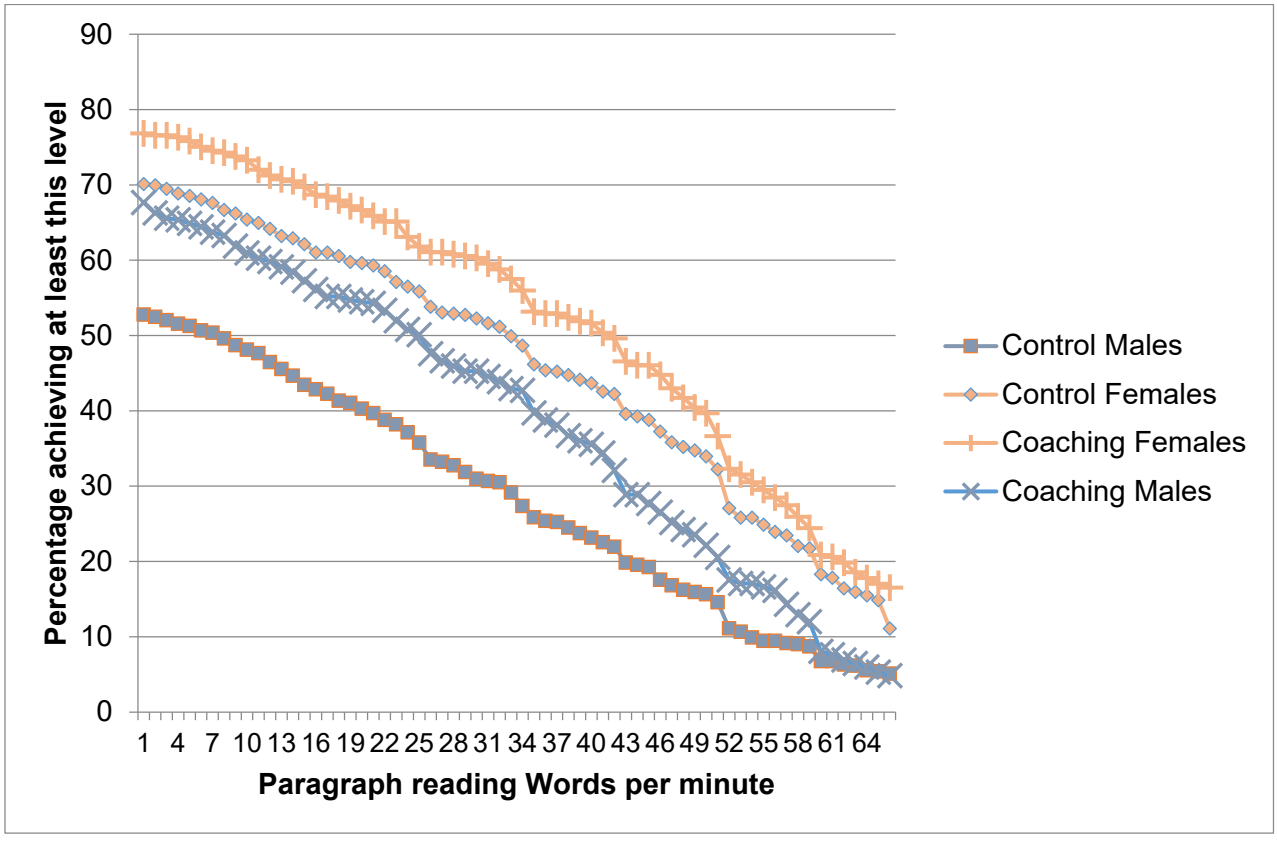
Table 14. Pupil-level heterogeneous treatment impacts

	(1)	(2)	(3)	(4)
VARIABLES	Female	Age	Reading proficiency	
Training	0.172**	0.0710	0.115	0.168**
	(0.0849)	(0.391)	(0.0775)	(0.0802)
Coaching	0.300***	-0.0485	0.245***	0.271***
	(0.0845)	(0.472)	(0.0772)	(0.0767)
Parents	0.158*	-0.0769	0.116	0.119
	(0.0815)	(0.464)	(0.0756)	(0.0812)
Training x group	-0.110	0.00780	0.00244	0.0589
	(0.0829)	(0.0587)	(0.0765)	(0.0925)
Coaching x group	-0.113	0.0458	0.0481	0.224**
	(0.0909)	(0.0728)	(0.0806)	(0.102)
Parents x group	-0.0935	0.0292	0.108*	0.153
	(0.0783)	(0.0693)	(0.0619)	(0.0928)
Training x group squared				-0.0853**
				(0.0374)
Coaching x group squared				-0.0548**
				(0.0258)
Parents x group squared				-0.00999
				(0.0230)
Observations	3,781	3,767	3,781	3,781
R-squared	0.174	0.173	0.175	0.193

Each columns represents a separate regression of treatment dummies and their interaction with a sub-group of interest on reading proficiency. The column heading indicates the variable name of the sub-group of interest. The additional controls are the same as in columns (2) to (5) in Table 12. Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Figure 21 shows a type of inverse cumulative distribution function – it indicates the percentage of children achieving at least at a certain level, in this case on paragraph reading (or Oral Reading Fluency) at the end of Year 2. The graph shows this for boys and girls in the control group and in the "coaching" group. In both groups of schools, girls are outperforming boys, yet the gap has narrowed for the treatment group. For example, about 30% of girls in the control group could not read a single word compared to nearly 50% of boys. In the "coaching" group, however, the percentage of boys who could not read even one word was just over 30% (not far off where girls are in the control group), while nearly 80% of girls could read at least one word. The reasons why this structured learning programme may be helping boys catch up to girls are difficult to identify with certainty. However, based on some of the changes we are observing in classroom practice (to be presented in a later section), it seems reasonable to suggest that the improved classroom management and increased individualized attention may assist boys to be engaged in learning activities. In contexts of large classes it seems especially likely that boys may be less engaged in active learning. Machin and McNally (2005: 363) note evidence from the United Kingdom that boys benefit from highly structured methods of teaching and are more likely than girls to "respond negatively to poor teaching through disengagement and indifference or through disruptive behaviour". One possible explanation which can be ruled out is simply that boys had a lower baseline average reading proficiency and therefore had more "space" to improve. On the contrary, as described later in this section, initially stronger learners benefited more from the programmes.

Figure 21: Oral Reading Fluency by gender and treatment group



Next, column (2) in table 14 shows the regression coefficients from models where the age of the learner is interacted with the intervention group. None of the interaction terms come through significantly either, suggesting that there is no clear story of a differential effect by age. To investigate this further, we ran another model (not shown here) in which age and it's interaction term were allowed to take on a quadratic functional form, instead of a linear functional form. There was some indication that the "coaching" intervention had a larger treatment effect for those in the mid-range of the age distribution, though again this result was not statistically significant.

The next important question with respect to differential treatment effects is whether the impact was different for stronger or weaker learners. If the impact of an intervention is larger for weaker learners this can be seen as an equity-enhancing programme. Nevertheless, the sample for this project has been restricted to non-fee paying schools, which serve about 70% of South Africa's learners. So the most affluent part of the South African population is not included in this analysis.

There are two main ways we investigate the question of whether the impact was different across the learner performance distribution. The first is to investigate whether the impact depended on baseline learner performance: Did initially weaker/stronger learners benefit more from the intervention? The second approach is to observe the performance distributions for treatment and control groups at the end of two years of interventions and see whether the differences in performance vary across the distribution. One would not expect the two approaches to yield completely opposing results, although the calculation is different leading to the possibility of such. Specifically, if there is a lot of rank mobility during the course of the intervention (initially weaker learners moving up the rank distribution and vice versa) but little difference in the overall shape of the distribution this would affect the first approach but not the second. From a policy point of view the second approach is arguably more relevant since the overall level of inequality is perhaps more important than *who* is moving up and down the distribution.[10]

The first approach can be followed by interacting the treatment variable with baseline learner scores. Column (3) in Table 14 shows the results of this approach for a regression in which the relationship between treatment impact and baseline score is assumed to be linear. For all three interventions similar overall effect sizes of the interventions are obtained as compared to the main results discussed earlier. The interaction terms are not significantly different to zero, indicating that there was no obvious linear relationship between baseline score and the impact of any intervention.

It is quite possible, however, that a non-linear relationship may exist (for example, where those in the middle of the distribution benefitted more than the initially weakest and strongest learners). To allow for this possibility, another model was fitted assuming a quadratic functional form. This is achieved by entering the square of the baseline score into the regression equation as well as the interaction between treatment and the squared baseline score. For both Interventions 1 and 2, it appears that the impact increases with learner baseline score but a declining rate, flattening off and even declining slightly at the very top end of the distribution.

[10] An exception would be if historically unjust patterns of inequality are being perpetuated, which would of course be a concern to the policymaker.

A final way to investigate and represent the differential impact of the interventions according to baseline scores, was to run separate regressions for four quartiles of baseline achievement. The results for the coaching intervention are presented graphically below (Figure 22). The results are in line with Table 14 above, as one would expect, but are perhaps easier to comprehend. It would appear from this method, that the impact of coaching may have been zero amongst initially weakest 25% of learners at baseline. The estimated effect then increases slightly with each subsequent quartile, with an estimated impact of 0.33 standard deviations for the initially strongest 25% of learners. It is possible, however, that part of the reason for initially weaker learners benefiting less was because they were more likely to repeat grade 1 in 2016 and thus only receive one year of the treatment. In order to assess this possibility, we reran the regression model excluding repeaters from the sample, on the basis that they did not receive the full two years of the programme. The results are presented in Figures 8 and 9. Clearly, the estimated effect sizes increase noticeably for the bottom two quartiles, where the highest proportions of repeaters are located. Although, the estimated effect of 0.15 standard deviations is still not statistically significant it is still a meaningful effect size. This would suggest that even initially weak learners did benefit from the coaching intervention, as long as they were exposed to two years of the intervention, but we cannot be 90% sure of this. Further, when excluding repeaters (Figure 23), the estimated effect of the intervention appears largest in the mid-range of the distribution. But again, the confidence intervals overlap considerably between the various quartiles so we cannot conclude this with any statistical certainty.

Figure 22: Impact of Coaching by quartiles of baseline performance

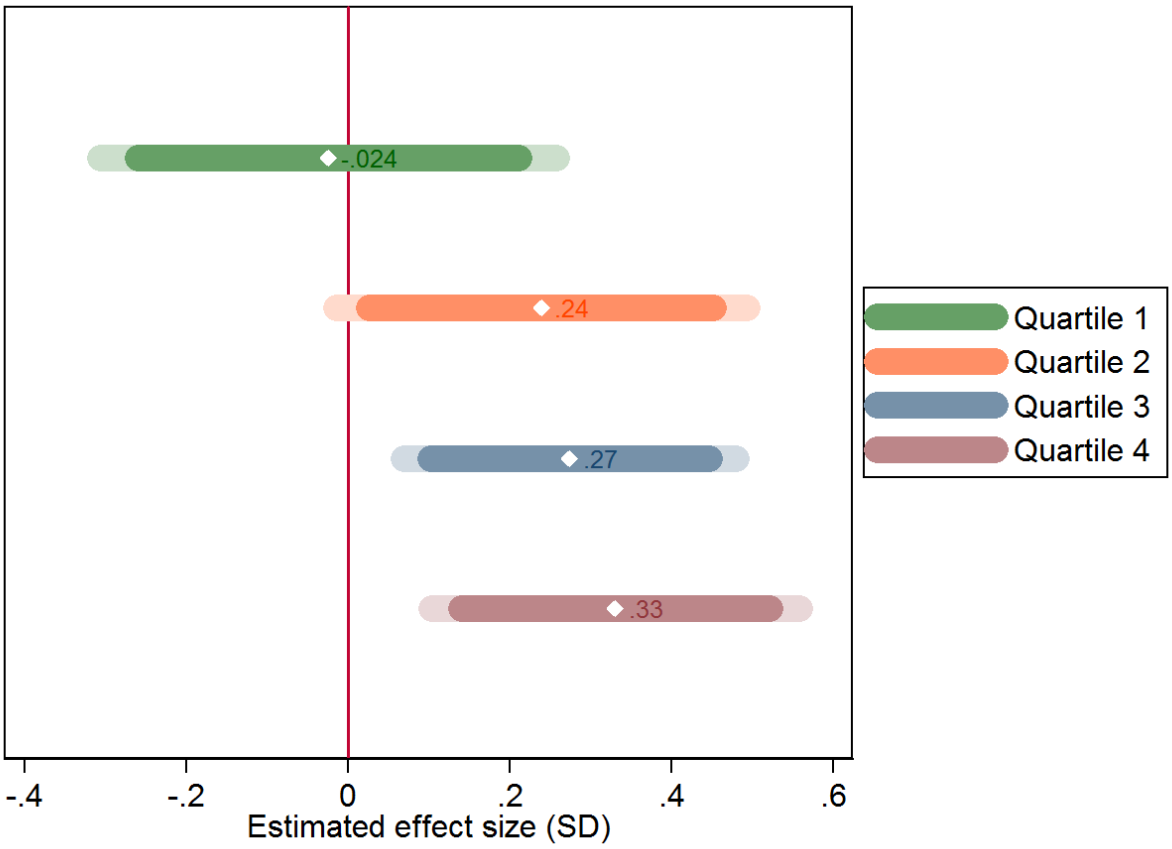
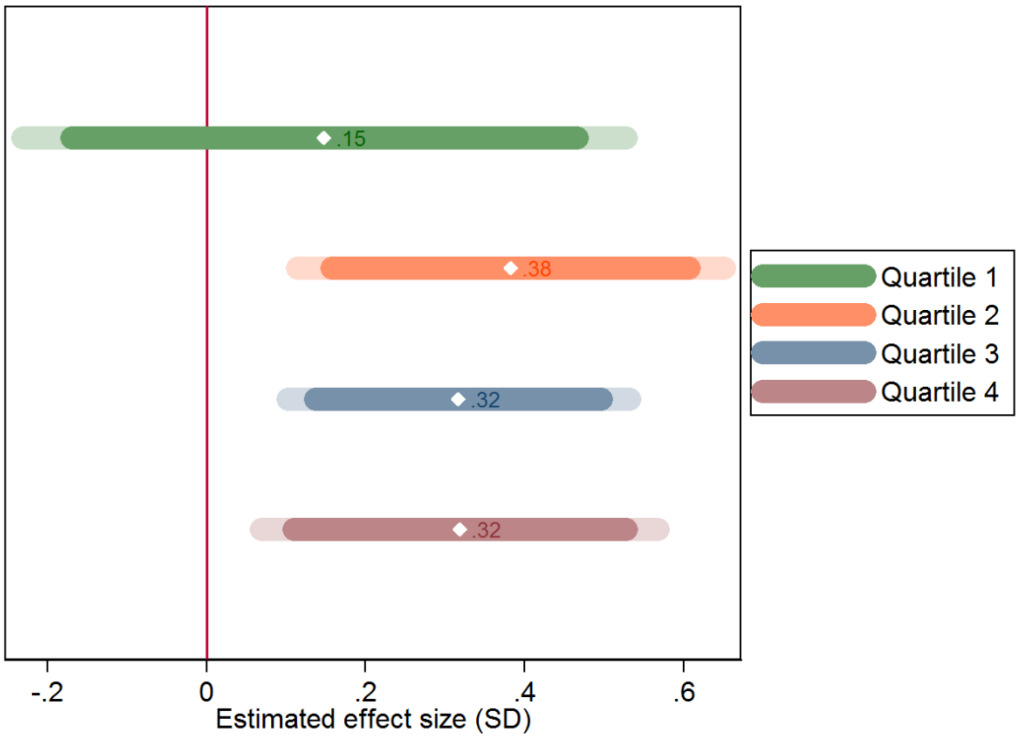


Figure 23: Impact of Coaching by quartiles of baseline performance (excluding repeaters)



The second approach, as discussed above, is to observe the performance distributions for treatment and control groups at the end of two years of interventions and see whether the differences in performance vary across the distribution. The method we use to do this is known as quantile regression. This estimates the effect of the intervention at various points in the distribution of the performance outcome. It asks, for example, what is the impact on the 10th percentile of performance, on the 20th percentile, on the 30th percentile, etc. We present the results of quantile regressions measuring the effect of the "coaching" intervention on the Midline (Year 1) composite test score and the Year 2 composite score in Figures 24 and 25, respectively. In both graphs the line plots the estimated effects across the performance distribution, while the shaded area represents the 95% confidence interval around the estimated effects.

The impact on both Year 1 and Year 2 scores was near zero at the 10th percentile of the distribution, and then quickly rose across the distribution, peaking at the 80th percentile in the case of the Year 1 results and the 50th percentile in the case of the Year 2 results. This confirms the earlier analysis indicating that the impact of the intervention appears to have been greatest in the middle to upper parts of the performance distribution, though not at the very top of the distribution. Importantly, there is no evidence of a negative effect for any part of the performance distribution.

One implication of this finding is that structured learning programmes, making use of lesson plans, may benefit a certain section of the performance distribution more, depending on the level at which the lessons are pitched, but at least in the case of this particular programme, no group was harmed and the level appears to have been pitched towards the middle of the learner proficiency range.

Figure 24: Quantile regression of "Coaching" Intervention impact on Year 1 scores

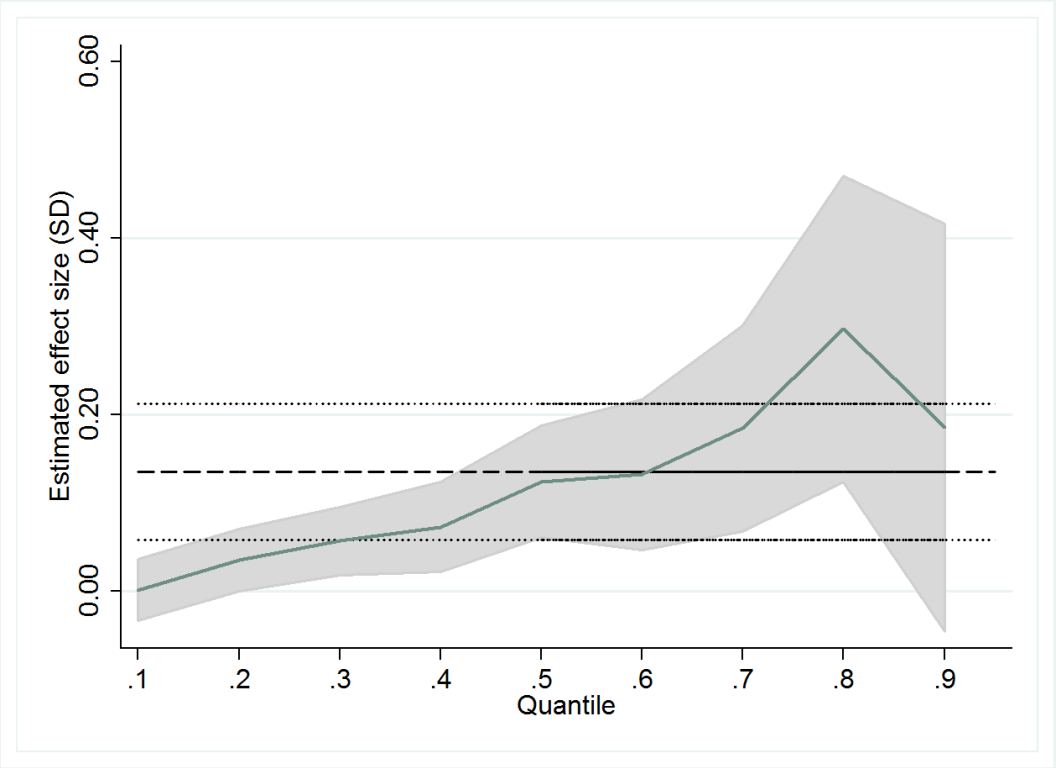
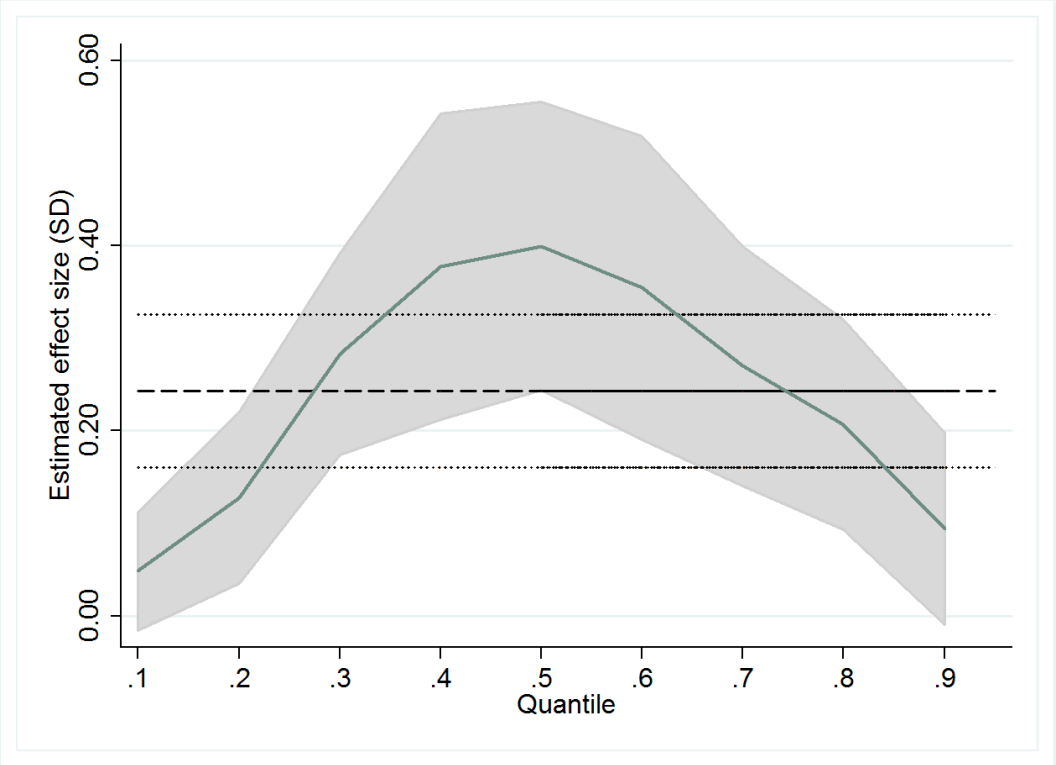


Figure 25: Quantile regression of "Coaching" Intervention impact on Year 2 scores



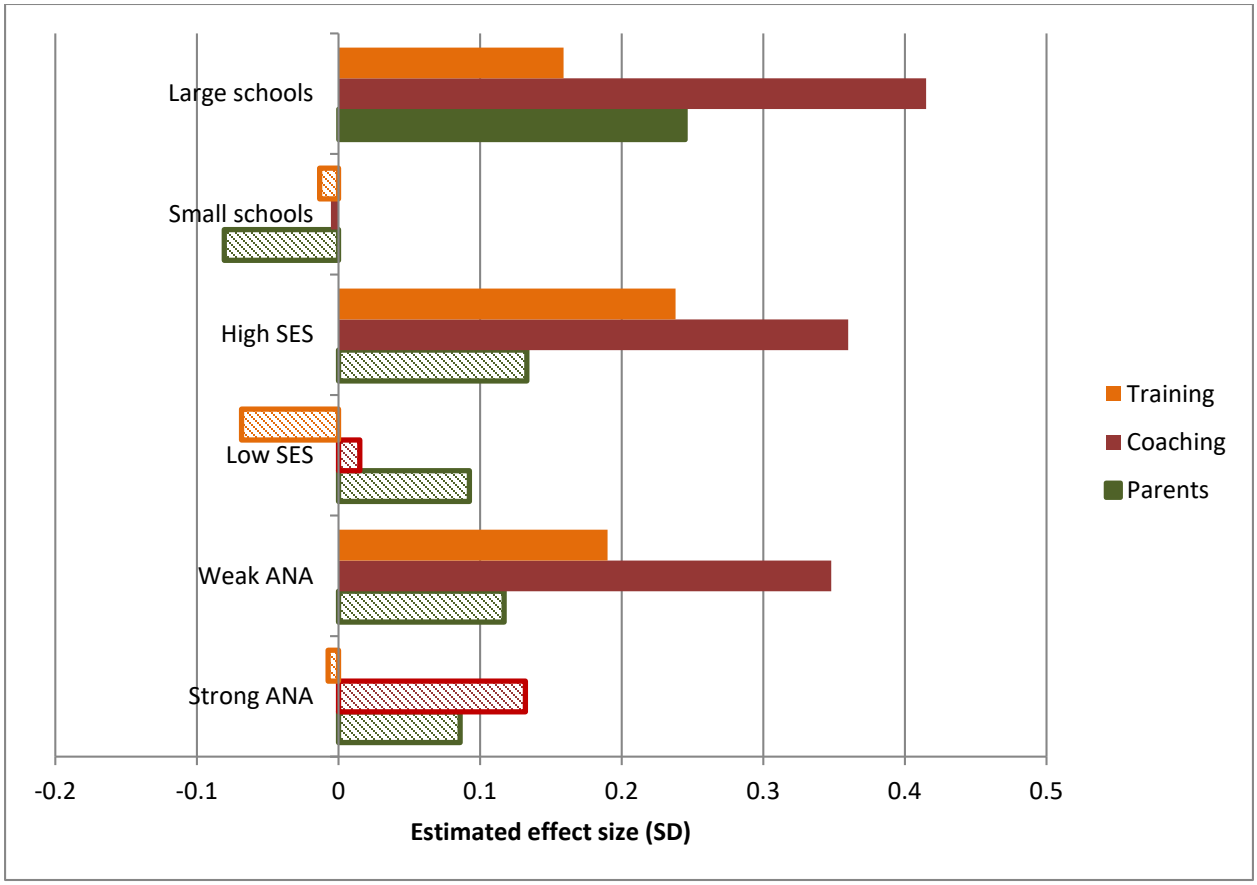
Differential intervention effects based on school characteristics

A number of school-level variables might be expected to influence the effectiveness of the EGRS interventions. The academic performance level of the school may be an indication of the school's readiness for a structured pedagogic programme. School size may affect the intervention in so far as it affects the number of classes in a grade, which affects the role of peer support amongst teachers in the school. School size also interacts with the rurality of a school since deep rural schools are considerably smaller on average. The EGRS has been implemented in two education districts, and the level of district support for the programme may influence its effectiveness. The condition of school facilities may be a proxy for community level poverty or for school management quality, both factors which may influence the effectiveness of an additional school support intervention. The levels of parental education and employment are also proxies for socio-economic status, which may influence the effectiveness of, in particular, the parental involvement intervention.

The EGRS randomization was stratified by school size, school socio-economic status and school academic performance (ANA2014), and so it made sense to split the sample according to the original strata (regression results reported in Figure 26). When running separate regressions for large schools and small schools it is apparent that none of the interventions had any significant effect within the strata consisting of smaller schools, but all three interventions had significant positive effects within the strata consisting of larger schools. Similarly, the training and coaching interventions had large statistically significant impacts within the strata consisting of relatively higher socio-economic status schools but no impacts within the strata consisting of relatively low socio-economic status schools. Somewhat unexpectedly the training and coaching interventions had clear positive impacts within the strata with weaker performance in the 2014 ANA, but no significant impact amongst schools with someone better performance in ANA. All three of these subgroup effects may well be interacting with another urban-rural subgroup effect, a result which is reported on in Table 15, where it becomes clear that no interventions had a significant positive effect within rural areas but all three had significant positive effects in the urban township settings. Urban schools are more likely to be large schools and also higher socio-economic status schools on average. It would also appear that the significant positive effects within those strata consisting of initially weaker performing schools is largely being driven by a group of initially low performing urban township schools, which experienced strong gains throughout the intervention period.

There is also a significant interaction between the education district and effectiveness of interventions. Again this may have some overlap with the urban-rural interaction since the district in which the interventions had greater effects is also the district with a higher proportion of urban township schools. No significant interaction between the condition of school facilities and the effect of interventions was observed.

Figure 26: Estimated treatment effects for various subgroups based on stratification of the sample



Note: Solid bars denote an estimated effect that is statistically significantly different to zero with at least 90% level of confidence, whilst a shaded bar denotes that we cannot be 90% sure that the effect is different from zero. The regression controls are the same as in columns (2) to (5) in Table 12. Standard errors were clustered at the school level.

Table 15: Estimated treatment effects for various subgroups of schools

	(1)	(2)	(3)
	Rural	District	Facilities - good condition
Training	0.446***	0.0560	0.0872
	(0.163)	(0.0860)	(0.0979)
Coaching	0.747***	0.164**	0.157*
	(0.148)	(0.0800)	(0.0884)
Parents	0.382***	0.0373	0.0573
	(0.133)	(0.0840)	(0.0884)
Training x group	-0.387**	0.330	0.0930
	(0.180)	(0.203)	(0.165)
Coaching x group	-0.625***	0.365*	0.278
	(0.176)	(0.199)	(0.172)
Parents x group	-0.309*	0.362**	0.227
	(0.160)	(0.181)	(0.181)
Observations	3,781	3,781	3,744
R-squared	0.181	0.178	0.178

Note. Each columns represents a separate regression of treatment dummies and their interaction with a sub-group of interest on reading proficiency. The column heading indicates the variable name of the sub-group of interest. The additional controls are the same as in columns (2) to (5) in Table 12. Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

It is difficult to know exactly what the main causal drivers were behind the somewhat overlapping subgroup effects across urban and rural locations, education districts, school size and school socio-economic status. Whilst rural schools are poorer and smaller than urban township schools, neither of these factors fully account for the urban-rural subgroup effect. Similarly we note that both learner and teacher absenteeism is higher in rural areas, but neither does this account for the urban-rural heterogeneous treatment effect. It may be that a combination of disadvantageous factors in deep rural settings precludes interventions from having a positive impact.

Table 16 shows that the level of parent education and parent employment rates, as estimated by the school principal, did not seem to play a significant role in determining the effectiveness of interventions.

Table 16: Treatment heterogeneity according to Principal's assessment of parental education and employment levels

	Low parent education (Wave 2)	Low parent education (Wave 3)	Low parent employment (Wave 2)	Low parent employment (Wave 3)
Training	0.0741	0.156	0.155	0.231**
	(0.135)	(0.139)	(0.106)	(0.110)
Coaching	0.178	0.232*	0.244***	0.304***
	(0.112)	(0.121)	(0.0898)	(0.107)
Parents	0.00939	0.138	0.182*	0.198*
	(0.113)	(0.125)	(0.101)	(0.107)
Training x group	0.109	-0.0608	-0.0592	-0.221
	(0.172)	(0.166)	(0.152)	(0.153)
Coaching x group	0.141	-0.0561	-0.0229	-0.216
	(0.151)	(0.154)	(0.177)	(0.172)
Parents x group	0.198	-0.0379	-0.156	-0.199
	(0.156)	(0.160)	(0.146)	(0.155)
Group	-0.123	-0.0980	-0.0238	0.00819
	(0.106)	(0.111)	(0.103)	(0.111)
Constant	-1.647***	-1.399***	-1.595***	-1.563***
	(0.458)	(0.462)	(0.433)	(0.459)
Observations	3,718	3,725	3,744	3,708
R-squared	0.176	0.177	0.175	0.178

Note. Each columns represents a separate regression of treatment dummies and their interaction with a sub-group of interest on reading proficiency. The column heading indicates the variable name of the sub-group of interest. The additional controls are the same as in columns (2) to (5) in Table 12. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Differential intervention effects based on teacher and classroom characteristics

We also investigate whether certain teacher characteristics influenced the effectiveness of interventions. Teacher Setswana reading proficiency was another characteristic hypothesized to potentially influence the effectiveness of interventions. This was measured in two ways. Firstly, we measured the reading fluency of teachers based on self-reported words per minute after reading from a given text in a minute. Secondly, we conducted a short comprehension test. The effectiveness of interventions did not depend on teacher fluency (self-reported words per minute). The teacher comprehension tests were conducted at midline endline, so after the implementation of the intervention. However, teacher comprehension scores were not significantly different across the four treatment groups.

Table 17 shows heterogeneous treatment impacts based on grade 2 teachers' performance in the comprehension test, teacher experience, and the class size (i.e. number of pupil taught by the teacher). Data is at a pupil level, but observations are reweighted so that each teacher receives equal weight (i.e. weighted observations by the inverse probability of a pupil being matched to a teacher, so

that observations with many pupils assigned to the same teacher receive a lower weight). Colum (1) shows that weak evidence that pupils taught by grade 2 teachers who performed better in the comprehension test benefited more from the Training.

Table 17. Heterogeneous treatment impact by grade 2 teacher characteristics

VARIABLES	(1) Comprehension	(2) Experience	(3) Class size	(4) Class size
Coaching	-0.0302	0.499***	-0.340	-1.829***
	(0.272)	(0.151)	(0.329)	(0.665)
Training	-0.389	0.182	-0.253	-2.732***
	(0.273)	(0.164)	(0.369)	(0.842)
Coaching x group	0.486	-0.0105	0.0153*	0.0977***
	(0.415)	(0.00705)	(0.00804)	(0.0327)
Training x group	0.833**	-0.00220	0.00905	0.146***
	(0.418)	(0.00731)	(0.00887)	(0.0434)
Coaching x group squared				-0.00105***
				(0.000390)
Training x group squared				-0.00176***
				(0.000540)
Observations	2,264	2,276	2,285	2,285
R-squared	0.173	0.170	0.177	0.190

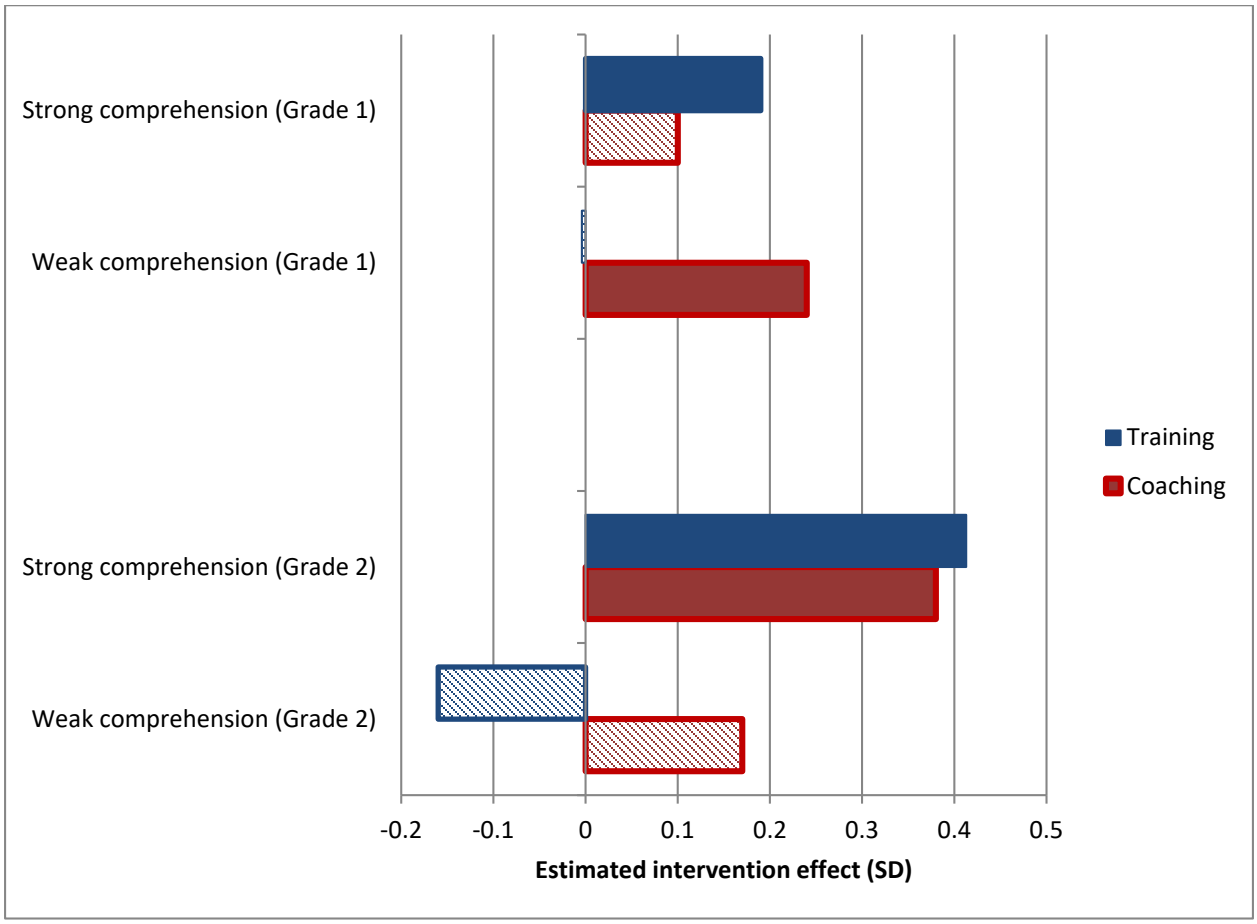
Note. Each columns represents a separate regression of treatment dummies and their interaction with a sub-group of interest on reading proficiency. The column heading indicates the variable name of the sub-group of interest. The additional controls are the same as in columns (2) to (5) in Table 12. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 27 graphically depicts the estimated effect sizes of the training and coaching interventions for split samples based on a binary distinction in teacher reading comprehension. We ran two sets of subgroup analysis, first splitting the sample based on grade 1 teacher comprehension scores and secondly splitting the sample based on grade 2 teacher comprehension scores. In both cases it would appear that the training intervention only worked with teachers who have stronger comprehension scores. However, for the coaching intervention the grade 1 and grade 2 analyses yield somewhat opposing results. In the light of these somewhat inconsistent results as well as the somewhat rudimentary nature of the teacher comprehension test, we feel the evidence around a differential treatment effect based on teacher comprehension scores remains a little thin.

Columns (3) and (4) show the interactions with class size. There is no strong linear relationship, but there is a very strong non-linear (concave) relationship between treatment effect size and class size for both Coaching and Training. To further unpack this non-linear relationship, Figure 28 depicts this visually, splitting the sample by tercile of class size. Both the teacher support interventions (“training” and “coaching”) had the largest impacts in relatively large classes (38 to 44 learners). In smaller

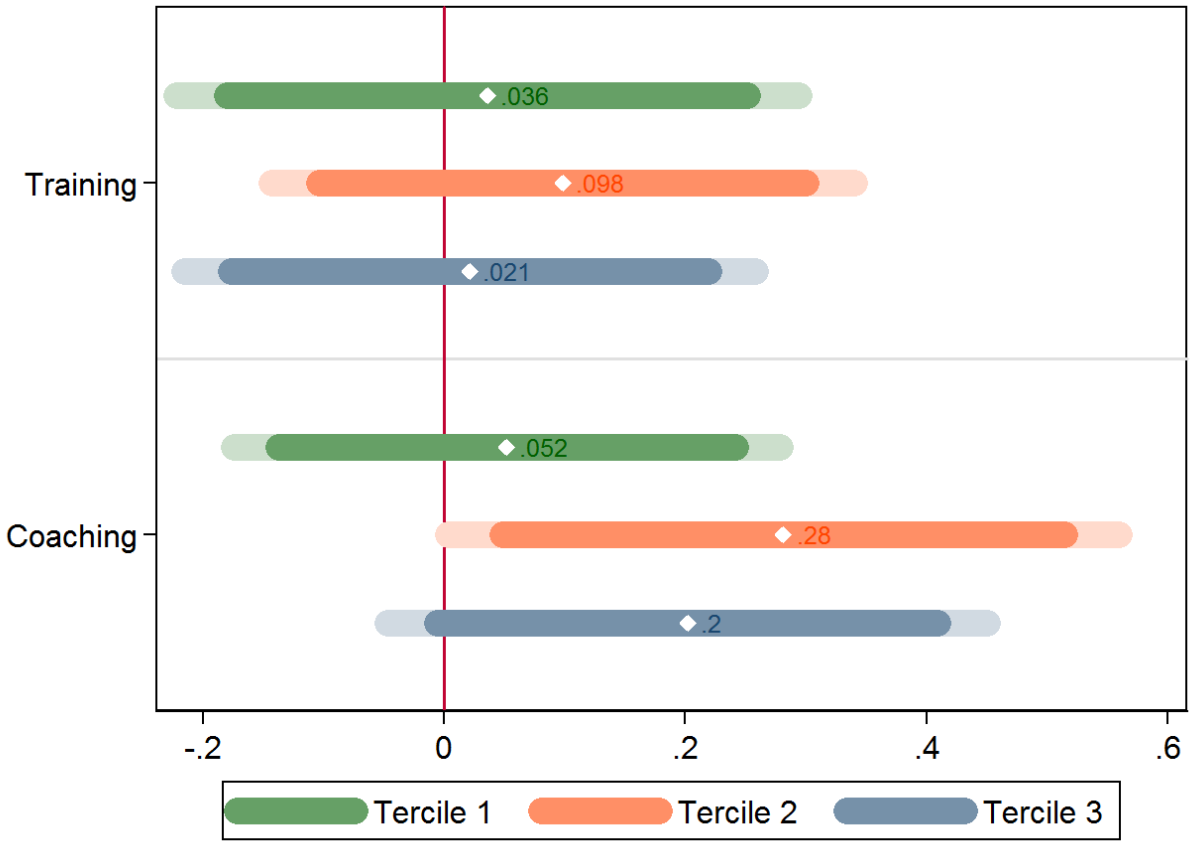
classes, it may be that teachers in the control schools are already able to effectively manage classrooms, provide structured learning and differentiated attention to a variety of learners, but in larger classes the EGRS interventions helped teachers to do so in a more challenging environment. In the very largest classes (50 plus learners), however, the EGRS interventions were somewhat less effective, possibly indicating that beyond a certain threshold it is very difficult to conduct effective teaching.

Figure 27: Estimated effects of Interventions 1 and 2 depending on teacher comprehension scores



Note: Solid bars denote an estimated effect that is statistically significantly different to zero with at least 90% level of confidence, whilst a shaded bar denotes that we cannot be 90% sure that the effect is different from zero. The regression controls are the same as in columns (2) to (5) in Table 12. Standard errors were clustered at the school level.

Figure 28: Treatment effects by tercile of class size



Notes: Children repeating grade 1 in 2016 are excluded from this analysis since we did not record class size in grade 1 classes of 2016; In the cases of 157 grade 2 learners (5% of the relevant sample) class size was missing and was imputed from the size of the other grade 2 classes in the same school. The results were negligibly different when excluding this 5% of learners.

Differential intervention effects based on parent characteristics

Finally we investigated whether any parent or guardian characteristics influenced the effectiveness of the interventions. Table 18 reports the results of regressions investigating whether parental identity, education and parental literacy (as proxied for by writing an open ended response at least twice) had any significant bearing on the treatment effect. Evidently, none of these factors appear to significantly determine the effectiveness of any of the interventions.

Table 18: Intervention effects based on parental identity, education and literacy

	Mother	Parent matric	Parent literacy proxy (wrote responses)
Training	0.0637	0.0776	0.0893
	(0.106)	(0.0863)	(0.0920)
Coaching	0.195*	0.234***	0.287***
	(0.111)	(0.0891)	(0.0883)
Parents	0.0454	0.0934	0.108
	(0.0959)	(0.0858)	(0.0854)
Training x group	0.0727	0.115	0.0615
	(0.0873)	(0.106)	(0.0980)
Coaching x group	0.0652	0.0590	-0.0750
	(0.0997)	(0.101)	(0.0828)
Parents x group	0.0891	0.0694	0.0276
	(0.0908)	(0.0913)	(0.0873)
Group	-0.0459	0.0871	0.0954*
	(0.0634)	(0.140)	(0.0516)
Constant	-1.617***	-1.598***	-1.680***
	(0.449)	(0.471)	(0.450)
Observations	3,781	3,418	3,781
R-squared	0.173	0.161	0.176

Notes: *Note*. Each columns represents a separate regression of treatment dummies and their interaction with a sub-group of interest on reading proficiency. The column heading indicates the variable name of the sub-group of interest. The additional controls are the same as in columns (2) to (5) in Table 12. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 19 reports the results of regression is investigating whether parental involvement in their child's education (as proxied for by reading with their child, checking homework and taking responsibility for the child's education) influence the effectiveness of interventions. Again, none of these factors appeared to have played a significant role, as evidenced by the non-significant coefficients on the various interaction terms.

Table 19: Intervention effects based on parental involvement in child's education

	Reads to Child	Checks homework	Takes responsibility
Training	0.215	0.169	0.135
	(0.160)	(0.203)	(0.105)
Coaching	0.413***	0.508***	0.205**
	(0.155)	(0.177)	(0.0971)
Parents	0.111	0.284	0.0545
	(0.157)	(0.175)	(0.0969)
Training x group	-0.0211	-0.00557	0.0988
	(0.0324)	(0.0367)	(0.134)
Coaching x group	-0.0426	-0.0520	0.0501
	(0.0336)	(0.0331)	(0.120)
Parents x group	0.00232	-0.0269	0.0253
	(0.0325)	(0.0338)	(0.128)
Group	0.0892***	0.0950***	-0.104
	(0.0223)	(0.0206)	(0.0798)
Constant	-1.853***	-1.917***	-0.960*
	(0.483)	(0.517)	(0.521)
Observations	3,059	2,660	1,933
R-squared	0.167	0.168	0.154

Note. Each columns represents a separate regression of treatment dummies and their interaction with a sub-group of interest on reading proficiency. The column heading indicates the variable name of the sub-group of interest. The additional controls are the same as in columns (2) to (5) in Table 12. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Unfortunately, it was not possible to estimate the impact of the parent involvement treatment on those who attended a larger number of parent meetings. One cannot simply compare those with high attendance to the control group since there may well be self-selection into high attendance. One possible approach that was considered was to run a two-stage least squares regression using assignment to the parent intervention as an instrument for attending at least one session. This approach relies on the assumption that any impact of the treatment would not have spilled over to those children whose parents did not attend any sessions. But more problematically, only about 20% of parents did not attend a single session in the two years of parent meetings. As a result, this instrumental variables approach did not significantly change the results.

Another approach that was considered was to predict parent attendance (obviously only using Treatment 3 observations) and thus to create a propensity score for likelihood of attendance for all observations across treatment groups based on characteristics of parents. Then one could match learners across treatment and control based on propensity to attend meetings and then see if there was a higher estimated impact of treatment on those with greater propensity to attend. However, as Table 20 below shows, we were unable to explain much of the variation in parent attendance, which caused this approach to be fruitless. It is perhaps interesting to note that parents who reportedly struggled to attend meetings due to work in fact were less likely to attend the Treatment 3 meetings, whilst children whose mother or grandmother responded to the parent

questionnaires were better represented at the Treatment 3 parent meetings (than those whose father, sibling or other person responded to the parent questionnaires). Future experiments testing parent involvement programmes may wish to include a "placebo" round of parent meetings in all schools including the control schools in order to identify a comparable group of "attenders" and thus estimate the treatment effect on the treated.

Table 20: Factors predicting attendance of Treatment 3 parent meetings

VARIABLES	(1) Parent attendance
Responded to Wave 1 questionnaire	1.187
	(1.115)
Parent responsible	1.201
	(1.076)
Difficulty attending due to work	-1.569*
	(0.836)
Parental education level: low	0.0427
	(1.171)
Parental employment level: low	-0.734
	(1.111)
Respondent: mother	1.432*
	(0.734)
Respondent: grandmother	2.106**
	(0.944)
Constant	4.735**
	(1.804)
Observations	633
R-squared	0.023

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

A separate report on family influences on early grade literacy is available and contains a lot more detail regarding the correlates of parent characteristics with early grade learning outcomes, based on the data from the study.

Explaining the rural-urban divide

The substantial difference in impact of between urban and rural schools requires some explanation. In this section we test for the potential explanations for the larger impact in urban schools, relative to rural schools. Large difference in impacts could either be due to:

I. Chance. The small number of urban schools in our sample means that results might be driven by outliers. With a small sample we are also more likely to have a “Type-M" error (Gelman and Carlin, 2014)

II. Lower quality of implementation in rural areas.

III. Lack of complementary inputs for (i) changing teaching practice and (ii) improving learning. Teachers in rural schools might have low capacity and motivation, leading to incomplete adoption of the new practices; or school management might be weak and school leaders do not provide sufficient support and oversight to facilitate adoption the new practices; or learning and support between early-grade teachers might be crucial to accelerate adoption, yet absent in very small schools with only one teacher per grade. Low adoption could also be driven by high teacher turn-over. Moreover, even if teaching practice does change, there might be other binding constraints in translating better teaching to lead to higher reading: illiterate parents may not be able to provide the necessary complimentary support at home; or pupils' baseline level of reading proficiency may be too low to benefit from the new teaching practices.

Chance

As a starting point, since there are only 52 urban schools in our sample (22 percent of schools), there is some concern that the large impacts in urban schools are driven by a few outliers. As a further robustness check, we perform a jackknife resampling technique, running the interaction regression multiple times and each time dropping a different school. Table 21 shows the descriptive statistics for the stored regression coefficients and p-values from the 230 regressions. We see that the distribution of interaction effects (i.e. the difference in treatment impacts between rural and urban areas) remain large for all treatment arms. However, the smallest possible estimate for the interaction effect with the parent treatment is statistically insignificant at the 10 percent level (p=0.11). This is not the case for Training and Coaches. This is especially the case for Coaching arm, where the coefficient on the interaction term varies between 0.56 and 0.71, with statistical significance at the 1 percent level.

Table 21. Summary statistics of the distribution of coefficients and p-values for the interaction between treatment arms and rural, using the jack-knife resampling method.

	(1)	(2)	(3)	(4)
	N	Mean	Min	Max
Coefficient: Rural x Training	230	-0.38722	-0.44596	-0.30887
P-value: Rural x Training	230	0.03374	0.01279	0.07442
Coefficient: Rural x Coaches	230	-0.62506	-0.70569	-0.55759
P-value: Rural x Coaches	230	0.00049	0.00002	0.00177
Coefficient: Rural x Parents	230	-0.30932	-0.37337	-0.23913
P-value: Rural x Parents	230	0.05575	0.01670	0.11806

Note: Descriptive statistics for the distribution of coefficients and p-values for 230 regressions of treatment dummies and their interaction with a rural dummy, each time iteratively dropping a different school from the sample. Regression controls are the same as in columns (2) to (7) in Table 12.

Quality of implementation

Next, we test if the quality of implementation was weaker in rural schools. Another factor which appears to be different between urban and rural areas is the extent to which schooling and coaching meetings were disrupted.

As a starting point for examining the quality of implementation, we first test if the differences between rural and urban schools are driven by an exceptional coach that mainly served urban schools. Column (1) in Table 22 shows regression results, where we control separately for each coach and its interaction dummy, estimated using the following equation:

$$y_{isb1} = \beta_0 + \sum_{t=1}^{3} \beta_t (\text{Coach t})_s + \sum_{t=1}^{3} \beta_{t+3} (\text{Coach t} \times \sigma)_s + X'_{isb0}\Gamma + \rho_b + \varepsilon_{isb1},$$

It is clear from column (1) in Table 22 that we can rule out a coach-effect: for all coaches the treatment impact is substantially larger in urban compared to rural schools.

Table 22: Quality of implementation - Coach visits

	Full sample		Rural schools
	(1)	(2)	(3)
Coach 1	0.475	-0.133	0.0444
	(0.335)	(0.151)	(0.285)
Coach 2	0.471*	-0.109	0.00705
	(0.264)	(0.182)	(0.282)
Coach 3	0.848***		
	(0.152)		
Coach 1 x Rural	-0.277		
	(0.353)		
Coach 2 x Rural	-0.420		
	(0.280)		
Coach 3 x Rural	-0.710*		
	(0.383)		
Rural	0.136		
	(0.157)		
Number of visits		0.0210	0.0472*
		(0.0245)	(0.0266)
Observations	2,140	815	604
R-squared	0.199	0.209	0.253

Note. The column heading indicates the variable name of the sub-group of interest. Standard set of controls included. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Next we investigate if the quality of implementation was worse in the rural schools.

Figure 29 shows the distribution of coach visits (based on coaches' own reports) to schools, broken down by urban vs rural. It is clear that rural teachers were visited less often by coaches. We see that teachers from rural schools were less likely to receive visits from reading coaches: the coaches reported to visit teachers from urban schools on average 10.4 times, compared to 8.7 in rural areas. Moreover, this trend was consistent across coaches: every coach paid rural schools at least one fewer visit. There is also much higher variation in the number of visits in rural schools, ranging from 4 to 13.

Figure 29: Number of coaching visits to urban and rural schools

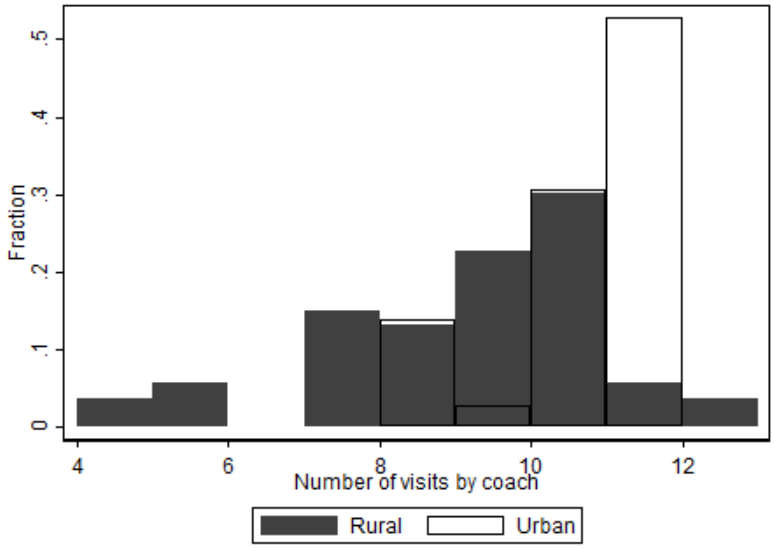


Figure notes. Distribution of the number of times that a school in the "Coaching" arm was visited by a reading coach, split by urban and rural schools.

Fewer visits in rural schools could be due to both the difficulty in reaching remote locations and weak management. The reasons coaches provided for fewer visits to some teachers include:

(i) Schools difficult to access in rainy season.
(ii) Riots and social unrest made school access difficult
(iii) High teacher turnover.
(iv) Absent teachers,[11] or teachers who leave early (for transport reasons), or who could not meet because of management responsibilities.[12] Teacher absence is also far more disruptive in small schools, since there are fewer teachers to fill in for the absent teacher.

Table 23 shows the percentage of schools in which coaching visits were disrupted for a variety of reasons, according to Class Act monitoring data. Factors such as difficulties in accessing schools in the rainy season, social unrest, teacher absenteeism, attending memorial services during normal school hours, choir competitions and sporting events interfering with normal teaching time, and difficulties in communicating with teachers due to poor mobile phone coverage, were all more prevalent in rural schools than in urban schools. It seems fair to conclude that if these factors commonly disrupted coaching visits they would also be likely to regularly disrupt normal teaching time in rural schools, and sufficient teaching time is probably a precondition for a structured pedagogic program following prescribed schedule to be successful.

[11] Based on the monitoring reports: "Absenteeism was noted as a coping mechanism by some teachers who wanted to avoid the issue of accountability which is linked to coaching."

[12] "In some small schools the school principal was also the Grade 1 or 2 teacher, resulting in some coaching visits being cancelled and/or postponed due to management responsibilities."

Table 23: Percentage of schools "Greatly affected" by various disruptions

	URBAN	RURAL
School access in rainy season	0%	48%
Disruptions as a result of unrest	7%	39%
Teacher absenteeism	7%	35%
Memorial services	0%	33%
Choir competitions	14%	21%
Sporting events	7%	24%
Poor cell phone coverage	0%	42%
OTHER (e.g. No electricity)	7%	21%

Source: Class Act monitoring data

To further examine the hypothesis that the number of visits played an important role, we test if there is an association between the size of the treatment impact and the number of visits. Columns (2) and (3) in Table 22 ("Quality of implementation - Coach visits") report results where we regress aggregate reading proficiency on the number of visits and workshops, controlling for each coach and the non-compliant school separately. In column (3) the sample is restricted to only rural teachers. We see that there is a statistically significant positive relationship in rural areas between the aggregate reading proficiency and the number of visits: one more visit is associated with a .047 standard deviation improvement in reading proficiency. Of course this result may be endogenously determined and should therefore be interpreted with caution.

Complementary inputs

Next, we examine if there are any complementary inputs to learning that are lacking in rural schools. To summarize results we do not find large differences between urban and rural schools for most characteristics. The most notable difference is that rural schools are more likely to have either very large or very small classes - precisely the settings where Training is less effective. However, even after accounting for all observed characteristics, we still cannot fully explain away the difference in impacts between urban and rural schools for the Coaching arm.

In order to test if these observed differences between urban and rural schools can explain the variation in treatment size between urban and rural schools, we include multiple interactions with observed characteristics to see if the rural interaction effect is reduced. Column (1) in Table 24. shows the results of interaction with rural, restricting the sample to schools that are not multi-grade, to pupils that are not repeaters, and schools where we have data on class size. This assures that the sample remains the same across the different models. Column (2) adds interactions with class size and class size squared; column (3) adds interactions with pupil baseline reading proficiency; and column (4) includes interactions with teacher experience. By comparing the coefficients of the (Rural x Training) and (Urban x Training) interaction terms, we see that the difference in impacts between rural and urban schools is reduced for both programs when we account for treatment heterogeneity due to class size. The reduction is large enough so that the treatment difference for Training is no longer significant. We do not see similar reductions when accounting for treatment heterogeneity for other variables. Notably, the size of the class interactions are robust.

Table 24. Accounting for the geographic treatment heterogeneity

	(1)	(2)	(3)	(4)
Training	0.521***	-1.948*	-2.042**	-2.081**
	(0.183)	(0.998)	(0.989)	(0.990)
Coaching	0.807***	-1.534*	-1.710*	-1.752*
	(0.178)	(0.893)	(0.893)	(0.912)
Rural x Training	-0.349*	-0.291	-0.303	-0.273
	(0.197)	(0.199)	(0.201)	(0.199)
Rural x Coaching	-0.577***	-0.486**	-0.451**	-0.427**
	(0.211)	(0.209)	(0.211)	(0.211)
Class size x Training		0.120**	0.124**	0.125**
		(0.0485)	(0.0479)	(0.0492)
Class size x Coaching		0.105**	0.111***	0.118***
		(0.0406)	(0.0408)	(0.0428)
Class size squared x Training		-0.00144**	-0.00148**	-0.00150**
		(0.000605)	(0.000596)	(0.000605)
Class size squared x Coaching		-0.00118**	-0.00123**	-0.00133**
		(0.000488)	(0.000493)	(0.000529)
Reading proficiency x Training			-0.0726	-0.0724
			(0.106)	(0.108)
Reading proficiency x Coaching			-0.0207	-0.0112
			(0.114)	(0.116)
Reading proficiency squared x Training			-0.0344	-0.0323
			(0.0377)	(0.0373)
Reading proficiency squared x Coaching			0.00493	0.00316
			(0.0298)	(0.0303)
Teacher experience x Training				0.00124
				(0.0101)
Teacher experience x Coaching				-0.00304
				(0.00338)
Teacher experience squared x Training				-4.87e-05
				(0.000352)
Teacher experience squared x Coaching				-4.56e-05
				(0.000232)
Observations	2,329	2,329	2,329	2,329
R-squared	0.178	0.197	0.212	0.214

Note. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

We also tried to investigate the differential programme effects between urban and rural schools through the case studies and lesson observation study but these did not provide conclusive evidence either. Overall, it seems likely that the combination of educationally disadvantageous factors which

are more prevalent in rural settings may have prevented programme impact. In particular, disruptions to coaching visits may have meant that this part of the causal chain was weakened. However, teacher attendance of Treatment 1 training sessions was high even for rural schools, and similar treatment heterogeneity was observed for this treatment arm. This suggests that a lower number of coaching sessions cannot be the full explanation. It may have been that with more time lost for teaching (e.g. teacher absenteeism and the other factors) it was harder to stick to the learning programme. The question of whether these kind of structured learning programmes can be effective in deep rural settings is ultimately one that we plan on investigating further in future extensions of this research agenda. Unfortunately, the results here do not point to any specific interventions that did work in rural schools, or any aspects that worked. For these rural settings it seems there may be a need to experiment with either a higher dosage of coaching, or with monitoring and other interventions designed to improve basic school functionality and reduce the loss in teaching time.

Intermediate outcomes

Intermediate outcomes for Training and Coaching.

In this section we investigate underlying change mechanisms by measuring how the learning environment, teaching practice, and classroom activities changed as a result of the program. For this purpose we draw from three different data-sources: the teacher questionnaire conducted in the full evaluation sample of 230 schools, the classroom and document inspection conducted in the same sample, and detailed lesson observations conducted in a stratified random sub-set of 60 schools.

A preliminary point is important to make: Table 25 shows that Teachers in the Coaching schools were considerably more likely to report feeling a high level of professional support than those in the control schools, with teachers in the Training group also somewhat more likely to experience high professional support. 82% of teachers in the coaching group felt supported and recognized for their work, compared with 53% of teachers in the control group. Similarly, 84% of teachers in the coaching group reported regularly meeting with people who provide mentoring and curriculum support, compared with 52% of teachers in the control group.[13] These proxies for professional support should to some extent be provided through other teachers in the same school, HODs and subject advisors, but are also a key part of the Theory of Change for the EGRS teacher support interventions, especially the on-site coaching intervention. It is therefore pleasing that as a first step in the process teachers actually felt more support as a result of the interventions.

[13] Note that the 52% of teachers in the control that receive “mentoring and support” could have received it from a subject leader or school principal in the same school, so it does not necessarily indicate external support.

Table 25: Teachers' experiences of professional support

	Control	Training	Coaching	Parents
I feel supported and recognised for my work	53%	62%	82%	49%
I regularly meet with people who provide mentoring and curriculum support	52%	57%	84%	45%

Notes. Data based on 275 teachers surveyed at endline.

Two main results are worth emphasizing. First, even though there is no large difference in access to graded readers, the lesson observations reveal that far more pupils are actually reading graded readers in the programme schools. This increase is substantially larger for teachers who received Coaching relative to teachers who received Training (even though they had received the exact same set of reading resources). Second, even though we find no change in the probability *that* pupils practice reading in the classroom, there is a noticeable difference in *how* they practice reading: Teachers in both Training and Coaching arms are more likely to enact group-guided reading, resulting in more opportunities for pupils to receive individual attention. The impact is, again, larger for teachers who received Coaching relative to Training. These results suggest that there are some teaching practices such as group-guided reading that are difficult to enact and require additional coaching to be effective. They also reveal an important interaction between resources and teaching practice: graded readers are only useful if teachers have developed the skills to use them effectively in the classroom.

We grouped the potential intermediate outcomes into six broad categories: (i) access to reading material in the classroom; (ii) adherence to the teaching routines as prescribed in the curriculum; (iii) curriculum coverage (or the extent of literacy activities conducted) (iv) breadth of reading opportunities in the classroom; (v) teacher-learner interactions related to group-guided reading; and (vi) learners' use of reading material in the classroom. The first two categories – access to reading material and adherence to the teaching routines – provide an indication of at least superficial fidelity to the programme. The third category tests if curriculum coverage has improved because of following the lesson plans. The subsequent two categories look at actual teaching activity in the classroom and tests for the enactment of different components of group-guided reading, an integral yet technically difficult activity prescribed in official curriculum documents. The final category captures what is arguably one of the most important requirements for learning to read: opportunities for pupils to individually practice reading text. For each category we construct a mean index out of the constituent indicators, using the method as specified by Kling and Liebman (2004).

The regression results are reported in Tables 26 to 30. In all specifications we include stratification fixed effects and cluster our standard errors at the school level, where necessary.[14] Many of the variables in this section are ordinal variables that were answered on a 4 or 5-point Likert scale. For ease of interpretation we convert these variables into binary variables when we report the results. In

[14] It is not necessary to cluster our standard errors with the subset of 60 teachers where we did lesson observations, because we only observed one teacher per school. The data is therefore already at the school level.

all cases results on statistical significance are essentially the same when running an ordered logistic model on the original variables. The mean index we report is always constructed using the comprehensive ordinal variables, so no information is lost in the mean index.

Access to print and adherence to teaching routine

Row (1) in Table 26 shows that there was a large and statistically significant improvement in overall access to reading material in the classroom: a 0.465 and 0.41 standard deviation increase for the Kling index in the Training and Coaching arms respectively. Rows (2) to (5) show results for indicators that constitute the mean index. There is a substantial increase in the probability that a classroom contains a well-stocked reading corner (a 25 and 26 percentage point increase), and exhibits a sufficient number of quality Setswana posters (25 and 21 percentage point increase) and flash cards on the classroom wall (an 18 and 17 percentage point increase). The magnitude of the impact is remarkably similar for both treatments. It is important to note, though, there is no impact on the probability that every pupil in the classroom has access to graded readers. This may reflect the complexity for a fieldworker who is not an educational expert to identify a *graded* reader (which refers to a set of booklets progressing incrementally in terms of difficulty) as opposed to any other reading books which may exist in the classroom. The 60-school classroom observation study in fact did reveal significant differences in access to graded readers between control and intervention schools.

Table 26. Access to print and adherence to teaching routine

	VARIABLES	Control	Training		Coaching			p-value Training = Coaches
		mean	Coef.	Std. Error	Coef.	Std. Error	Obs	
	Access to reading material							
(1)	*Kling index*	0	0.465***	(0.120)	0.410***	(0.114)	264	0.651
(2)	All have graded readers	0.416	0.114	(0.0921)	0.0327	(0.0904)	263	0.449
(3)	Reading corners	0.486	0.252***	(0.0854)	0.260***	(0.0806)	253	0.930
(4)	Setswana posters	0.316	0.249***	(0.0821)	0.206**	(0.0865)	263	0.651
(5)	Flash cards	0.752	0.177***	(0.0564)	0.166***	(0.0592)	263	0.828
	Routine							
(6)	*Kling index*	0	0.300***	(0.0811)	0.497***	(0.0652)	276	0.0209
(7)	Group-guided reading	0.241	0.124*	(0.0738)	0.197***	(0.0674)	274	0.363
(8)	Spelling test	0.696	0.155**	(0.0627)	0.238***	(0.0509)	273	0.143
(8)	Phonics	0.491	-0.0708	(0.0745)	0.171**	(0.0720)	274	0.00195
(9)	Shared reading	0.422	0.183**	(0.0728)	0.171**	(0.0711)	274	0.872
(10)	Creative writing	0.310	0.301***	(0.0715)	0.383***	(0.0681)	274	0.286

Notes. Each row represents a separate regression, including stratification fixed effects. The first column indicates the outcome variable. Data is at the teacher level. Standard errors are clustered at the school level *** p<0.01, ** p<0.05, * p<0.1

Next we test for evidence that teachers are more likely to follow the routine specified in the scripted lesson plans. In the teacher survey, we asked them to report how frequently they conduct various

types of teaching activities: group-guided reading, spelling tests, phonics, shared reading, and creative writing.[15] Recall that the frequencies of doing these activities are stipulated in the official curriculum, so in principle the teachers in the control should be performing them at the same frequency. We find that for all of these activities teachers in both Training and Coaching schools are more likely to perform the activity at the *appropriate* frequency. It is important to note that the treated teachers are not stating that they are more likely to perform all activities. They are more likely to perform activities that are required to be performed on a daily basis – group-guided reading and phonics – but state they are less likely than the control group to perform the activity that should only take place on a weekly basis – correcting spelling. These results can therefore not be attributed to pure experimenter demand effect of over-reporting all teaching activities.

Group-guided reading

We have learnt that teachers who received the scripted lesson plans appear more likely to follow the right routine, and as a result are more likely to teach phonics and facilitate group-guided reading in the classroom. Next we unpack the type of teaching activities related to group-guided reading. Recall that there are three important components of group-guided reading: individual opportunities to read out loud, individual assessment, and sorting reading groups by ability. We asked about each one of these indicators separately in the teacher questionnaire.

Table 27. Group-guided reading, individual attention, assessment, and sorting by ability

	VARIABLES	Control	Training		Coaching			P-value
		mean	Coef.	Std. error	Coef.	Std. error	Obs	Training = Coaching
	From teacher questionnaire							
(1)	Kling index	0	0.210**	(0.0880)	0.415***	(0.0772)	276	0.0124
(2)	Teacher can provide list of groups	0.430	0.168*	(0.0987)	0.344***	(0.0815)	232	0.0748
(3)	Listens to each pupil read out loud	0.578	0.0324	(0.0772)	0.237***	(0.0638)	273	0.00714
(4)	One-on-one reading assessment	0.655	0.0877	(0.0755)	0.161**	(0.0638)	274	0.296
(5)	Sort groups by ability	0.718	0.107*	(0.0579)	0.144**	(0.0580)	261	0.527
	From lesson observations							
(6)	Kling index	0	0.722***	(0.237)	0.760***	(0.213)	60	0.863
(7)	Pupils split into groups	0.211	0.365**	(0.169)	0.555***	(0.160)	52	0.252
(8)	Pupils read aloud in groups	0.444	0.140	(0.194)	0.410**	(0.158)	54	`
(9)	Pupils read individually to teacher	0.176	0.334*	(0.186)	0.515***	(0.183)	51	0.317
(10)	Individual reading assessment	0.158	0.295*	(0.170)	0.125	(0.177)	55	0.340
(11)	Individual phonics assessment	0.0556	0.175	(0.143)	0.0622	(0.118)	56	0.487
(12)	Reading groups, different texts	0.105	0.0919	(0.133)	0.247	(0.161)	52	0.415

Notes. Each row represents a separate regression, including stratification fixed effects. The first column indicates the outcome variable. Data is at the teacher level. Standard errors are clustered at the school level *** p<0.01, ** p<0.05, * p<0.1

[15] Options were: Less than once a week, once a week, 2-4 times a week, every day, twice a day

Rows (1) to (5) in Table 27 show results from the teacher questionnaire, which was administered in all 230 schools. There was an overall increase for both treatment arms in the activities that relate to group-guided reading, with a consistently larger impact for Coaching relative to Training. First, as a confirmation of the self-reported increase in conducting group-guided reading, we find that teachers in the Coaching arm are 34.4 and 17.6 percentage points more likely to be able provide a list of the reading groups relative to the Control and Training respectively. We further find that teachers who received Coaching are more likely to state that they listen almost daily to pupils reading out loud (23.7 and 20.4 percentage point increase compared to Control and Training respectively); more likely to perform one-on-one reading assessment at least weekly (16.1 and 7.3 percentage point increase compared to Control and Training respectively); and more likely to state that they sort groups by ability (14.4 percentage point increase relative to Control). The fact that most of these activities are more likely to take place with teachers who received Coaching vs Training suggests that group-guided reading is a pedagogical skill that requires the additional monitoring and feedback provided from coaches to develop. This is also suggestive evidence that these activities related to group-guided reading are at least part of the explanation for faster acquisition of reading proficiency in the Coaching arm relative to Training.

The results from the teacher survey provide evidence that group-guided reading was more likely to take place in both treatment arms, with the largest increase observed for teachers who received Coaching. Moreover, the larger change seems to come from individual attention, rather than sorting by ability. However, these results are all self-reported. To test if these practices actually changed in the classroom we next turn to results from the lesson observations.

Rows (6) to (12) in Table 17 shows that the results from the teacher survey on group-guided reading are broadly supported by the lesson observations: there is a large, statistically significant increase in the mean index of 0.72 and 0.76 standard deviations in the Training and Coaching groups respectively. When examining the individual indicators that constitute the mean index, we see that there is a large increase in the Coaching arm in the probability that the pupils are split into groups (55.5 percentage point increase), that pupils read aloud in groups (41 percentage point increase), and that the pupils read individually to the teacher (51.5 percentage point increase). The impact for these three indicators is smaller for the Training arm, and not always statistically significant. However, we do not find strong evidence for any improvement in the probability of providing individual assessment and grouping by ability. There is a small increase in the probability of providing individual assessment, which is statistically significant only in the Training arm. Teachers that received Coaching are 24.7 percentage points more likely to have different reading groups assigned to different graded readers (compared to a 9.2 percentage point increase for teachers that received Training), but the difference is not statistically significant.

Taken together we see strong evidence that there was an increase in group-guided reading in both treatment arms, with the largest change observed for teachers that received the Coaching. This coincided with more individual attention by the teacher and opportunities to read out loud in groups, but there is weaker evidence for any change in individual assessment and sorting by ability.

Frequency of opportunities to read

Next, we look at the frequency of reading opportunities in the classroom. The fieldworkers were asked to record how many pupils in the classroom are involved with reading letters, words, sentences, or extended texts. The answers were recorded on a 5-point Likert scale, ranging from no pupils to all of the pupils. Results are reported in rows (1) to (9) in Table 28. There is only weak evidence that more pupils in the classroom are practicing different reading activities. Pupils in the Training and Coaching arms are more likely to read extended texts, but the mean index is not significantly different across intervention groups. However, it is important to note that these records do not indicate how the pupils are practicing reading. The pupils might be practicing reading through whole-class chorusing, and might not actually have been provided with individual opportunities to read.

Table 28: Opportunities to read

	VARIABLES	Control	Training		Coaching		Obs	P-value Training = Coaching
		mean	Coef.	Std. error	Coef.	Std. error		
	Reading frequency (lesson observations)							
(1)	Kling index	0	0.0767	(0.149)	0.148	(0.150)	60	0.300
(2)	Letters	0.625	-0.126	(0.185)	0.105	(0.174)	49	0.231
(3)	1-2 words	0.471	-0.0408	(0.176)	0.229	(0.227)	44	0.378
(4)	3-10 words	0.667	-0.0582	(0.148)	0.0905	(0.129)	52	0.425
(5)	10+ words	0.133	0.0772	(0.151)	0.111	(0.170)	40	0.406
(6)	1-2 sentences	0.529	-0.269	(0.201)	-0.115	(0.214)	44	0.268
(7)	3-5 sentences	0.333	0.389**	(0.178)	0.441***	(0.161)	48	0.360
(8)	5+ sentences	0.188	0.352**	(0.173)	0.363**	(0.177)	49	0.330
(9)	Extended texts	0.579	0.0262	(0.181)	0.148	(0.182)	55	0.237

Notes. Each row represents a separate regression, including stratification fixed effects. The first column indicates the outcome variable. Data is at the teacher level. Standard errors are clustered at the school level *** p<0.01, ** p<0.05, * p<0.1

Curriculum coverage, assessment and opportunities to write

Table 29 looks at curriculum coverage and teacher assessment of written work. Fieldworkers were required to count the number of days that writing exercises were completed in the exercise book, and the number of pages completed in the government workbook.[16] To minimize risk of bias due to strategic selection of exercise and workbooks, the teacher was asked to provide books of one of the most proficient pupils in his/her class. Table 29 indicates that the amount of written work was higher in both Training and Coaching schools relative to the control group, but there was no statistically

[16] To reduce data capture error, we asked the fieldworker to only count pages completed for three specific days. We chose three days that should have been covered by teachers by the end of the year, regardless of their choice of sequencing.

significant difference between Training and Coaching schools. The frequency of teachers marking learner work was not significantly different across the groups.

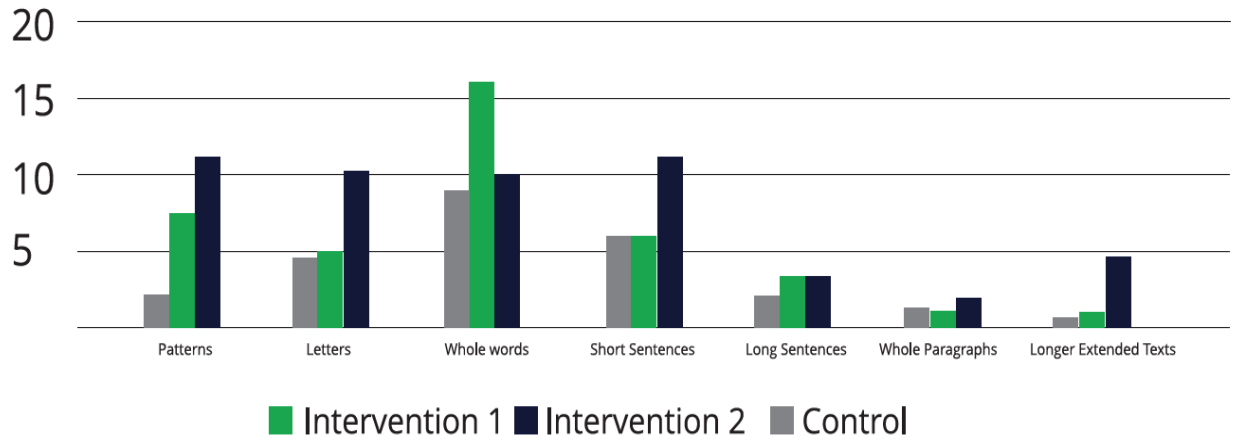
Table 29: Curriculum coverage and assessments

	VARIABLES	Control	Training		Coaching			P-value Training = Coaching
		mean	Coef.	Std. error	Coef.	Std. error	Obs	
	Curriculum coverage							
(1)	Mean index	0	0.469***	(0.128)	0.317**	(0.139)	271	0.343
(2)	Days pupil completed any exercises	23.57	16.64***	(3.348)	5.007	(3.778)	270	0.00679
(3)	Days pupil completed writing exercises	19.08	8.532***	(3.046)	6.306*	(3.478)	270	0.581
(4)	Days pupil completed full sentence writing exercises	14.11	9.736***	(3.155)	5.539*	(3.044)	270	0.264
(5)	Proportion of pages completed	0.761	-0.0441	(0.0555)	0.0840**	(0.0423)	258	0.0185
	Assessment							
(6)	Learner has marked book	0.939	0.0197	(0.0336)	0.0197	(0.0308)	267	0.999
(7)	All exercises marked	0.400	0.0201	(0.0851)	0.0182	(0.0781)	256	0.984

Notes. Each row represents a separate regression, including stratification fixed effects. The first column indicates the outcome variable. Data is at the teacher level. Standard errors are clustered at the school level *** p<0.01, ** p<0.05, * p<0.1

The 60-school lesson observation data revealed clearer differences across intervention groups with respect to opportunities to write. Figure 30 shows that for most categories of written work there were more writing exercises completed in the exercise books of children in Training and Coaching schools compared to the control group. The more advanced skill of writing extended text was virtually non-existent in control and Training schools, whereas an average of nearly 5 pieces of extended writing was observed in the books of children in the Coaching group.

Figure 30: Average number of writing exercises per writing type



Source: 60-school lesson observation data

Pupil use of reading material

As a final measure, we also test if pupils have more individual opportunities to handle and read books. During the 60-school lesson observation study fieldworkers were required to count how many pupils actually handled books (excluding the government workbooks) and how many pupils read graded reading books during the lesson. Even though there was no difference in access to graded readers between treatment arms, we see a substantial increase in use of reading material, especially in the number of children who have opportunities to read. These results are reported in Table 30. Strikingly, in the control schools only one pupil in one school actually read from a book, leading to an average of 0.05 pupils reading a book in the control. The average number of pupils who read increased by 2.3 and 5.1 in the Training and Coaching arms, respectively. There is also a marked difference between the treatment arms: far more pupils in the Coaching arm handle and read books.

Table 30: Opportunities to handle and read books in class

	VARIABLES	Control	Training		Coaching			P-value
		mean	Coef.	Std. error	Coef.	Std. error	Obs	Training = Coaching
	Use of reading material							
(12)	Kling index	0	4.859*	(2.551)	12.15***	(2.532)	60	0.004
(13)	No. learners handle books	1	0.717	(0.988)	2.542**	(1.001)	59	0.0145
(14)	No. learners read readers	0.0526	2.329**	(1.098)	5.093***	(1.067)	57	0.009

Notes. Each row represents a separate regression, including stratification fixed effects. The first column indicates the outcome variable. Data is at the teacher level. Standard errors are clustered at the school level *** p<0.01, ** p<0.05, * p<0.1

These results reveal the important interaction between resources, teaching practice, and use of resources. Access to graded readers is high in all the evaluation arms, including the control. However the purpose of the graded readers is to provide individual opportunities to practice reading. Pupils are provided this opportunity during group-guided reading, an activity that teachers find challenging to implement. These resources therefore cannot be used without appropriate enactment of a new teaching method. As a result very few pupils are actually reading graded readers in the control schools.

More detail on the intermediate outcomes observed in Training and Coaching groups is available in the separate report on the 60-school classroom observation study.

Intermediate outcomes for parent involvement

The next table investigates the extent to which dimensions of parental involvement may have shifted in response to the intervention, and does so using a set of Ordinary Least Squares (OLS) regressions. Each regression predicts an intermediate outcome that could have shifted. The key explanatory variable is being in Intervention 3 (relative to being in the control group of schools), although a set of control variables for baseline learner scores, learner age and gender, school and community poverty are included (though not reported on in the table). The intermediate outcomes are a parent's frequency of reading to the child, the frequency of checking homework, the frequency of playing games with the child, the number of parent meetings at the school that the parent attended, whether the parent feels that they are primarily responsible for their child's education (as opposed to the school or the government), the frequency of learner absence from school, how often the parent checks their child's school bag, the typical bed time of the child and whether the child sometimes stay up beyond 9pm to watch television.

Table 31 reveals that only the number of parent meetings attended was significantly higher in the Intervention group. This is a somewhat mechanical outcome since regular parent meetings were the mechanism through which any change would have occurred. The fact that no other indicators shifted substantially confirms that on average, there was no large change in parental behavior in response to the intervention.

Table 31: Intermediate outcomes for parent involvement

	(1) readtochild	(2) checkHW	(3) games	(4) Meetings	(5) Responsible	(6) absence	(7) Check bag	(8) Bed time	(9) Late tv
Parents	-0.0182	-0.200	0.0191	0.466***	-0.0364	-0.0766	-0.0284	-0.00381	-0.0299
	(0.119)	(0.140)	(0.0793)	(0.116)	(0.0271)	(0.0674)	(0.0757)	(0.0445)	(0.0307)
Constant	4.217***	4.494***	2.842***	1.423	0.217	2.207***	1.288**	2.011***	0.266
	(0.851)	(1.039)	(0.561)	(0.922)	(0.201)	(0.604)	(0.562)	(0.335)	(0.223)
Observations	1,727	1,502	1,806	1,328	2,574	1,783	1,795	1,427	1,158
R-squared	0.028	0.090	0.020	0.049	0.122	0.038	0.067	0.052	0.049

Notes: Each row represents a separate regression on treatment assignment, with the same set of controls as in Table 12. Data is restricted to the Control and Parent arm. Column headings indicate the development variable. Cluster robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Robustness checks

We are not aware of any reasons to expect a systematic upward bias in the estimated treatment effects. Moreover, since attrition is not statistically significantly different across treatment group we do not need to attempt the Lee bounding exercise as proposed in our Pre-Analysis Plan. Therefore, in this section we test the robustness of the main results to two possible issues, both of which might be expected to attenuate the estimated intervention effect sizes, namely the influence of a few multi-grade classrooms in the sample (where applying the grade-specific lesson plans would be tricky), and the possibility of contamination of control group classrooms due to sharing of lesson plans.

The main risk of contamination of the control group in this experiment arises through the possibility of sharing the EGRS lesson plans. We can rule out the possibility that coaching would have taken place at the control schools or that control school teachers would have attended the central training sessions. Similarly, we know that the additional reading books, flash cards and posters were only provided to treatment schools.

Table 32 shows the percentage of teachers in each of the intervention groups who reported borrowing or sharing lesson plans with teachers in other schools. From the table it is clear that the sharing of lesson plans across schools appears to be a relatively common phenomenon with the percentages ranging between 20 and 30%, although these lesson plans may well include lesson plans other than those provided through the EGRS. This means that control schools may well have, at least to some extent, been exposed to the EGRS lesson plans. Whilst this is only one component of the intervention, it appears important to test the sensitivity of our results to the exclusion of potentially contaminated control schools.

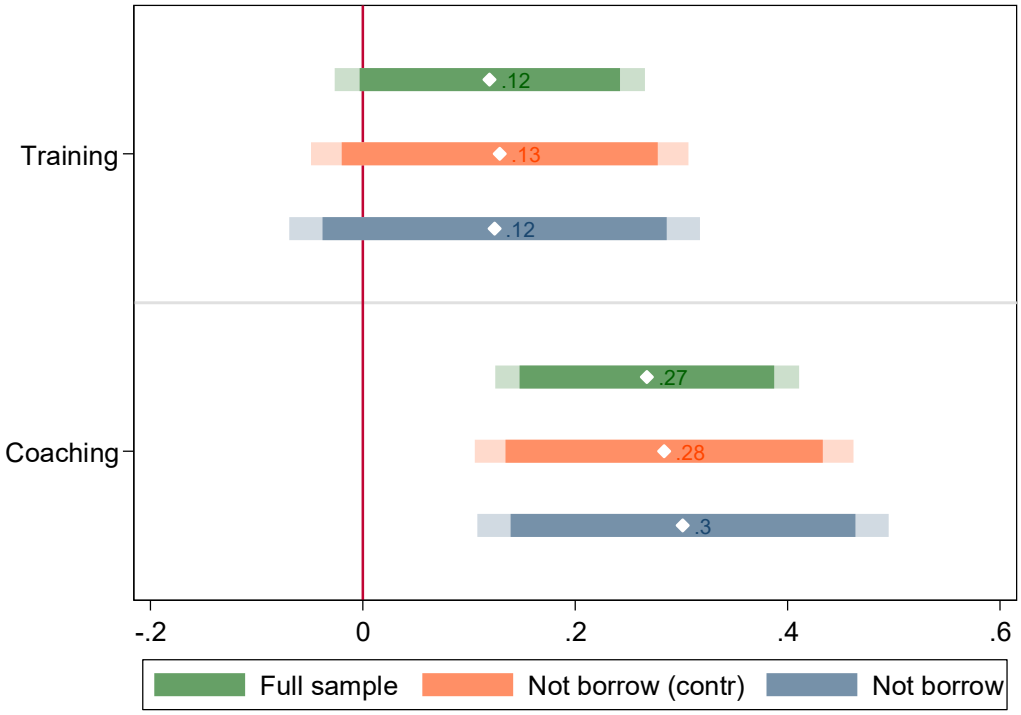
Table 32: Borrowing and sharing of lesson plans across schools

	Receive/borrow lesson plans from other schools	Share lesson plans with other schools
Control	23.28	45.69
Training	19.74	31.58
Coaching	20.73	31.71
Parents	30.38	46.84

Note: The table shows percentages of endline teachers

Figure 31 diagrammatically presents the results of our main regression models but excluding control schools who reported to have borrowed lesson plans from teachers at other schools (orange bars), as well as excluding all borrowers of lesson plans in both control and treatment schools (blue bars). It is clear that the estimated effect sizes hardly change when one excludes borrowers from the analysis. Therefore, there is no evidence that potential contamination of the control group has led to an under-estimation of treatment effects.

Figure 31: Sensitivity of results to exclusion of potentially contaminated control schools



Note: All graphs exclude learners repeating grade 1 since teachers in grade 1 were not interviewed. Orange bars indicate treatment coefficients of model where control schools that borrowed lesson plans are excluded from the sample; blue bars indicate model where all schools that borrowed lesson plans are excluded from the sample.

Cost-effectiveness Analysis

In thinking about which interventions are suitable to scale up, we need to consider both the impacts and costs of the programmes. In this section we outline different ways of doing cost-effectiveness analysis. In all scenarios the Coaching programme is most cost-effective.

For cost estimates we use the budget from the second year of implementation, where more of the implementation/budgeting kinks have been worked out.

A challenge in allocating costs is that one organization was responsible for implementing all three interventions, so some fixed costs (such as quality assurance and management, and material development) were spread across all treatment arms. We made the following judgment calls:

- Exclude fixed costs of material development (e.g. development of lesson plans and training manuals), and reference group meetings with provincial officials. The motivation is that when this program is scaled up - both across schools, and over time- the fixed costs will be nominal.
- Allocate the budget items of (i) program management, (ii) administration and (iii) quality assurance in the following proportions: 30%, 50%, 20% for Training, Coaching and Parents arms respectively. This is based on the best estimates of the service provider. These costs would conceivably increase proportionally if the program is scaled.

Table 33 shows a breakdown of costs by treatment group and budget item. It is clear that salary was the largest cost driver in the Coaching and Parents arm, but accommodation/venue was the largest cost driver in the Training arm. This is because Teachers had to travel to a site to participate in the training.

The total costs of implementation (excluding cost for material development and reference group meetings) for the Training, Coaching and Parents interventions were R1.48M, R2.08M, and R1.1M respectively. Since these programmes were each implemented in 50 schools and the average number of grade one pupils in our sample of schools at the start of the programme was 75, the per-pupil costs are R397 (30.58 USD), R557 (42.91 USD), and R295 (22.75 USD). Given the impacts of 0.12, 0.246 and 0.114 SD increase for the respective programs, we can conclude that the Coaching arm was most cost-effective with a **0.57 SD** increase for $100 spent per pupil per year, compared to **0.39 and 0.50 SD** increase per $100 spent per pupil per year for the Training and Parents arms respectively. There is no large difference between Coaching and Training since Coaching is roughly twice as effective and twice as expensive.

Table 33. Breakdown of costs in the year two, by treatment arm and line item

	Training		Coaching		Parents	
	Sub-total	%	Sub-total	%	Sub-total	%
Materials Provision	272,817	15%	257,396	11%	231,991	20%
Salary	614,340	35%	1,397,900	59%	712,278	61%
Transport	47,940	3%	419,491	18%	40,567	3%
Accommodation and venue	548,563	31%	7,023	0%	80,325	7%
Catering	1,074	0%	1,074	0%	39,711	3%
Sub-total	**1,484,734**	**84%**	**2,082,883**	**88%**	**1,104,872**	**94%**
Material Development	280,987	16%	280,987	12%	68,091	6%
Total	**1,765,721**	**100%**	**2,363,870**	**100%**	**1,172,962**	**100%**

Note: Cost data constructed through discussion between Research Team and Class Act

The above Cost-Effectiveness Analysis looked at the full evaluation sample, but one should arguably use the estimates in the sub-section of schools where the programme can be expected to have the largest impact, namely urban township schools. In these schools the average impacts were 0.45, 0.75, and 0.38 standard deviations in the Training, Coaching and Parents interventions respectively. The

Coaching intervention remains more cost-effective with .146 standard deviations increase per dollar spent per pupil per year.

Finally we consider another metric of performance, beyond standard deviations increase: the increase in the number of pupils who pass the comprehension test. The treatment impact on the probability of passing the comprehension test is 1.59, 9.88, and 1.131 p.points for the Training, Coaching and Parents arms respectively. Now the Coaching is far more cost-effective. A learner is 0.23 percentage points more likely to pass the comprehension test per dollar spent per pupil per year, compared to .052 and .057 percentage points in the Training and Parents arms respectively.

Summary of the 60-school lesson observation study

The main strength of Randomised Control Trials (RCTs) is their internal validity in measuring the causal impact of particular programmes. In other words, if outcomes end up higher in a group that received an intervention, we know that this is because of the intervention and we can make a quantitative estimate of that impact. But in order to gain a deeper understanding of why and how a programme may or may not have achieved its desired outcomes, one needs to complement the quantitative estimates of causal impact with mixed methods research. To better understand which mechanisms were affecting the change in Interventions 1 and 2, a classroom observation study was commissioned.

The study was conducted in 60 of the schools that participated in the EGRS. A stratified random sample of 20 schools from each of the Control, Intervention 1 and Intervention 2 groups was chosen to form part of the study. In each of the schools, three different types of evidence were collected: (1) lesson observations; (2) evidence of work done in learners' workbooks and exercise books, as well as the review of various teaching documents and; (3) information from the teacher based on an interview.

Comparing the three different groups of schools, it emerged that the intervention schools were performing notably better than the control schools in the following themes: 'Teaching and Learning Environment'; 'Planning and Curriculum Coverage' and 'Classroom Management'. The main differences in the 'Teaching and Learning Environment' were the increased availability of display material (for example flashcards), a classroom arrangement that is more conducive to reading, and increased availability of reading books in the intervention classrooms.

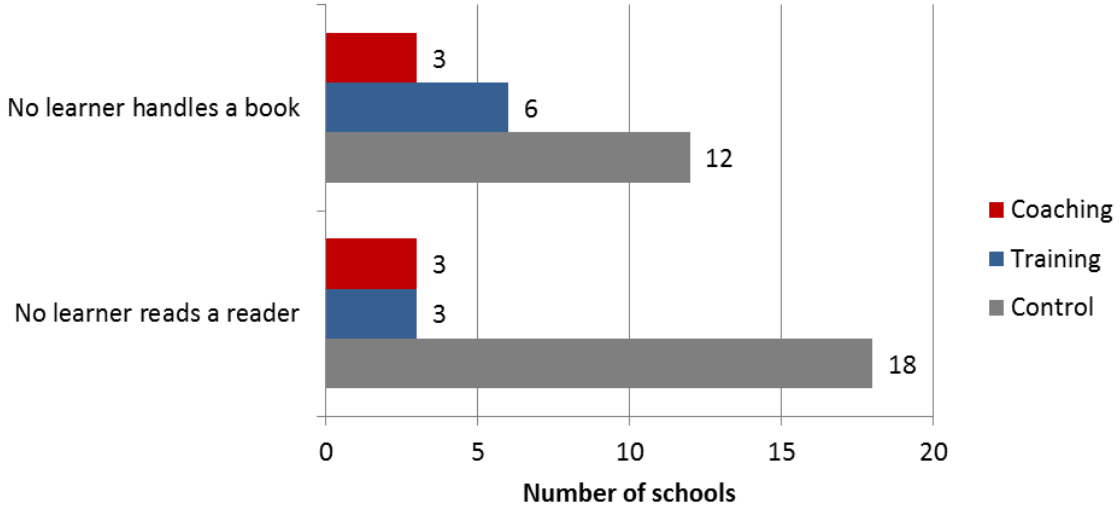
The scripted lesson plans provided through the programme proved to be hugely beneficial in translating the curriculum into daily lessons with detailed activities, which in turn improved 'Planning and Curriculum Coverage'. The specificity of the EGRS lesson plans was visibly different from the lesson plans used by the Control group's teachers and included important aspects such as vocabulary development. The benefit of greater specificity is especially clear with regards to vocabulary development, where teachers in Intervention 1 and 2 schools were much more likely to engage the learners in vocabulary development during the observed lesson. The EGRS lesson plans also provided teachers with a more accurate understanding of the size and

scope of the curriculum that needs to be covered across the year, and provided them with a mechanism for tracking their own progress. The teachers in Intervention 2, however, were more likely to actually track their own progress and to be up to date in covering the curriculum. Evidence of increased curriculum coverage in Intervention 2 schools was found in the lessons observed, as well as in the learners' workbooks. The increased curriculum coverage meant that learners were more often engaged in writing activities and therefore learners in the intervention schools were less often observed being uninvolved in class. Although teachers in the intervention schools were observed to have a more realistic understanding of the curriculum scope, they still did not necessarily have a sufficient understanding of the cognitive demand required by the curriculum.

With regards to classroom management it was found that in 90% of the Intervention 2 classrooms no time was lost due to learners not being involved, whereas this was the case in 75% of the Intervention 1 classrooms. The evidence of more writing exercises in the learners' workbooks in the Intervention 1 and 2 schools corroborates the finding of learners being more involved and suggests that the improved classroom management is leading to increased curriculum coverage. In only 55% of the Control classrooms was no time lost due to learners not being involved.

In relation to the themes 'Opportunities to Write' and 'Use of Learning and Teaching Support Material' there were notable differences between Intervention 1 and Intervention 2 schools. Differences in the 'Use of Learning and Teaching Material' can be largely attributed to the prevalence of learners using storybooks and readers in class, as well as to the use of resources such as flashcards and charts by teachers during lesson observations in the Intervention 2 schools. In 90% of the Control schools not a single learner was observed reading a graded reader, whereas this was commonly observed in the intervention classrooms (see Figure 32). These findings suggest that the EGRS interventions have been successful, not only in providing classrooms with the necessary readers, but also in ensuring that teachers make effective use of these resources.

Figure 32: Use of Learning and Teaching Support Material during lessons



Data Source: 60-school classroom observation study

With regards to ‘Opportunities to Write’, learners in Intervention 2 schools completed more writing exercises on average, specifically exercises pertaining to writing letters, short sentences and extended texts. Learners in Intervention 2 classrooms were also engaged in a wider variety of writing exercises overall and were more likely to have their personally created dictionaries and to do more cursive writing exercises than learners in Intervention 1 classrooms.

Group guided reading provides a valuable opportunity for individualised and small group attention and was observed to occur more often in intervention classrooms. From evidence in the lesson observations it appears that in the majority of Control classrooms, the teachers’ in-class reading and phonics assessment was based on the class as a whole, rather than on individual learner proficiency.

The evidence found through the Lesson Observation Study suggests that the reading coaches played a critical role with regard to two aspects: (1) providing teachers with a more in-depth understanding of the enactment of the methodologies they were taught during the training; and perhaps more importantly, (2) supporting and motivating teachers to persist with the implementation of the programme. As mentioned above, there is significant evidence that Intervention 2 teachers were implementing the lesson plans as intended. Intervention 2 teachers were also more frequently seen providing different levels of readers to different ability groups in the lessons observed; doing a wider variety of writing activities during the Home Language lessons; covering the required pages in the DBE workbooks and covering more challenging aspects of the grade 2 writing curriculum, especially writing sentences and extended texts.

Although Intervention 1 brought about significant changes in teachers’ instructional practices, it seems that the reading coach component of Intervention 2 was the essential ingredient to encourage persistence in the curriculum-aligned learning programme. Available evidence therefore suggests that the ‘triple cocktail’ of lesson plans, high quality materials and coaching is necessary to affect real change in teachers’ instructional practices.

Summary of the case studies

A set of case studies was undertaken in four schools – two Training and two Coaching schools. Each case study involved lesson observations, teacher interviews and document reviews. A full report is available on these case studies.

A number of successful areas of the EGRS programmes were highlighted. Firstly, teachers were making daily use of the EGRS scripted lesson plans and regular use of the EGRS curriculum coverage trackers. Secondly, regular phonics, handwriting, group guided reading instruction and individual seatwork (writing) was taking place in EGRS classrooms. Thirdly, the provisioning of writing activities in the EGRS lesson plans was playing a role in motivating teachers to give classes more writing tasks, and learners were completing written work on most school days.

The case studies also identified several factors inhibiting programme impact. Particularly large classes made it difficult for teachers to provide learners with the individual attention they required. Secondly, there appeared to be an absence of a culture of reading for enjoyment and limited exposure of grade 2 learners to books besides the graded readers provided through EGRS and the DBE workbooks. Thirdly, teachers displayed a 'restricted' understanding of what it means to teach children to read independently – there was still an over-reliance on teacher-directed strategies (e.g. telling learners what words were).

A second set of case studies was conducted by Dr Kerryn Dixon and Prof Brahm Fleisch in an additional four schools. These were all Coaching schools, selected at the extreme ends of the improvement spectrum based on the average performance on Wave 2 data. As in Dr Reeves' case studies, Dr Dixon and Prof Fleisch observed lessons, interviewed teachers, principals and other school staff, and reviewed classroom documents. A summary report is available on these case studies.

This report focuses on the complexities and nuances associated with the teachers' engagement with the various components and methods of the Coaching intervention. Although teachers lacked the vocabulary to talk about the five components of reading contained in the lesson plans, i.e. Phonological Awareness, Phonics, Vocabulary, Fluency and Comprehension (and writing), the strength of the lesson plans is that they incorporate all of these components in a set of standardised lessons, with simple, systematic routines. The lesson plans impacted both macro (across the academic year) and micro (within each lesson) pacing. Teachers singled out the positive types of learning that occurred during the coaching process, and signaled that a unique and helpful emotional environment was created by the coach. It was also found that the new learning materials substantially contributed to improved instruction. The comprehensive set of 'word' flashcards were used extensively. Their popularity may be linked to teachers' familiarity with the 'look and say method' for teaching sight/high frequency words. The Vula Bula books were received very favourably by teachers and were observed in use. Teachers specifically noted that the books were pitched at the correct level and were appropriately sequenced. A number of weaknesses were also observed. The phonics programme was not well understood by teachers. Group-guided reading, a key method for teaching reading was also not properly understood and was inadequately practiced. Whilst group-guided reading was essentially non-existent in Control schools (as evident in the 60-school lesson observation study), this indicates that even in the Coaching intervention there is a long way to go before reaching high quality instructional practice.

8. Discussion

This evaluation has compared the effectiveness of two structured pedagogic programmes (in which only the modality of teacher support varied) and a parent involvement programme, all of which aimed to improve home language reading acquisition. The first approach to teacher support (Training) followed the traditional model of a once-off training conducted at a central venue. In the second approach (Coaching), teachers were visited on a monthly basis by a specialist reading

coach who observed their teaching, provided feedback, and modelled correct teaching practices. We find that Coaching had a large and statistically significant impact on learner reading proficiency, more than twice the size of the Training arm. Coaching was also more cost-effective. We also find that teachers in both treatments are more likely to practice a difficult teaching technique called group-guided reading, although this impact was far larger for teachers who received Coaching. We also observe substantial heterogeneity in treatment size: the impact for both programmes is far larger in urban and relatively more affluent areas. Coaching is also far more effective in large classes.

Although there may have been some impact of the parental involvement programme on phonological awareness (one key component in learning to read), the overall effect on reading outcomes was indistinguishable from zero. A key obstacle to programme impact was parent attendance, with nearly 50% of children's parents not even attending once per year, about 20% of parents never attending throughout the two years, and only about a third of parents attending at least three sessions per year. Although parent involvement is clearly an important factor in a child's literacy development, it has proven difficult to shift parent involvement in a substantial way.

A plethora of recent meta-analyses and systematic reviews allows us to benchmark the results of this study. Kraft et al (2018) conducted a meta-analysis of 44 Coaching programs in the United States and found a pooled effect size of .11 SD of on academic achievement, for large-scale effectiveness studies with 100 teachers or more. Conn (2017) examined studies in sub-Saharan Africa and found that the average impact of pedagogical interventions (that were credibly evaluated) was 0.228 standard deviations. McEwan (2012) found a mean effect of teacher professional development programs of 0.12 standard deviations. Another review by Snilstveit et al (2016) concluded that structured pedagogical programs are some of the most effective: They averaged over 21 studies and found an average impact of 0.23 standard deviations. Taken together, our estimated effect size of 0.232 standard deviations for Coaching is in line, and perhaps slightly larger than, similar interventions implemented in developing countries.

Seen in the context of other evaluations of similar programmes, we feel it is likely that these results are relevant to – and in line with findings from – other contexts, at least within sub-Saharan Africa. Other studies in sub-Saharan Africa have found that the combination of reading coaches and supporting learning material can improve pupils' proficiency in early-grade reading (Piper, Zuilkowski & Mugenda, 2014,, Piper & Korda, 2011, Lucas, 2014, Kerwin, 2018). Moreover, a previous quasi-experimental evaluation of a similar coaching program in a different province in South Africa also found positive impacts (Fleisch & Schöer, 2016). By comparing the impacts of identical learning programmes, but with different forms of teacher support, this study sheds light on which components are uniquely responsible for the success of these sorts of programmes, which are often bundled.

Discussions around external validity inevitably involve some degree of speculation, and ultimately one has to be cautious in extrapolating findings from one study site to another. Firstly, we acknowledge the possibility that there may have been some element of a Hawthorne effect amongst teachers receiving training and coaching support, who knew that they were part of a

study. It is possible that the same interventions may have been less effective if they were not implemented as part of a study. We also acknowledge that our sampling frame excluded more affluent schools, very small schools, very large schools, and those not using Setswana as LOLT in the Foundation Phase. However, these exclusions are of minority groups of schools in the North West Province and are of those groups of schools that are not the policy focus for an intervention such as this. Grade-specific lesson plans are in any case not appropriate for multi-grade teaching, which is what is usually practiced in very small schools. Although particularly large schools are indeed relevant for the interventions being tested, we do observe that within our sample the treatment effect for training and for coaching increases with school size. Therefore, it seems fair to extrapolate that the treatment effects in the schools excluded due to being large would be at least as large as the average treatment effects reported in our study and probably would be larger.

We should also note that we do not yet have a policy warrant for extending any of these programmes in deep rural settings, but that more likelihood of success can be expected in urban township areas, of which there are many throughout South Africa.

Finally, the work of EGRS 2 seeks to confirm the effectiveness of the coaching model in another province and in another subject area, namely English as a First Additional Language. Lastly, when reading these results together with other studies done in Kenya and elsewhere, a case for structured learning programmes with on-site coaching is building with strong external validity.

At the same time, this study also points to the importance of context in moderating impacts. We find that the programme is dramatically more effective in urban schools relative to rural schools, and the impact in rural schools was indistinguishable from zero. This is even more striking since implementation was tightly controlled – there were only three reading coaches, and each went to both urban and rural schools – and the study setting was relatively homogeneous: the study was set in one province, all the sampled schools had the same language of instruction (Setswana), and are drawn from the bottom 70% percent of schools in terms of the official school poverty classification. This large heterogeneity in impacts underscores the importance of understanding why a programme was successful and under which conditions it is likely to succeed in another context.

As government or donors consider how to expand coaching programmes into more schools, there are various constraints to consider. Even though the costs of the intervention are low relative to overall government expenditure per pupil (7-8%), this translates into a high proportion of per pupil non-personnel spending (70-80%), since nearly 90% of government spending goes to salaries. Another constraint to scale-up would be finding enough suitable reading coaches. One option is to make use of the district-level government officials, known as “subject advisors”. se subject advisors are already supposed to visit schools and provide pedagogical support. They might, in principle, perform the role of coaches. However, the ratio of schools to subject advisors is too high for them to visit each school more than once or twice in a year. In some districts, there are several hundred schools allocated to one subject advisor. Moreover, the recruitment process, oversight

structure and full set of responsibilities of subject advisors means that we do not think the same coaching function can be fulfilled through this structure.

But it is worth noting that scaling a programme does not mean that it needs to be implemented in all schools at the same time. It is unlikely that the program can be simultaneously implemented in all poor schools in the country without substantially reducing the quality. Therefore, we believe that the coaching programme should be implemented in a prioritized sub-set of schools (say 20% – 30% of schools per province), for two or three years at a time, and that the recruitment and oversight of coaches should be managed by literacy NGOs or an independent institution so as to protect the role that they play and the relationship of trust that is created between teacher and coach. The next section presents a concrete proposal for taking the successful programme to the next level of scale in the districts in which the EGRS was conducted.

Overall, the evidence from this program provides promising evidence of a successful set of interventions to improve early-grade reading in South Africa, but more research is needed to understand the type of program that can work in the most difficult and poorest settings.

Aside from the specific findings emerging from this study, it is worth also noting the significance of this project as a prototype government-led evaluation, at least in the South African education sector, as well as to note some of the lessons learned through such a process. Since this project was the first of its kind in South Africa there was no existing budget for its implementation nor was there a section within the Department of Basic Education responsible for it. Essentially, it was championed by specific individuals within the department and over time the work was incorporated into the operational plan of the section responsible for research, monitoring and evaluation. As a result fund-raising took a lot of effort and the project ultimately required contributions from a number of local and international donors, something which resulted in a fairly heavy reporting burden. In spite of this, the partnerships with donors, the University of the Witwatersrand, the HSRC and external academics on the research team, were a valuable aspect of the project leading to progress in the way government and other stakeholders work to advance the goals of the education sector.

The work of project management, analysis, instrument development, stakeholder engagement, ensuring that findings begin to influence government planning, etc, all fell to a small team of individuals in the DBE, with some help from key individuals outside. Certain bottlenecks seemed unavoidable, since those working in the DBE had access to data, networks, etc. This was demanding on time, but it was also a great strength of the project since the links between the various aspects of the work may well have broken otherwise.

The second Early Grade Reading Study, which is taking place in Mpumalanga (2017-2019), has built on many of the lessons learned during the first EGRS. For instance, the baseline data collection was not a smooth process and many lessons were learned resulting in much more detailed Terms of Reference for service providers conducting subsequent data collections. It was decided that a minimum of three days was needed for fieldworker training. Lessons around

instrument development were also learned. For example, certain components of reading may be really important but are hard to measure on a large scale (e.g. phonological awareness). There are many lessons still to be learned about running impact evaluations through government, especially when it comes to responding to the findings. This is the phase we are currently in and is described in the next section.

9. Engagement and promoting buy-in and demand

The Policy Influence Plan (PIP) contains extensive details about the various stakeholder engagement activities that have been ongoing throughout the EGRS. Currently, a lot of dissemination work has been taking place, including local and international conferences, presentations to various government forums and to our South African donors for the project – the ZENEX Foundation as well as to UNICEF. The recent highlight was the launch event of the Year 2 evaluation findings held on the 16th and 17th of August 2017 at the Department of Basic Education. This was well attended by departmental officials, provincial education departmental officials, researchers, literacy NGOs, and those in other government departments. An international speaker, Dr Benjamin Piper, also attended the event and shared experiences from similar early grade reading work being done in Kenya.

One key process that will be followed in the months to come is to follow the procedures associated with the National Evaluation System (NES). The EGRS has been registered on the National Evaluation Plan (NEP), which is administered by the DPME. All evaluations on the NEP require that the custodian department (in this case the DBE) provide a formal management response to the results of the evaluation and that the department develops an improvement plan. These responses together with the evaluation results and recommendations are then presented to the national Cabinet Ministers for approval, meaning that follow-up decisions will carry the highest level of political backing. The DBE will then be monitored by the DPME in terms of implementing the improvement plan.

Although we recommend that all provincial education departments and partners consider responding to the various recommendations above, it is most strategic to build on the momentum in the districts of Ngaka Modiri Molema and Dr Kenneth Kaunda by scaling up the implementation of the structured learning programme, materials and coaching in these districts. As a next step, therefore, the DBE anticipates conducting further work in the North West province in two ways:

1. Delivering the materials for grade 1 to 3 (lesson plans, Vula Bula graded readers, posters and flashcards) to all Quintile 1-3 schools in the two districts (Ngaka Modiri Molema & Dr Kenneth Kaunda) including the 230 schools that have been part of the EGRS - total estimated number of schools is 450. All subject advisors should also receive training in order to understand the materials and support the implementation in schools.

2. Providing the additional on-site coaching component to a total of 180 schools; specifically the 80 control group schools, the 50 intervention I schools and the 50 intervention II schools. This is intended to firstly provide the most successful programme to the control group schools that have participated, and secondly to continue to build on the momentum gained in the 100 intervention schools.

The coaching would take place across the entire Foundation Phase for a minimum of 2 years.

The idea of scaling the programme up in this manner is to bring it in line with what full scale implementation would ultimately look like, at least within these two districts. Given potential constraints in recruiting, managing and financing large numbers of reading coaches, it is unlikely that this model could be implemented in all primary schools. However, it should be implemented at a level of scale which is large enough to make a significant difference if well targeted. Moreover, having some schools receiving the on-site coaching at the same time as district-wide support being provided to all schools, could be a catalyst to shifting the norms of instructional practice at scale.

Several factors inform the choice of 180 schools for additional on-site coaching. Firstly, this would represent a significant step up in terms of scale – from 50 schools (3 coaches) to 180 schools (probably between 12 and 15 coaches). Secondly, there is an ethical obligation to work in the 80 schools that served as a control group in the EGRS. Thirdly, it makes sense to build on the momentum that exists in the 100 schools that received the structured learning programmes under EGRS. Fourthly, by working in this group of 180 schools the 50 schools that received the less successful parent involvement programme can serve as a future control group for evaluation purposes.

10. Specific findings for policy and practice

A number of specific findings and recommendations can be made based on the results of this impact evaluation:

Recommendations and DBE Plans for moving forward

1. **The DBE should take steps to provide Foundation Phase teachers with a curriculum-aligned structured learning programme using lesson plans and integrated reading support materials.**
    a. Although the learning programme implemented in the EGRS was clearly more effective when supported by on-site coaching, the fact that intervention 1 had some impact suggests that the learning programme itself (lesson plans and integrated materials) can offer benefits to curriculum delivery.
    b. In order for lesson plans and additional integrated reading materials to be provided to all language groups and into the future, the DBE should set up processes to ensure that lesson plans and materials are revised and approved based on effective quality assurance activities.
2. **As far as possible, schools using the structured learning programme should be supported by on-site specialised reading coaching.**
    a. The main finding of the EGRS is that a structured learning programme aligned to the NCS, together with additional high quality reading support materials (graded reading books, flash cards, posters), can make a significant difference to learning outcomes, if accompanied by effective and carefully monitored support to teachers in the form of on-site coaching.

b. Coaching is the best alternative: Whereas previously very little evidence existed about effective large-scale teacher support modalities in South Africa, we now have evidence that on-site coaching to Foundation Phase teachers can shift learning outcomes, and that this is a cost-effective strategy. Modelling of lessons, in a safe space, as they navigate the lesson plans for teaching learners to read is critical.
c. Direct in-service training better than 'train-the-trainer' models: Direct in-service training of teachers (4 two-day workshops over the course of 2 years), while less effective than on-site coaching, is in turn likely to have more impact than "cascade" models where specialists "train the trainers" who then interact with teachers.
d. Existing subject advisers cannot fulfil the role of a coach: The low ratio of subject advisors to schools (especially in the Foundation Phase) makes it impossible for subject advisors to fulfil the role of reading coaches, as implemented in EGRS; nor do we recommend increasing the number of subject advisors to allow this since the recruitment process, oversight structures and modus operandi of the coaches is different to that of subject advisors.
e. In order to realize effective on-site coaching, the DBE should develop guidelines for on-site coaching as well as institutional support for the learning programme provided by subject advisors and curriculum leaders within schools.
f. Prioritizing schools is a viable option: On-site coaching interventions could be implemented in priority schools (e.g. 100 or 500 schools in a province) on a temporary basis (e.g. 2 years at a time) and through independent contracting and oversight structures. The cost for 100 schools would be about R6 million at current prices.

3. **Provinces should ring-fence finances for the implementation of the structured learning programme and on-site coaches**
    a. A finance review should be commissioned of the cost implications of the implementation of the structured learning programme and the on-site coaching. The review should provide the detailed costs involved of implementing the programme in a province, as well as investigate potential ways in which provinces could fund the implementation of the programme.

4. **Large-scale implementation should be immediately pursued in the districts of Ngaka Modiri Molema and Dr Kenneth Kaunda.**
    a. While systems are being set up to facilitate widespread implementation of the structured learning programme and on-site coaching, further implementation should build on the momentum created by the EGRS in Ngaka Modiri Molema and Dr Kenneth Kaunda.
    b. All schools in these districts should receive the lesson plans and integrated reading materials together with a short 1-day orientation training workshop. Subject advisors should receive training on the programme so that they are equipped to provide instructional support to teachers.
    c. A select group of schools should be identified to receive on-site coaching.

d. The expanded implementation in the North West should continue to be evaluated to establish whether the same level of impact can be achieved on a larger scale.

5. **Further research is needed to build on the initial findings of the EGRS.**
    a. Develop reading norms in the African languages: Reading norms cannot simply be adapted from one language into another due to differences in language structures. It is a complex exercise requiring longitudinal data. Therefore, the EGRS data could be used towards the development of reading norms in the African languages.
    b. Parental involvement needs further research and may be promising: Whilst parental involvement is a hugely deterministic factor in a child's learning outcomes, the biggest challenge from a policy perspective is how to shift parent involvement at scale. Given the potential cost-effectiveness of such interventions, researchers and policy-makers should continue to investigate mechanisms to do so.
    c. Learning what works in deep rural settings: Formative research and subsequent impact evaluation is required to figure out what kinds of school support programmes make a meaningful difference in deep rural settings.
    d. Measuring long-run EGRS impacts: The DBE is planning to administer subsequent data collections on the same sample of learners to measure the long run impacts of these reading interventions. It remains an open question whether a home language reading intervention that impacted on reading proficiency in the early grades will yield a persistent benefit to those learners who were impacted. It is conceivable both that the effect may dissipate over time or that it may compound if a solid reading foundation enables more learning later on and in different subject areas. It may be that this will depend on the quality of teaching received in subsequent years.
    e. EGRS for EFAL in Mpumalanga: A second phase of the Early Grade Reading Study (EGRS 2) is underway in the Mpumalanga province, since the start of 2017. This project aims to investigate the effectiveness of two alternative interventions on English as First Additional Language in the Foundation Phase.
    f. Early Grade Mathematics Study: The DBE is embarking on an Early Grade Mathematics Study over the next 2 to 5 years, with the first activity being a detailed scoping study to identify and design promising interventions with strong theories of change as well as cost-structures that would be sustainable on a large scale.

6. References

Adukia, A. (2017). Sanitation and Education. *American Economic Journal: Applied Economics, 9*(2), 23-59.

Altman, E. (2013). *Childhood nutrition in Chile: An assessment of a national school food program.* Atlanta: Thesis submitted to the Emory College of Arts and Sciences of Emory University in partial fulfilment of the requirements of the degree of Bachelor of Arts with Honors.

Andrabi, T., Das, J., & Khwaja, A. (2015). Report cards: The impact of providing schools and child test scores on educational markets. Unpublished.

Araujo, M., Carneiro, P., Cruz-Aguayo, Y., & Schady, N. (2016). Teacher quality and learning outcomes in kindergarten. *The Quarterly Journal of Economics, 131*(3), 1415–1453.

Banerjee, A., Banerji, R., Duflo, E., Glennerster, R., & Khemani, S. (2010). Pitfalls of Participatory Programs: Evidence from a Randomised Evaluation in Education in India. *American Economic Journal: Economic Policy, 2*(1), 1-30.

Banerji, R., Bhattacharjea, S., & Wadhwa, W. (2013). The Annual Status of Education Report (ASER). *Research in Comparative and International Education, 8*(3), 387-396.

Conn, K. (2017). Identifying effective education interventions in Sub-Saharan Africa: A meta-analysis of rigorous impact evaluations. *Review of Educational Research* .

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., & Sundararaman, V. (2011). School inputs, household substitution and test scores. *National Bureau of Economic Research Working Paper 16830*.

Deaton, A., & Cartwright, N. (2016). Understanding and misunderstanding randomised controlled trials. *WP 22595. National Bureau of Economic Research* .

Department of Basic Education. (2012). *Ministerial Audit of Provincial Reading Strategies.*

Department of Basic Education. (2015). *Action Plan to 2019: Towards the realisation of Schooling 2030.* Pretoria.

Diagne, A., Lô, M., Sokhna, O., & Diallo, F. (2014). Evaluation of the impact of school canteen programmes on internal effieciency of schools, cognitive acquisitions and learning capacities of students in rural primary schools in Senegal. . *Partnership for Economic Policy working paper*.

Ebenezer, R., Gunawardena, K., Kumarendran, B., Pathmeswaran, A., Jukes, M., Drake, L., & Silva, N. (2013). Cluster-randomised trial of the impact of school-based deworming and iron supplementation on the cognitive abilities of school children in Sri Lanka's plantation sector. . *Tropical Medicine and International Health, 18*, 942-951.

Fleisch, B., & Schöer, V. (2014). Large-scale instructional reform in the Global South: Insights from the mid-point evaluation of the Gauteng Primary Language and Mathematics Strategy. *South African Journal of Education, 34*(3), 1-12.

Fleisch, B., Schöer, V., Roberts, G., & Thornton, A. (2016). System-wide improvement of early-grade mathematics: New evidence from the Gauteng Primary Language and Mathematics Strategy. *International Journal of Educational Development, 49*, 157-174.

Fleisch, B., Taylor, S., Schöer, V., & Mabogoane, T. (2017). Failing to catch up in reading in the middle years: The findings of the impact evaluation of the Reading Catch-Up Programme in South Africa. *International Journal of educational Development, 53*, 36-47.

Fleisch, B., Taylor, S., Schöer, V., & Mabogoane, T. (2017). The value of large-scale randomised control trials in system-wide improvement: The case of the Reading Catch-Up Programme. *South African Journal of Education, 37*(1).

Freeman, M., Greene, L., Dreibelbis, R., Saboori, S., Muga, R., Brumback, B., & Rheingans, R. (2011). 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 and International Health, 17*(3), 380-391.

Fryer, R. (2017). The production of human capital in developed countries: Evidence from 196 randomized field experiments. *Handbook of Economic Field Experiments, 2*, 95–322.

Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspectives on Psychological Science, 9*(6), 641 - 651.

Glewwe, P., Hanushek, E., Humpage, S., & Ravina, R. (2014). School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010. In P. Glewwe (Ed.), *Educational Policy in Developing Countries.* Chicago and London: University of Chicago Press.

Glewwe, P., Illias, N., & Kremer, M. (2010). Teacher incentives. *American Economic Journal: Applied Economics, 2*(3), 205-227.

Glewwe, P., Kremer, M., & Moulin, S. (2009). Many children left behind? Textbooks and test scores in Kenya. *American Economic Journal: Applied Economics, 1*(1), 111-135.

Glewwe, P., Kremer, M., Moulin, S., & Zitzewitz. (2004). Retrospective vs. prospective analysis of school inputs: the case of flip charts in Kenya. *Journal of Development Economics, 74*, 251-268.

Gove, A., & Wetterberg, A. (2011). *The Early Grade Reading Assessment: Applications and Interventions to Improve Early Grade Literacy.* RTI Press.

Haddaway, N., Woodcock, P., Macura, B., & Collins, A. (2015). Making lierature reviews more reliable through application of lessons from systematic reviews. *Conservation Biology, 29*(6), 1596-1605.

Hanushek, E. (2010). The economic value of higher teacher quality. *National Center for Analysis of Longitudinal Data in Education Research, Working Paper No.56*.

Harris, D., & Sass, T. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics, 95*(7), 798–812.

He, F., Linden, L. L., & MacLeod, M. (2009). *A better way to teach children to read? Evidence from a randomized control trial.* Retrieved September 12, 2012, from https://webspace.utexas.edu/linden/www/Teach%20Children%20to%20Read.pdf

Hoadley, U. (2010). *What do we know about teaching and learning in primary schools in South Africa?* Stellenbosch: Appendix B to Van der Berg, S; Meyer, H; Reeves, C; van Wyk, C; Hoadley, U; Bot, M; & Armstrong, P 2010. 'Grade 3 Improvement Project: Main report and Recommendations" for Western Cape Education Department.

Ismail, S., Jarvis, E., & Borja-Vega, C. (2012). *Guyana's hinterland community-based school feeding program (SFP).* Washington, DC: World Bank.

Jackson, C., & Makarin, A. (2016). Can Online Off-The-Shelf Lessons Improve Student Outcomes? Evidence from A Field Experiment. *NBER Working Paper No. 22398*.

Jacob, B., & Lefgren, L. (2004). The impact of teacher training on student achievement quasiexperimental evidence from school reform efforts in Chicago. *Journal of Human Resources, 39*(1), 50–79.

Jukes, M., Zuilkowski, S., Parawan, A., & Lee, S. (2014). Evaluating the cognitive impacts of school health in the Philippines. *International Journal of Child Health and Human Development, 7*, 55-66.

Kennedy, M. (2016). How does professional development improve teaching? *Review of Educational Research, 86*(4), 945–980.

Kerwin, J., & Thornton, R. (2018). Making the Grade: The Sensitivity of Education Program Effectiveness to Input Choices and Outcome Measures. Retrieved January 30, 2018, from SSRN: https://ssrn.com/abstract=3002723 or http://dx.doi.org/10.2139/ssrn.3002723

Kraft, M., Blazar, D., & Hogan, D. (2016). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Education Research*.

Kremer, M., Brannen, C., & Glennerster, R. (2013). The Challenge of Education and Learning in the Developing World. *Science, 340*, 297-300.

Krishnaratne, S., White, H., & Carpenter, E. (2013). *Quality education for all children? What works in education in developing countries.* New Delhi: International Initiative for Impact Evaluation.

Lieberman, E. P. (2013). *Does Information Lead to More Active Citizenship? Evidence from an Education Intervention in Rural Kenya.* MIT Working Paper.

Lucas, A., & McEwan, P. N. (2014). Improving early-grade literacy in East Africa: Experimental evidence from Kenya and Uganda. *Journal of Policy Analysis and Management, 33*(4), 950-976.

Luo, R., Shi, Y., Zhang, L., Liu, C., Rozelle, S., Sharbono, B., . . . Martorell, R. (2012). Nutrition and educational performance in rural China's elementary schools: Results of a randomised control trial in Shaanxi province. *Economic Developmnet and Cultural Change, 60*, 735-772.

Machin, S., & McNally, S. (2005). Gender and student achievement in English schools. *Oxford Review of Economic Policy, 21*(3), 357-372.

McEwan, P. (2012). Improving learning in primary schools of developing countries: A meta-analysis of randomised experiments. *Review of Educational Research, 20*(10), 1-42.

McEwan, P. (2013). The impact of Chile's school feeding program on education outcomes. *Economics of Education Review, 32*, 122-139.

Miguel, E., & Kremer, M. (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica, 82*, 159-217.

Muralidharan, K., & Sundararaman, V. (2011). Teacher performance pay: Experimental evidence from India. *Journal of political economy, 119*(1), 39-77.

Murnane, R., & Ganimian, A. (2014). Improving educational outcomes in developing countries: Lessons from rigorous evaluations. *NBER Working Paper No 20284, National Bureau of Economic Research, Cambridge, MA*.

Nadel, S., & Pritchett, L. (2016). Searching for the Devil in the Details: Learning about Development Program Design. *Worping Paper 434. Center for Global Development*.

National Education Evaluation and Development Unit. (2012). *National Report 2012.*

Nguyen, T., & Lassibille, G. (2008). Improving management in education: Evidence from a randomised experiment in Madagascar. Unpublished manuscript 6.

Nitsaisook, M., & Anderson, L. (1989). An experimental investigation of the effectiveness of inservice teacher education in Thailand. *Teacher and Teacher Education , 5*(4), 287-302.

Nonoyama-Tarumi, Y., & Bredenberg, K. (2009). Impact of school readiness program interventions on children's learning in Cambodia. *Internation Journal of Educational Development, 29*(1), 39-45.

Pandey, P., & Goyal, S. a. (2011). Does information improve school accountability? Results of a large randomised trial. Washington, DC: World Bank.

Pandey, P., Goyal, S., & Sundararaman, V. (2009). Community participation in public schools: Impact of information campaigns in three Indian states. *Education Economics, 17*(3), 355-375.

Piper, B. (2009). *Integrated Education Program: Impact Study of SMRS Using Early Grade Reading Assessment in Three Provinces in South Africa.* Pretoria: United States Agency for International Development.

Piper, B., & Korda, M. (2011). *EGRA Plus Liberia: Program Evaluation Report.* RTI.

Piper, B., Zuilkowski, S., & Mugenda, A. (2014). Improving reading outcomes in Kenya: First-year effects of the PRIMR Initiative. *International Journal of Education Development, 37*(11-21), 11-21.

Popova, A., Evans, D., & Arancibia, V. (2016). Inside In-Service Teacher Training: What works and how do we measure it? Washington, DC.: Rise Programme.

Prinsloo, C. (2008). Practice makes perfect - Limpopo study shows the value of exercising literary muscles. *HSRC Review, 6*(4), 5-6.

Rivkin, S., Hanushek, E., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417–458.

Ross, K., & Zuze, T. (2004). Traditional and alternative views of school system performance. *IIEP Newsletter*(October-December), 8-9.

Sabarwal, S., Evans, D., & Marshak, A. (2014). The Permanent Input Hypothesis: The Case of Textbooks and (No) Student Learning in Sierra Leone. *World Bank Policy Research Working Paper 7021*.

Sailors, M., Hoffman, J., Pearson, P., Beretvas, S., & Matthee, B. (2010). The Effects of First- and Second-Language Instruction in Rural South African Schools. *Bilingual Research Journal*, 21-41.

Simeon, D., Grantham-McGregor, S., Callender, J., & Wong, M. (1995). Treatment of Trichuris trichiura infections improves growth, spelling scores and school attendance in some children. *Community and International Nutrition*, 1875-1883.

Snilstveit, B., Stevenson, J., Menon, R., Philips, D., Gallagher, E., Geleen, M., . . . Jimenez, E. (2016). *The impact of education programmes on learning and school participation in low- and middle-income countries.* London: International Initiative for Impact Evaluation.

Snilstveit, B., Stevenson, J., Menon, R., Phillips, D., Gallagher, E., Geleen, M., . . . Jimenez, E. (2016). *The impact of education programmes on learning and school participation in low- and middle-income countries: a systematic review summary report, 3ie Systematic Review Summary 7.* London: International Initiative for Impact Evaluation (3ie).

Spaull, N., Van der Berg, S., Wills, G., Gustafsson, M., & Kotze, J. (2016). *Laying Firm Foundations: Getting Reading Right.*

Taylor, N., Van der Berg, S., & Mabogoane, T. (2013). *Creating Effective Schools.* Cape Town: Pearson.

Taylor, S., & Von Fintel, M. (2016). Estimating the impact of language of instruction in South African primary schools: A fixed effects approach. *Economics of Education Review, 50*, 75-89.

Van der Berg, S. (2008). How effective are poor schools? Poverty and educational outcomes in South Africa. *Studies in Educational Evaluation, 34*, 145-154.

Watkins, W., Cruz, J., & 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.

Wong, H., Shi, Y., Luo, R., Zhang, L., & Rozelle, S. (2014). Improving the health and education of elementary school children in rural China: Iron supplementation versus nutritional training for parents. *Journal of Development Studies, 50*, 502-519.