International Initiative for Impact Evaluation

Rapid evidence assessment brief

Education

© Basic Education Sector Transformation Program / DFAT

Rapid evidence assessment of teachers' training programs in low-and middle-income countries

International research suggests that teacher professional development is an effective way of increasing teachers' competencies and, ultimately, improving student outcomes (Asian Development Bank, 2021; World Bank, 2016). However, developing countries face issues that hinder teacher knowledge and skills development, one of which relates to the quality of pre-and in-service teacher training (Global Partnership for Education, 2019).

Given the importance of teacher quality and adaptability, especially in blended settings after the pandemic (Saavedra et al., 2020), decision-makers have shown an interest in reviewing their teacher professional development policies. In fact, improvements in teacher quality and educational outcomes are one of the priority evaluation agenda of the National Economic and Development Authority (NEDA), a premier socioeconomic planning body of the government of the Philippines. They commissioned 3ie to conduct a rapid evidence assessment (REA) to answer the following questions: what are the effects of pre-service and in-service training programs for teachers on students learning outcomes; do these effects vary by context, training types, delivery modality, or other key moderators; and which are the monitoring and evaluation systems used to ensure the quality of the teacher training programs identified?

© Dominic Chavez / World Bank

State of the evidence on professional development programs in L&MICs

The search for relevant studies initially identified 11,565 records. **We included 101 evaluations of professional development programs for analysis.** The publication of these studies shows an increase since 2007, particularly in the last five years. Further details on the project flowchart are presented in figure A1 (see Appendix 1).

Most of the studies evaluated professional development programs using experimental designs (n = 79). Sixteen studies reported cost-effectiveness analyses. Studies were usually conducted in Sub-Saharan Africa (n = 33), Latin America (n = 31) and East Asia and the Pacific (n = 19). By country, studies in China (n = 12), India (n = 8), and Kenya (n = 8) were the most prevalent.

More than half of the studies evaluated multicomponent professional development programs (n = 63), a combination of teachers' training with other components, such as the provision of books or technology and school community engagement activities, to improve students' academic performance. **The remaining assessed single-component professional development programs (n = 38),** mainly focused on building or improving teachers' capacities through training. The average program duration was 8 months.

Almost all programs provided in-service training only (n = 97). Few implemented pre-service training (n = 2), or both pre- and in-service training (n = 2). **The programs were most frequently delivered in person (n = 74),** and some used both in-person and online delivery modes (n = 12).

The professional development programs focused most commonly on primary school levels (n = 73), followed by secondary (n = 23), and preschool levels (n = 14). Sixty-two studies reported the program setting: these focused on rural areas (n = 27), both rural and urban areas (n = 20), and urban areas only (n = 15).

The professional development programs were frequently funded by government agencies (n = 44), such as DFID and USAID. Other funders were multilateral organizations (n = 19), including World Bank, and NGOs (n = 19), such as 3ie and Save the Children. **Common program implementers were government agencies (n = 52),** NGOs (n = 27), and academic organizations (n = 17), such as J-PAL.

Evidence of effects of professional development programs in L&MICs

Among the subset of professional development programs focused on in-service training at primary school levels, conducted in middle-income countries and evaluated through experimental designs, **single-component programs result in lower math scores than multicomponent programs** (SMD = -0.04, 95%CI = -0.07, -0.01), though the difference is small. When comparing language scores, there is also an advantage of multicomponent programs, although the difference with single-component programs is not statistically significant (SMD = -0.02, 95%CI = -0.07, 0.02).

In addition, we looked at the characteristics of a sample of single- and multicomponent programs that reported larger positive effects on student learning. Tables A2 and A3 (Appendix 2) present summaries of these example programs, which show some common characteristics:

- These professional development programs were usually long-term, covering the academic year.
- These professional development programs frequently provided a group training session plus ongoing support throughout the length of the program. Examples of this support include classroom observation, provision of feedback, and regular text messages to teachers.
- Considerations for programming and policy **professional development programs tend to have positive effects on student's learning outcome.** While there is variation in the effects across programs, there are also ample opportunities to have effective professional development programs.
- Our analyses suggest that **the design of the programs matter.** These are some options to consider when planning teacher professional development programs to improve student learning:
  - Consider implementing multicomponent programs, in which teacher training is complemented with the provision of resources and other school community activities aimed to support teachers and students. Comprehensive programs may work better at tackling complex education issues, particularly those that include individual components.
  - Consider implementing long-term programs with continued teacher support throughout the program. This could help ensure that the implementation of the program is appropriate and may provide teachers with more learning opportunities. While this finding was drawn by looking at the characteristics of programs with larger effects, it is also aligned with recent work around international teacher professional development programs (Popova et al., 2022).
  - Consider promoting active engagement with academic organizations and the program research teams. It is likely that the involvement of these stakeholders encourages the design of evidence-based programs.
  - Programs implemented locally may work better compared to national and regional programs. Local programs may be more adept at addressing particular local needs.

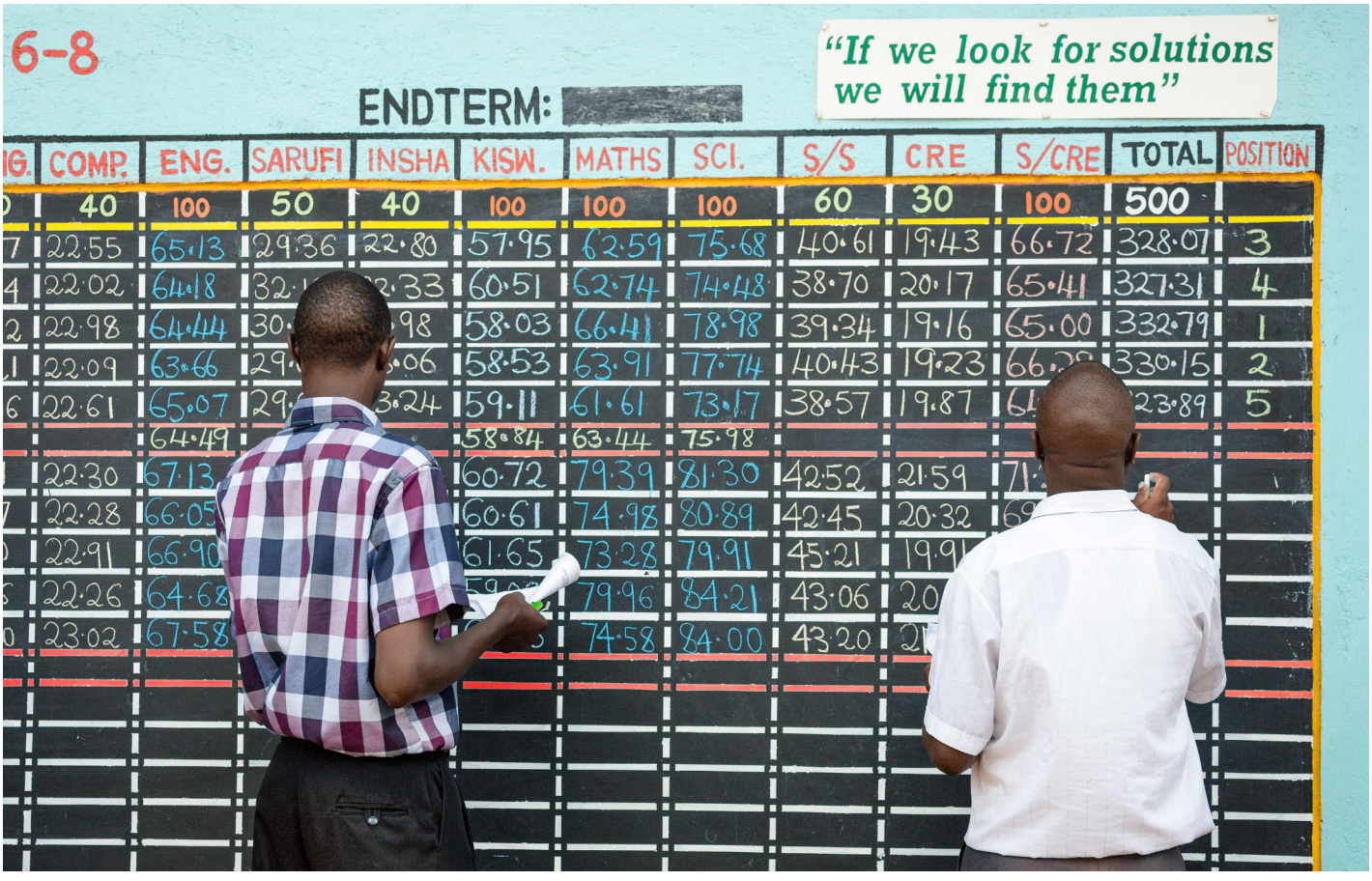
Evidence of effects of professional development programs in L&MICs

We extracted data from a subset of single- and multicomponent Professional Development programs (n = 31 single-component and n = 16 multicomponent), which focused on in-service teacher training and were evaluated using experimental designs. We conducted meta-analyses and moderator analyses to identify the effects of these professional development programs. The full list of program characteristics used in the moderator analysis, as well as the individual findings, are presented in Table A1 (Appendix 2).

When considering all studies and outcomes for all subjects, the evidence suggests that **professional development programs can have positive effects on students' learning outcomes** (Standardized Mean Difference (SMD) = 0.11, 95% Confidence Interval (CI) = 0.06, 0.16). Among these studies, professional development programs delivered in-person have larger effects compared to hybrid programs (SMD = 0.08, 95%CI = 0.02, 0.14), and those implemented at the national scale result in smaller effects compared to local professional development programs (SMD = -0.21, 95%CI = -0.35, -0.06).

Single-component professional development programs show a small positive effect on language but not on math outcomes (SMD = 0.08, 95%CI = 0.03, 0.12; and SMD = 0.02, 95%CI = -0.01, 0.04; respectively). Single-component programs **funded by academic organizations** have larger effects in both math and language scores (SMD = 0.16, 95%CI = 0.08, 0.24; and SMD = 0.21, 95%CI = 0.11, 0.32; respectively). National and regional programs result in smaller effects on math outcomes compared to **programs implemented locally** (SMD = -0.13, 95%CI = -0.20, -0.06; and SMD = -0.14, 95%CI = -0.22, -0.07; respectively), and **programs including lectures and where teachers were trained by the research team or government officials** have higher language scores (SMD = 0.29, 95% CI = 0.16, 0.43; SMD = 0.30, 95%CI = 0.17, 0.43; and SMD = 0.26, 95%CI = 0.14, 0.39; respectively).

Multicomponent professional development programs show positive but small effects on both language and math outcomes SMD = 0.07, 95%CI = 0.04, 0.09; and SMD = 0.05, 95%CI = 0.03, 0.07; respectively). Professional development programs with the following characteristics – compared to programs without these features – have larger effects on language scores: including **individual activities**, such as developing lesson plans or reflecting on practice after receiving feedback, has the largest effect (SMD = 0.50, 95%CI = 0.07, 0.93). However, using a **cascade training model**, where some teachers/mentors receive training first and then they train participating teachers (SMD = 0.10, 95%CI = 0.07, 0.14), and embedding **coaching support** throughout the program (SMD = 0.07, 95%CI = 0.02, 0.12) show positive but much smaller effects on language scores. In turn, programs with funding from private organizations and coaching activities have slightly smaller effects on math scores (SMD = -0.05, 95%CI = -0.08, -0.02; and SMD = -0.04, 95%CI = -0.075, -0.003; respectively).



© Kelley Lynch / GPE

Considerations for future research

Teacher professional development is an important topic for policy development, which could be explored further. Based on the findings from this REA, we identify several areas where future research could be directed:

- **Incorporate the full body of evidence** identified at the search stage of the REA (n = 101 evaluations). The analyses presented here cover around half of this global literature.
- **Incorporate the analysis of subskill measures.** We combined similar outcomes (e.g. letter identification and reading, or addition and geometry) to create "language" and "math" outcome measures for each program. Additional analysis using specific indicators could help understand if professional development programs have different effects at the subskill level.
- **Incorporate teacher-level outcomes** to explore the theory of change of professional development programs. We presented analyses on student-level outcomes only, but we also identified and excluded 92 studies due to reporting other types of outcomes, including measures of teacher performance, attitudes, and beliefs.
- **Incorporate additional information from included professional development programs.** Many times, the impact evaluation reports did not include comprehensive information about the program design, implementation and monitoring. An additional search of program documentation and other evaluations, such as process evaluations, could enrich the analyses.
- **Incorporate rigorous impact evaluations as part of the design of pre-service professional development programs.** This would contribute to building up the scarce evidence around the effects of these programs in L&MICs.

© Kelley Lynch / GPE

© Pierre Prakash / EU / ECHO

Additional evidence of professional development programs

Table 1: Summary of professional development programs conducted in the Philippines

Study	San Antonio et al., 2011	Abeberese et al., 2014
Context	Setting: Not specified School level: Primary Grade: 6 Scale: Regional (Bicol region)	Setting: Not specified School level: Primary Grade: 4 Scale: Regional (Tarlac province)
Professional development program	Single-component Module-Based Professional development for Teachers (MBPDT), aimed to improve teaching quality in math. MBDPT was an in-service training program, delivered in-person. Teachers participated in an orientation workshop focusing on modules to enhance their professional competence, including: (1) Learning Activities for Different Learners, (2) Teaching Approaches, (3) Developing Higher Order Thinking Skills in Teaching Mathematics, (4) Mathematics (Part I), and (5) Mathematics (Part II).	Multicomponent Sa Aklat Sisikat Read-a-Thon, a reading program with in-service training, delivered in-person, and based on three components: (1) teacher training to incorporate reading in the curriculum, 2) provision of age-appropriate books, and (3) delivering 31-day "read-a-thons" to encourage children to read and support teachers to incorporate reading into their classes. The school system expects students to develop reading fluency by the fourth grade, hence the program targeted 4th grade students.
Program length (exposure)	5 weeks	3 months
Training length	5 weeks	2 days
Frequency of training and implementation	Teachers attended an orientation workshop on ways of using modules to enhance their professional competence. Every week for five weeks, teachers received new material and follow-ups from the school heads and supervisors in terms of how they were studying the modules. Teachers were trained by the research team: local educational practitioners.	Sa Aklat Sisikat provided a two-day training session focused on how to teach reading in an engaging way, how to increase reading at school and at children's homes, and how to implement a "read-a-thon." Sa Aklat Sisikat is an NGO based in Manila and has implemented reading programs in all provinces of the country.
Summary of results	The MBPDT program did not lead to significant effects on teachers' commitment levels, teachers' professional content knowledge and students' math proficiency levels. The sample size in this study was small (50 classrooms), and the authors highlighted that it was low-cost with possibilities to be expanded.	With a sample size of 5,510 students in 100 schools, the authors found immediate improvements in reading test scores of 0.13 standard deviations (SD) and of 0.06 SD three months later. More reading in school also led to a small increase in the number of books children read at home. The authors suggested creating teachers' incentives and strategies to emphasize the importance of reading.

Government monitoring of professional development programs

The studies did not usually report information about the monitoring of professional development programs. However, professional development programs conducted in India (Banerjee et al., 2016; Duflo et al., 2015) embedded a government-led monitoring system. The Haryana state government supported the programs' implementation through their existing school monitoring system: Associate Block Resource Coordinators (ABRCs). The ABRCs acted as block and district supervisors and field-level monitors of schools. Although the ABRCs were initially established to ensure better professional development management by monitoring and guiding school teachers, the lack of practical training on the operationalization of monitoring processes limited their role and they were generally involved in administrative tasks. Hence, the state government asked the research team to help revive the ABRCs. For this, the Haryana government supported various activities, including:

- Creating a new system within their body, where academic leaders within the Haryana government can monitor and advise teachers during the programs' implementation
- Assigning district officials, the management of monitoring tools and the training of block-level officials on monitoring teachers, analyzing the data, and writing the monitoring reports
- Supporting district officials to facilitate monthly meetings with program stakeholders, including block-level officials, the State Council for Education Training and Research (SCERT), the research team, and the NGO implementing the professional development programs.
- Requesting senior government officials to take action on issues being discussed during these monthly meetings.

References

Abeberese, A. B., Kumler, T. J., & Linden, L. L. (2014). Improving Reading Skills by Encouraging Children to Read in School: A Randomized Evaluation of the Sa Aklat Sisikat Reading Program in the Philippines. The Journal of Human Resources, 49(3), 611–633. https://www.jstor.org/stable/23798604

Ashraf, N., Banerjee, A., & Nourani, V. (2020). Learning to Teach by Learning to Learn [Job Market Paper]. https://www.povertyactionlab.org/sites/default/files/research-paper/working-paper-3647_Learning-to-teach_Banerjee_Nov2020_0.professional developmentf

Asian Development Bank. (2021). Teacher Professional development Case Studies: K-12, TVET, and Tertiary Education. Asian Development Bank. https://doi.org/10.22617/SPR210293

Baker-Henningham, H., Scott, Y., Bowers, M., & Francis, T. (2019). Evaluation of a Violence-Prevention Programme with Jamaican Primary School Teachers: A Cluster Randomised Trial. International Journal of Environmental Research and Public Health, 16(15), 2797. https://doi.org/10.3390/ijerph16152797

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukherji, S., Shotland, M., & Walton, M. (2016). Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of "Teaching at the Right Level" in India. Working paper No. 22746. National Bureau of Economic Research. https://doi.org/10.3386/w22746

Beg, S., Halim, W., Lucas, A. M., & Saif, U. (2022). Engaging Teachers with Technology Increased Achievement, Bypassing Teachers Did Not. American Economic Journal: Economic Policy, 14(2), 61–90. https://doi.org/10.1257/pol.20200713

Castro, J. F., Glewwe, P., Heredia-Mayo, A., Montero, R., & Castro, J. F. (2021). Work with What You've Got: Improving Teachers' pedagogical Skills at Scale in Rural Peru. Staff Working Paper P21-4. University of Minnesota, Department of Applied Economics. https://ageconsearch.umn.edu/record/316662

Chen, G., Chan, C. K. K., Chan, K. K. H., Clarke, S. N., & Resnick, L. B. (2020). Efficacy of video-based teacher professional development for increasing classroom discourse and student learning. Journal of the Learning Sciences, 29(4–5), 642–680. https://doi.org/10.1080/10508406.2020.1783269

Ding, H., & Rubie-Davies, C. M. (2019). Teacher expectation intervention: Is it effective for all students? Learning and Individual Differences, 74, 101751. https://doi.org/10.1016/j.lindif.2019.06.005

Duflo, E., Berry, J., Mukerji, S., & Shotland, M. (2015). A wide angle view of learning: Evaluation of the CCE and LEP programmes in Haryana, India (3ie Impact Evaluation Report No. 22). International Initiative for Impact Evaluation (3ie). https://www.3ieimpact.org/evidence-hub/publications/impact-evaluations/wide-angle-view-learning-evaluation-cce-and-lep

Fisher, Z., & Tipton, E. (2015). Robumeta: An R-package for robust variance estimation in meta-analysis. https://doi.org/10.48550/arXiv.1503.02220

Fuje, H., & Tandon, P. (2018). When do in-service teacher training and books improve student achievement? Experimental evidence from Mongolia. Review of Development Economics, 22(3), 1360–1383. https://doi.org/10.1111/rode.12387

Global Partnership for Education. (2019). GPE Results Report 2019. https://www.globalpartnership.org/sites/default/files/2019-06-gpe-results-report-2019.professional developmentf

He, F., Linden, L., & MacLeod, M. (2008). How to Teach English in India: Testing the Relative Productivity of Instruction Methods within the Pratham English Language Education Program [Working Paper]. JPAL. https://www.povertyactionlab.org/sites/default/files/research-paper/153_318_How%20to%20teach%20English_He_Linden_Macleod_%20July2008.professional developmentf

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. Res Synth Methods, 1, 39–65. https://doi.org/10.1002/jrsm.5

Jukes, M. C. H., Turner, E. L., Dubeck, M. M., Halliday, K. E., Inyega, H. N., Wolf, S., Zuilkowski, S. S., & Brooker, S. J. (2017). Improving Literacy Instruction in Kenya Through Teacher Professional development and Text Messages Support: A Cluster Randomized Trial. Journal of Research on Educational Effectiveness, 10(3), 449–481. https://doi.org/10.1080/19345747.2016.1221487

Popova, A., Evans, D. K., Breeding, M. E., & Arancibia, V. (2022). Teacher Professional development around the World: The Gap between Evidence and Practice. The World Bank Research Observer, 37(1), 107–136. https://doi.org/10.1093/wbro/lkab006

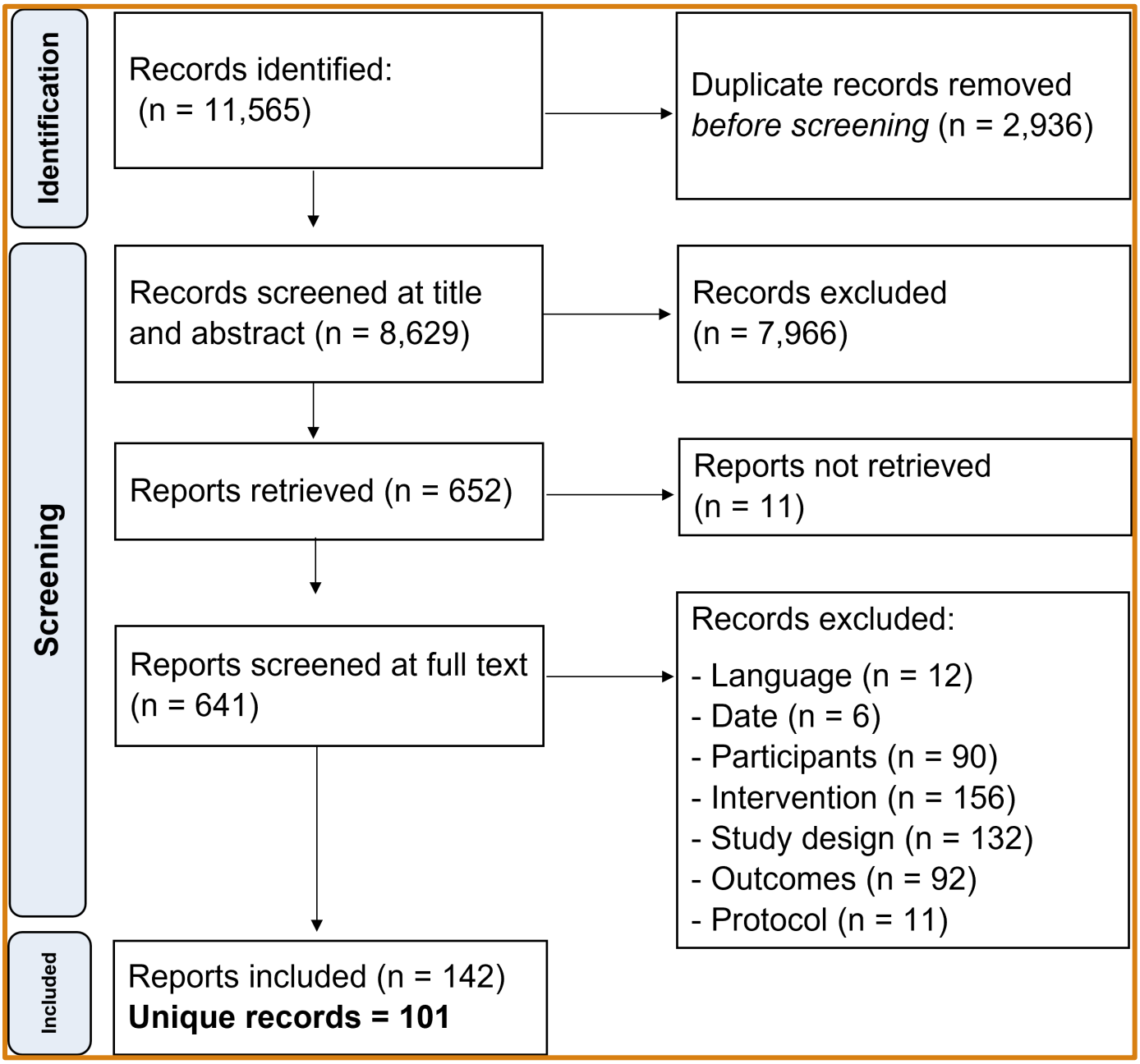
Saavedra, J., Aedo, C., Arias, O., Pushparatnam, A., Gutiérrez Bernal, M., & Rogers, H. (2020). Realizing the Future of Learning: From Learning Poverty to Learning for Everyone, Everywhere. World Bank Group. http://documents.worldbank.org/curated/en/250981606928190510/Realizing-the-Future-of-Learning-From-Learning-Poverty-to-Learning-for-Everyone-Everywhere

San Antonio, D. M., Morales, N. S., & Moral, L. S. (2011). Module□based professional development for teachers: A cost□effective Philippine experiment. Teacher Development, 15(2), 157–169. https://doi.org/10.1080/13664530.2011.571496

World Bank. (2016). Developing a Proficient and Motivated Teacher Workforce in the Philippines (Philippines Education Note No. 3; p. 16). World Bank. http://hdl.handle.net/10986/24746

Appendix 1

Figure A1: Project flowchart



Appendix 2

Table A1: Findings for main models and moderator analyses

	All programs All outcomes (1)	Single-component programs Math outcomes (2)	Multicomponent programs Math outcomes (3)	Single-component programs Language outcomes (4)	Multicomponent programs Language outcomes (5)
Main model (without moderators)	SMD = 0.11, 95% CI = 0.06, 0.16 Positive, statistically significant	SMD = 0.02, 95% CI = -0.01, 0.04 Not statistically significant	SMD = 0.05, 95% CI = 0.03, 0.07 Positive, statistically significant	SMD = 0.08, 95% CI = 0.03, 0.12 Positive, statistically significant	SMD = 0.07, 95% CI = 0.04, 0.09 Positive, statistically significant
Moderators					
Continent (compared to East Asia and Pacific)	Insufficient degrees of freedom, hence unreliable results	Not sufficient programs in each moderator category	South Asia: SMD = -0.07, 95% CI = -0.17, 0.02 Not statistically significant Sub-Saharan Africa: SMD = -0.08, 95% CI = -0.19, 0.03 Not statistically significant	Latin America and Caribbean: SMD = -0.06, 95% CI = -0.25, 0.12 Not statistically significant Sub-Saharan Africa: SMD = -0.03, 95% CI = -0.21, 0.15 Not statistically significant	Not sufficient programs in each moderator category
Country	Not sufficient programs in each moderator category	Not sufficient programs in each moderator category	Not sufficient programs in each moderator category	Not sufficient programs in each moderator category	Not sufficient programs in each moderator category
Setting (compared to both: urban and rural)	Rural: SMD = -0.06, 95% CI = -0.18, 0.06 Not statistically significant Urban: SMD = 0.03, 95% CI = -0.15, 0.21 Not statistically significant	Not sufficient programs in each moderator category	Rural: SMD = 0.02, 95% CI = -0.03, 0.08 Not statistically significant	Rural: SMD = 0.02, 95% CI = -0.08, 0.11 Not statistically significant Urban: SMD = 0.05, 95% CI = -0.05, 0.14 Not statistically significant	Rural: SMD = 0.00, 95% CI = -0.11, 0.11 Not statistically significant
Scale of the program (compared to local)	National programs: SMD = -0.21, 95% CI = -0.35, -0.06 Negative, statistically significant Regional programs: SMD = -0.14, 95% CI = -0.29, 0.02 Not statistically significant	National programs: SMD = -0.13, 95% CI = -0.20, -0.06 Negative, statistically significant Regional programs: SMD = -0.14, 95% CI = -0.22, -0.07 Negative, statistically significant	National programs: SMD = -0.03, 95% CI = -0.17, 0.10 Not statistically significant Regional programs: SMD = -0.04, 95% CI = -0.14, 0.05 Not statistically significant	National programs: SMD = -0.12, 95% CI = -0.34, 0.09 Not statistically significant Regional programs: SMD = -0.13, 95% CI = -0.27, 0.00 Not statistically significant	Not sufficient programs in each moderator category
Program funder: Academic organization (binary indicator)	SMD = 0.03, 95% CI = -0.12, 0.18 Not statistically significant	SMD = 0.16, 95% CI = 0.08, 0.24 Positive, statistically significant	SMD = -0.03, 95% CI = -0.06 0.01 Not statistically significant	SMD = 0.22, 95% CI = 0.11, 0.32 Positive, statistically significant	SMD = 0.02, 95% CI = -0.03, 0.07 Not statistically significant
Program funder: Government agency (binary indicator)	SMD = 0.03, 95% CI = -0.06, 0.13 Not statistically significant	SMD = 0.04, 95% CI = -0.03, 0.10 Not statistically significant	SMD = -0.03, 95% CI = -0.06, 0.01 Not statistically significant	SMD = -0.01, 95% CI = -0.13, 0.11 Not statistically significant	SMD = 0.03, 95% CI = -0.02, 0.08 Not statistically significant
Program funder: Multilateral organization (binary indicator)	Insufficient degrees of freedom, hence unreliable results	Not sufficient programs in each moderator category	Not sufficient programs in each moderator category	SMD = -0.07, 95% CI = -0.20, 0.06 Not statistically significant	Not sufficient programs in each moderator category
Program funder: Private organization (binary indicator)	Insufficient degrees of freedom, hence unreliable results	Not sufficient programs in each moderator category	SMD = -0.05, 95% CI = -0.08, -0.02 Negative, statistically significant	SMD = -0.06, 95% CI = -0.20, 0.07 Not statistically significant	SMD = 0.02, 95% CI = -0.03, 0.07 Not statistically significant
Program funder: NGO (binary indicator)	Insufficient degrees of freedom, hence unreliable results	SMD = -0.05, 95% CI = -0.11, 0.02 Not statistically significant	SMD = -0.01, 95% CI = -0.06, 0.04 Not statistically significant	SMD = 0.01, 95% CI = -0.18, 0.21 Not statistically significant	SMD = -0.01, 95% CI = -0.09, 0.07 Not statistically significant

	All programs All outcomes (1)	Single-component programs Math outcomes (2)	Multicomponent programs Math outcomes (3)	Single-component programs Language outcomes (4)	Multicomponent programs Language outcomes (5)
Program implementer: Academic organization (binary indicator)	SMD = 0.03, 95% CI = -0.14, 0.19 Not statistically significant	SMD = -0.01, 95% CI = -0.10, 0.09 Not statistically significant	SMD = -0.02, 95% CI = -0.09, 0.06 Not statistically significant	SMD = -0.07, 95% CI = -0.18, 0.05 Not statistically significant	SMD = -0.01, 95% CI = -0.15, 0.13 Not statistically significant
Program implementer: Government agency (binary indicator)	SMD = -0.10, 95% CI = -0.20, 0.00 Not statistically significant	SMD = -0.03, 95% CI = -0.10, 0.04 Not statistically significant	SMD = 0.03, 95% CI = -0.02, 0.08 Not statistically significant	SMD = -0.04, 95% CI = -0.17, 0.09 Not statistically significant	SMD = -0.03, 95% CI = -0.12, 0.07 Not statistically significant
Program implementer: Multilateral organization (binary indicator)	SMD = -0.01, 95% CI = -0.06, 0.04 Not statistically significant	Not sufficient programs in each moderator category	Not sufficient programs in each moderator category	Not sufficient programs in each moderator category	Not sufficient programs in each moderator category
Program implementer: Private organization (binary indicator)	Insufficient degrees of freedom, hence unreliable results	SMD = 0.02, 95% CI = -0.17, 0.21 Not statistically significant	Not sufficient programs in each moderator category	SMD = -0.02, 95% CI = -0.18, 0.14 Not statistically significant	Not sufficient programs in each moderator category
Program implementer: NGO (binary indicator)	Insufficient degrees of freedom, hence unreliable results	SMD = 0.05, 95% CI = -0.20, 0.30 Not statistically significant	SMD = -0.04, 95% CI = -0.08, 0.01 Not statistically significant	SMD = 0.00, 95% CI = -0.19, 0.19 Not statistically significant	SMD = 0.01, 95% CI = -0.07, 0.09 Not statistically significant
School level: Preschool (binary indicator)	Insufficient degrees of freedom, hence unreliable results	Not sufficient programs in each moderator category	Not sufficient programs in each moderator category	SMD = 0.09, 95% CI = -0.27, 0.46 Not statistically significant	Not sufficient programs in each moderator category
School level: Primary (binary indicator)	SMD = -0.05, 95% CI = -0.21, 0.10 Not statistically significant	SMD = -0.02, 95% CI = -0.09, 0.05 Not statistically significant	Not sufficient programs in each moderator category	SMD = 0.01, 95% CI = -0.15, 0.17 Not statistically significant	Not sufficient programs in each moderator category
School level: Secondary (binary indicator)	SMD = -0.06, 95% CI = -0.14, 0.03 Not statistically significant	SMD = -0.01, 95% CI = -0.08, 0.05 Not statistically significant	SMD = 0.02, 95% CI = -0.05, 0.09 Not statistically significant	SMD = -0.06, 95% CI = -0.19, 0.08 Not statistically significant	Not sufficient programs in each moderator category
Delivery mode (compared to hybrid)	In-person delivery: SMD = 0.08, 95% CI = 0.02, 0.14 Positive, statistically significant	In-person delivery: SMD = -0.01, 95% CI = -0.07, 0.06 Not statistically significant	In-person delivery: SMD = -0.04, 95% CI = -0.14, 0.05 Not statistically significant	In-person delivery: SMD = 0.08, 95% CI = -0.13, 0.30 Not statistically significant	In-person delivery: SMD = 0.06, 95% CI = -0.27, 0.15 Not statistically significant
Length of the program (exposure in months)	Insufficient degrees of freedom, hence unreliable results	SMD = -0.00, 95% CI = -0.01, 0.00 Not statistically significant	SMD = -0.00, 95% CI = -0.00, 0.00 Not statistically significant	SMD = -0.00, 95% CI = -0.01, 0.01 Not statistically significant	SMD = -0.00, 95% CI = -0.01, 0.00 Not statistically significant
Length of the training (exposure in days)	SMD = -0.00, 95% CI = -0.00, 0.00 Not statistically significant	SMD = 0.00, 95% CI = -0.00, 0.00 Not statistically significant	SMD = -0.00, 95% CI = -0.00, 0.00 Not statistically significant	SMD = 0.00, 95% CI = -0.00, 0.00 Not statistically significant	SMD = -0.00, 95% CI = -0.00, 0.00 Not statistically significant
Did the program include group activities (e.g. learning circles)? (binary indicator)	SMD = 0.06, 95% CI = -0.06, 0.17 Not statistically significant	SMD = 0.05, 95% CI = -0.01, 0.11 Not statistically significant	SMD = -0.03, 95% CI = -0.08, 0.02 Not statistically significant	SMD = 0.08, 95% CI = -0.02, 0.19 Not statistically significant	SMD = -0.02, 95% CI = -0.10, 0.06 Not statistically significant
Did the program include individual activities (e.g. self-study)? (binary indicator)	SMD = -0.06, 95% CI = -0.15, 0.02 Not statistically significant	SMD = 0.04, 95% CI = -0.02, 0.10 Not statistically significant	SMD = 0.26, 95% CI = -0.11, 0.62 Not statistically significant	SMD = -0.05, 95% CI = -0.17, 0.07 Not statistically significant	SMD = 0.50, 95% CI = 0.07, 0.93 Positive, statistically significant
Did the program include coaching or mentoring of individual teachers? (binary indicator)	SMD = 0.02, 95% CI = -0.06, 0.10 Not statistically significant	SMD = -0.00, 95% CI = -0.07, 0.06 Not statistically significant	SMD = -0.04, 95% CI = -0.07, -0.00 Negative, statistically significant	SMD = -0.00, 95% CI = -0.12, 0.11 Not statistically significant	SMD = 0.07, 95% CI = 0.02, 0.12 Positive, statistically significant
Did the program include a cascade model of training? (binary indicator)	Insufficient degrees of freedom, hence unreliable results	SMD = 0.04, 95% CI = -0.33, 0.40 Not statistically significant	SMD = -0.02, 95% CI = -0.08, 0.03 Not statistically significant	SMD = -0.02, 95% CI = -0.21, 0.17 Not statistically significant	SMD = 0.10, 95% CI = 0.07, 0.14 Positive, statistically significant

	All programs All outcomes (1)	Single-component programs Math outcomes (2)	Multicomponent programs Math outcomes (3)	Single-component programs Language outcomes (4)	Multicomponent programs Language outcomes (5)
Who conducted the training: Other school teachers (binary indicator)	SMD = -0.03, 95% CI = -0.15, 0.08 Not statistically significant	SMD = -0.01, 95% CI = -0.08, 0.05 Not statistically significant	SMD = 0.03, 95% CI = -0.02, 0.08 Not statistically significant	SMD = -0.10, 95% CI = -0.20, 0.01 Not statistically significant	SMD = -0.04, 95% CI = -0.14, 0.06 Not statistically significant
Who conducted the training: University professors or post-graduate students (binary indicator)	SMD = -0.06, 95% CI = -0.16, 0.05 Not statistically significant	SMD = -0.03, 95% CI = -0.09, 0.04 Not statistically significant	Not sufficient programs in each moderator category	SMD = -0.08, 95% CI = -0.21, 0.04 Not statistically significant	Not sufficient programs in each moderator category
Who conducted the training: Researchers (binary indicator)	SMD = 0.21, 95% CI = -0.01, 0.43 Not statistically significant	SMD = 0.03, 95% CI = -0.04, 0.10 Not statistically significant	Not sufficient programs in each moderator category	SMD = 0.30, 95% CI = 0.17, 0.43 Positive, statistically significant	Not sufficient programs in each moderator category
Who conducted the training: Government officials (binary indicator)	Insufficient degrees of freedom, hence unreliable results	SMD = -0.03, 95% CI = -0.09, 0.04 Not statistically significant	Not sufficient programs in each moderator category	SMD = 0.26, 95% CI = 0.14, 0.39 Positive, statistically significant	Not sufficient programs in each moderator category
Who conducted the training: Others (e.g. NGO officials) (binary indicator)	Insufficient degrees of freedom, hence unreliable results	SMD = 0.06, 95% CI = -0.29, 0.41 Not statistically significant	SMD = -0.04, 95% CI = -0.08, 0.01 Not statistically significant	SMD = 0.12, 95% CI = -0.23, 0.46 Not statistically significant	SMD = 0.01, 95% CI = -0.06, 0.08 Not statistically significant
Did the program include follow-up activities (e.g. SMS)? (binary indicator)	SMD = 0.04, 95% CI = -0.04, 0.12 Not statistically significant	SMD = 0.04, 95% CI = -0.01, 0.10 Not statistically significant	SMD = 0.02, 95% CI = -0.02, 0.06 Not statistically significant	SMD = 0.08, 95% CI = -0.03, 0.19 Not statistically significant	SMD = -0.01, 95% CI = -0.06, 0.05 Not statistically significant
Did the program include workshops? (binary indicator)	SMD = 0.03, 95% CI = -0.05, 0.10 Not statistically significant	SMD = 0.03, 95% CI = -0.03, 0.09 Not statistically significant	SMD = 0.02, 95% CI = -0.04, 0.07 Not statistically significant	SMD = -0.07, 95% CI = -0.18, 0.04 Not statistically significant	SMD = -0.02, 95% CI = -0.11, 0.07 Not statistically significant
Did the program include lectures? (binary indicator)	Insufficient degrees of freedom, hence unreliable results	SMD = -0.02, 95% CI = -0.09, 0.04 Not statistically significant	SMD = 0.04, 95% CI = -0.02, 0.11 Not statistically significant	SMD = 0.29, 95% CI = 0.16, 0.43 Positive, statistically significant	SMD = 0.03, 95% CI = -0.11, 0.17 Not statistically significant
Was the program based on needs diagnostics or designed with local stakeholders? (binary indicator)	SMD = -0.03, 95% CI = -0.14, 0.07 Not statistically significant	SMD = -0.04, 95% CI = -0.11, 0.03 Not statistically significant	SMD = -0.04, 95% CI = -0.18, 0.10 Not statistically significant	SMD = 0.00, 95% CI = -0.15, 0.15 Not statistically significant	SMD = -0.01, 95% CI = -0.09, 0.07 Not statistically significant
Did the program focus on subject content? (binary indicator)	SMD = 0.02, 95% CI = -0.09, 0.13 Not statistically significant	SMD = -0.03, 95% CI = -0.09, 0.03 Not statistically significant	SMD = -0.01, 95% CI = -0.06, 0.03 Not statistically significant	SMD = -0.03, 95% CI = -0.16, 0.10 Not statistically significant	SMD = -0.01, 95% CI = -0.06, 0.05 Not statistically significant
Did the program focus on pedagogy? (binary indicator)	Insufficient degrees of freedom, hence unreliable results	SMD = 0.04, 95% CI = -0.02, 0.11 Not statistically significant	SMD = -0.03, 95% CI = -0.07, 0.01 Not statistically significant	Not sufficient programs in each moderator category	SMD = 0.02, 95% CI = -0.03, 0.08 Not statistically significant
Did the program include the use of technology? (binary indicator)	SMD = -0.08, 95% CI = -0.16, 0.01 Not statistically significant	SMD = -0.00, 95% CI = -0.07, 0.06 Not statistically significant	SMD = 0.03, 95% CI = -0.01, 0.08 Not statistically significant	SMD = -0.04, 95% CI = -0.17, 0.10 Not statistically significant	SMD = -0.01, 95% CI = -0.08, 0.05 Not statistically significant

Note: Column 1 shows the results of robust variance estimation models (Fisher & Tipton, 2015; Hedges et al., 2010), which allow to incorporate all programs and all learning outcomes, regardless of the subject measured. The results of single- and multicomponent programs on math learning outcomes are shown in columns 2-3. The findings of single- and multicomponent programs on language learning outcomes are shown in columns 4-5. All columns present statistically significant effects at 95% confidence level. The data extracted from **professional development** programs and moderators used in these analyses were based on Popova and colleagues' work (2022).

Table A2: Examples of single-component professional development programs with positive and precise effects on student test scores

Study	Chen et al., 2020 – China	Castro et al., 2021 – Peru	Baker-Henningham et al., 2019 – Jamaica	Ashraf et al., 2020 – Uganda	Ding & Rubie-Davies, 2019 – China
Context	Setting: Not specified School level: Secondary Grades: 6 and 7 Scale: Local	Setting: Rural School level: Primary Grades: 2 and 4 Scale: National	Setting: Urban School level: Primary Grade: 1 Scale: Local	Setting: Not specified School level: Primary Grade: 1 Scale: Local	Setting: Urban School level: Primary Grade: 8 Scale: Local
professional development Program	Teachers were filmed in the classroom, assessed by coaches and received feedback on teaching quality.	Teachers were observed in the classroom, assessed by coaches and received feedback on teaching quality.	Violence-prevention training, including use of demonstrations, practice activities with feedback and group discussions.	1) Teachers' training 2) Classroom visits	Teacher training on three areas of teaching: challenging tasks, detailed feedback, and personal regard (immediacy).
Program length (exposure)	1 school year	1 school year	8 months	1 school year	4.5 months
Training length	1 school year	1 school year	8 months	Every term for 2 weeks	4.5 months
Frequency of training and implementation	Five 2-hour sessions every month or two. Coaches used selected parts of the classroom video observations to discuss and reflect with teachers on areas of improvements. Teachers developed plan to implement these improvements.	Coaches (selected among “top performers” and experienced teachers) provided feedback every month. Based on coaches’ observations of teachers’ in their classroom, they developed improvement plans. Discussions between coaches and teachers also took place following the plans.	Teachers were offered 12 hours of workshop sessions, as well as one in-class support session per month. All sessions included three steps: planning discussion, supporting the teacher in the classroom, and debriefing and goal setting.	Sessions focused on pedagogical skills such as using precise language, introducing new concepts, and understanding how children learn conducted by trained tutors. After the training, monthly class visits occurred to observe how teachers engage with students and reflect on how to improve they practice.	A two-week training to present the program theory and three subsequent workshops, each one month apart.
Summary of results	Teachers increased their pedagogical skills which increased student learning.	1) Teachers increased their pedagogical skills which increased student learning after 1 year. 2) Despite the high cost of coaching, the program showed long-lasting effects on teachers' skills.	1) Teachers used significantly less violence against children, leading to a better classroom environment. 2) Children improved their early learning skills, especially oral language and self-regulation skills.	Improvements in student outcomes were attributed to teachers’ improved pedagogy. Students were more engaged and inquisitive. Teachers had more desire to learn.	Increased achievement of all students and improved self-concept of students with low and med performance in the entrance examination. Low-expectation students benefited the most.

Table A3: Examples of multicomponent professional development programs with positive and precise effects on student test scores

Study	Beg et al., 2022 – Pakistan	Jukes et al., 2017 – Kenya	He et al., 2008 – India	Fuje & Tandon, 2018 – Mongolia	Banerjee et al., 2016 – India
Context	Setting: Not specified School level: Secondary Grade: 6 and 8 Scale: Regional	Setting: Rural School level: Primary Grade: 1 Scale: Local	Setting: Rural and Urban School level: Primary Grade: 2 and 3 Scale: Regional	Setting: Rural School level: Primary Grade: Not specified Scale: Regional	Setting: Rural and Urban School level: Primary Grade: 1 to 5 Scale: Regional
Professional development program	1) eLearn Classrooms: teachers received a preloaded tablet, and they could project the content to a project-installed LED screen. 2) eLearn Tablets: students received individual tablets with math and science content, but only science teachers received tablets. Teachers could not display the content to the class. In both intervention arms, teachers were given video lectures on a math and scientific concept (less than 30 hours of content).	1) Teacher training to improve literacy 2) Malaria screening and treatment program	1) Machines arm: Interactive educational machines to be individually used by students. 2) Activities arm: Activities designed using 440 flashcards and teachers' manuals with recommended drills. The educational content was the same across "Machines" and "Activities".	1) Provision of children books 2) Teachers' training to improve their skills to support students in math, reading, and writing activities	1) Summer Camps: provide remedial instruction to academically weak children 2) TM: Teacher training and Materials 3) TMV: Training (on how to improve basic reading and arithmetic), Materials, and Volunteering Support. 4) TaRL: Teaching at the Right Level and government monitoring
Program length (exposure)	3-4 months	2 school years	1 school year	1 school year	1) 1 month 2) 2 school years 3) 2 school years 4) 1 school year
Training length	2 days plus 30 hours of video lectures	2 school years	5 days plus regular contact with program monitors	1 week, under cascade model	Not specified
Frequency of training and implementation	In both intervention: one day on orientation to the new technologies and another day on how to combine classroom teaching with technology-enabled multimedia content.	A 3-day workshop with guided opportunities to create new instructional materials, a problem-solving workshop 4 months after the start of the school year, and a refresher training the following school year. Ongoing support for teachers for two years through weekly text messages providing brief instructional tips and motivation.	School teachers and teaching assistants received a 5-day training session to learn how to implement the program. Schools also had regular access to external program monitors to assist teachers. Pratham, the program developer, provided assistance to teachers, but could not mandate the implementation of the curriculum as designed.	Cascade model: mentor teachers were trained for 4 days by well-qualified national trainers and conducted 2 visits per school to mentor fellow teachers. Teachers received a 3 day training session.	Details of the training were not specified. For TM, and TMV classrooms were never organized around initial learning levels. In TaRL, the allocation of a specific hour of the day facilitated the organization on students by learning levels.
Summary of results	1) eLearn Classrooms increased student achievement. 2) eLearn Tablets decreased student achievement on combined math and science exams designed for the projects.	Teachers used written text more often and focused more on letters and sounds. Student scores increased in word reading and oral reading fluency after two years. School dropout was also reduced. Effects were greater on girls than boys. Anti-malaria intervention did not affect educational achievement. The authors emphasized that this type of training was low cost and can be expanded.	Overall higher students' test scores from the teacher training program. Weaker students benefited more when teachers directed "Activities", while stronger students benefited more from "Machines".	1) Provision of books with teacher training raised students' achievement. 2) When provided alone, teacher training and books were not effective.	1) Summer camp, TMV and TaRL interventions significantly improved both language and math scores. 2) TM interventions had no effect.

© Deepa Srikantaiah / GPE

What is a rapid evidence assessment?

A rapid evidence assessment is a targeted systematic review. Similar to a systematic review, it uses a systematic approach to search and screen studies for inclusion. To make it rapid, the search strategy could be limited to certain databases and the scope may be narrowed to focus only on a few intervention types.

About the rapid evidence assessment

This brief summarizes the findings of a rapid evidence assessment prepared for the government of the Philippines as part of 3ie's Philippines Evidence Program, funded by Australia's Department of Foreign Affairs and Trade. Authors of this brief include findings from 101 evaluations of teacher professional development programs. A protocol was developed for this REA, which is available upon request.

About this brief

This brief was developed with contributions from Constanza Gonzalez Parrao (3ie), Romeo Arahan (3ie), Pierre Marion (3ie), Sanghwa Lee (3ie), Birte Snilstveit (3ie), Laurenz Mahanta-Langer (South Africa Centre for Evidence – SACE), Promise Mucharesva Nduku (SACE), Andile Madonsela (SACE), Tanya Mdlalose (SACE), John Ategeka (SACE), and Chris Cooper (University College London). The contents are the responsibility of the 3ie and do not necessarily reflect the views of the Philippines government or the Australian government. This brief was designed and produced by Akarsh Gupta and Tanvi Lal.

The International Initiative for Impact Evaluation (3ie) develops evidence on how to effectively transform the lives of the poor in low- and middle-income countries. Established in 2008, we offer comprehensive support and a diversity of approaches to achieve development goals by producing, synthesizing and promoting the uptake of impact evaluation evidence. We work closely with governments, foundations, NGOs, development institutions and research organizations to address their decision-making needs. With offices in Washington DC, New Delhi and London and a global network of leading researchers, we offer deep expertise across our extensive menu of evaluation services.

For more information on the 3ie's Rapid evidence assessment brief, contact **info@3ieimpact.org** or visit our site web.

3ieimpact.org

January 2023

@3ieNews /3ieimpact 3ieimpact /company/3ieimpact /3ievideos