October 2023

Numeracy at Scale: Final Report

Photo credit: JICA TAFITA, Madagascar

RTI INTERNATIONAL

RTI Project Number
0216840.000.000

Numeracy at Scale: Final Report

October 2023

Prepared for

Center for Global Development
1 Abbey Gardens
Westminster, London
SW1P 3SE

Prepared by

Jonathan Stern, Rachel Jordan, Yasmin Sitabkhan, Wendi Ralaingita, Bidemi Carrol, Kellie Betts, Matthew Jukes, Julianne Norman, Peggy Dubeck, Joseph DeStefano, Jessica Mejia, Laiba Bahrawar, Chris Cummiskey, Nameen Jalloh, Jonathan Pamel, Robert Momanyi

RTI International
3040 E. Cornwallis Road
Research Triangle Park, NC 27709
USA

CONTENTS

1 Executive Summary 1

1.1 Introduction 1
1.2 Research Questions 1
1.3 Findings 2

1.3.1 Overview of programs 2
1.3.2 High-level analyses 4
1.3.3 Instructional practice findings 4
1.3.4 Instructional support findings 5
1.3.5 Systems findings 6

1.4 Recommendations and Considerations 6

1.4.1 Pedagogy 7
1.4.2 Instructional support 7
1.4.3 Systems 7

2 Introduction 8

2.1 Background 8
2.2 Report Outline 8
2.3 Program Selection 8

2.3.1 Final selection 9

2.4 Research Questions 9

3 Research Design 11

3.1 Numeracy at Scale Research Methods 11

3.1.1 Hypotheses 11
3.1.2 Research methods 13
3.1.3 Data collection and analyses 15

3.2 Sample 15

3.2.1 ESMATE – El Salvador 16
3.2.2 R-Maths – South Africa 17
3.2.3 GKA – India 17
3.2.4 Nanhi Kali – India 18
3.2.5 RAMP – Jordan 19
3.2.6 TAFITA – Madagascar 20

3.3 Instruments 23

3.3.1 Mathematics instruments 23
3.3.2 Tool development 23
3.3.3 Tool piloting 25
3.3.4 Instrument translation and rendering 26

3.4 Study Limitations 27

4 Data Collection 28

4.1 Program Entry Visits 28
4.2 Data Collector Training 28

4.2.1 Ensuring data reliability and accuracy 29
4.3 Data Collection 30
4.3.1 School-level data collection: Quantitative 30
4.3.2 School-level data collection: Qualitative 30
4.3.3 Systems-lead data collection: Qualitative 30
4.4 Program Details 31
5 Findings 33
5.1 Program Overviews 33
5.1.1 ESMATE 33
5.1.2 R-Maths 35
5.1.3 GKA 37
5.1.4 Nanhi Kali 41
5.1.5 RAMP 45
5.1.6 TAFITA 50
5.2 High-Level Analysis: Program Matrix 54
5.2.1 Introduction 54
5.2.2 Most frequently implemented program elements 55
5.2.3 Program design 56
5.2.4 Key elements of Numeracy at Scale programs 58
5.3 School-Level Analysis 62
5.3.1 What classroom ingredients lead to learning in programs that are effective at scale? (research question 1) 62
5.3.2 What methods of training and support lead to teachers adopting effective classroom practices? (research question 2) 76
5.3.3 Teacher knowledge and attitudes 90
5.3.4 Findings on education technologies to support numeracy: Mindspark 91
5.4 Systems-Level Analysis 94
5.4.1 Coding and analysis process 94
5.4.2 Common characteristics across education systems 95
5.4.3 Challenges to scaling and sustaining interventions 102
5.4.4 Remediation and recovery efforts: COVID-19 104
5.4.5 Other keys to program success 105
6 Discussion and Recommendations 107
6.1 Pedagogy Considerations 109
6.2 Instructional Support Considerations 110
6.3 Systems Considerations 111
References 113
Annex A. Additional Tables from Interviews 115
A.1 Classroom Observation (Additional Items) 115
A.2 Coaching/Mentoring Interview (Additional Items) 118
A.3 Trainer Interview 125
A.4 Teacher Interview (Additional Items) 132
A 5. Head Teacher Interview 139
Annex B. Learning at Scale Program Interview Questions 147

Annex C. Final Program Selection......148

Annex D. Full Data on Program Elements and Key Elements for Program Implementation......150

Annex E. Classroom Observation Instruments developed for Numeracy at Scale......155

E.1 Quantitative Observation Tool 155
E.2 Qualitative Observation Tool 163

LIST OF FIGURES

Figure 1. Numeracy at Scale programs......1
Figure 2. Learning at Scale theory of change...... 11
Figure 3. Example of a causal link from the theory of change 14
Figure 4. ESMATE's theory of change 33
Figure 5. ESMATE's problem-solving approach...... 34
Figure 6. GKA's theory of change and program model...... 39
Figure 7. Key actors in Nanhi Kali's implementation...... 42
Figure 8. Operation of academic support centers...... 43
Figure 9. Mindspark math structure (grades 1–9)...... 45
Figure 10. RAMP internal evaluation results...... 47
Figure 11. RAMP's theory of change...... 48
Figure 12. TAFITA's approach 52
Figure 13. Numeracy at Scale domains, by program 57
Figure 14. Example of TAFITA session...... 64
Figure 15. Models supporting development of ESMATE place value concept 64
Figure 16. Observation: Who used materials during independent work? 65
Figure 17. Materials and strategies used in RAMP...... 65
Figure 18. Examples of teacher encouragement of multiple strategies and discussion of a problem 68
Figure 19. Percent of lessons in which the teacher connected math concepts to real-life examples or the lives of students, by program 69
Figure 20. Percentage of lessons in which all, most, or few students were engaged for the whole time, by program 69
Figure 21. Percentage of lessons in which each response type was observed: never, 1–10 times, or more than 10 times 70
Figure 22. Percent of lessons in which the teacher called on one student, a few of the same students, or a variety of students, by program 71
Figure 23. Tasks assigned during independent time, shown in percent of lessons where each task was observed, by program 72
Figure 24. A group solves problems using sticks, bundles, and place value chart 73
Figure 25. Teacher shows students how to play a number game and then ensures that all students get a turn...... 73
Figure 26. Most useful program supports reported by teachers, by program 76
Figure 27. Most useful training content areas, as reported by teachers, by program 77
Figure 28. Percentages of teachers reporting on the training methods that the program trainings employed more of compared to previous trainings...... 79
Figure 29. Most useful training methods, as reported by teachers, by program...... 80

Figure 30. Comparison of program teacher materials with prior materials, as reported by teachers, by program 80
Figure 31. Comparison of programs' student materials with prior materials, as reported by teachers: Average across programs 82
Figure 32. Frequency of coaching visits, as reported by teachers (percentage), by program 83
Figure 33. Differences between coaching before and during programs, as reported by teachers (average across programs) 84
Figure 34. Most helpful aspects of training on coaching, according to coaches, by program 86
Figure 35. Percent of teachers who report attending teacher meetings, by program 87
Figure 36. Most useful aspects of teacher meetings, according to teachers (average across programs) 87
Figure 37. Changes that head teachers have made because of participating in their respective program, as reported by the head teachers 89
Figure 38. How Mindspark works 92
Figure 39. Example of Mindspark teaching strategy 93
Figure 40. Student action after incorrectly answering a question 94

LIST OF TABLES

Table 1. Criteria for programs to be considered for inclusion in the Numeracy at Scale study 8
Table 2. Instruction hypotheses 12
Table 3. Training hypotheses 12
Table 4. Teacher support hypotheses 12
Table 5. System hypotheses 13
Table 6. Quantitative sample achieved in El Salvador (ESMATE) 16
Table 7. Qualitative sample achieved in El Salvador (ESMATE) 16
Table 8. System-level interviews in El Salvador (ESMATE) 16
Table 9. Quantitative sample achieved in South Africa (R-Maths) 17
Table 10. Qualitative sample achieved in South Africa (R-Maths) 17
Table 11. System-level interviews in South Africa (R-Maths) 17
Table 12. Quantitative sample achieved in Karnataka, India (GKA) 17
Table 13. Qualitative sample achieved in Karnataka, India (GKA) 17
Table 14. System-level interviews in Karnataka, India (GKA) 18
Table 15. School sample achieved in Andhra Pradesh, India (Nanhi Kali) 18
Table 16. Quantitative sample achieved in Jordan (RAMP) 19
Table 17. Qualitative sample achieved in Jordan (RAMP) 19
Table 18. System-level interviews in Jordan (RAMP) 19
Table 19. Quantitative sample achieved in Madagascar (TAFITA) 20
Table 20. Qualitative sample achieved in Madagascar (TAFITA) 20
Table 21. System-level interviews in Madagascar (TAFITA) 20
Table 22. Final instrument count, by program 22
Table 23. IRR results during piloting 26
Table 24. Translation needs, by program 26
Table 25. Program entry visits, in chronological order 28
Table 26. Results of trainee accuracy and reliability measures 29
Table 27. TAFITA (Madagascar) 31
Table 28. GKA (India) 31

Table 29. Nanhi Kali (India) 32
Table 30. RAMP (Jordan) 32
Table 31. ESMATE (El Salvador) 32
Table 32. R-Maths (South Africa) 32
Table 33. Most-implemented elements across domains 55
Table 34. Key program elements named by implementers, by domain 58
Table 35. Key program elements as described by Learning at Scale programs (minimum three programs each) 61
Table 36. Observation: Which materials were used during independent work? 63
Table 37. Percentage of students who used counters, drawings, or fingers to solve problems 66
Table 38. Teachers' responses to incorrect answers during whole-class instruction 67
Table 39. Teacher interview: How has your regular class instruction changed since you started working with this program? 67
Table 40. Independent work, by program 71
Table 41. Teacher interview: How has your regular class instruction changed since you started working with this program? 72
Table 42. Teachers' responses to students who finished an activity early, by program 74
Table 43. Teacher interview: What part of your instruction has had the biggest impact on student learning? 75
Table 44. Most important difference between program trainings and others, as reported by teachers, by program 78
Table 45. Helpful activities in a typical coaching visit, as reported by teachers, by program 84
Table 46. Topics covered in post-observation discussions with teachers, according to coaches, by program 85
Table 47. Mathematical knowledge for teaching 90
Table 48. Teacher attitude scales 90
Table 49. Role of ministry and other stakeholders, by program 95
Table A-1. Classroom Observations 115
Table A-2. Coaching/Mentoring Interview 118
Table A-3. Trainer Interview 125
Table A-4. Teacher Interview 132
Table A-5. Head Teacher Interview 139

LIST OF ABBREVIATIONS

ASC	academic support center [India]
ASER	Annual Status of Education Report
CAPS	Curriculum Assessment Policy Statement [South Africa]
COVID-19	coronavirus disease of 2019
DREN	Direction Régionale de l'Education Nationale [Madagascar]
EGMA	early grade mathematics assessment
EGRA	early grade reading assessment
ESMATE	Project for the Improvement of Mathematics Teaching in Primary and Secondary Education [El Salvador]
IRR	inter-rater reliability
GKA	Ganitha Kalika Andolana
JICA	Japan International Cooperation Agency
MKT	mathematical knowledge for teaching
n/a	not applicable
NCCD	National Centre for Curriculum Development [Jordan]
NGO	nongovernmental organization
RAMP	Early Grade Reading and Mathematics Initiative [Jordan]
R-Maths	Grade R Mathematics and Language Improvement Programme [South Africa]
RTI	RTI International (registered trademark and trade name of Research Triangle Institute)
TAFITA	Tantsoroka ho an'ny Fitantananany sekoly [Madagascar]
TaRL	Teaching at the Right Level
USAID	United States Agency for International Development

ACKNOWLEDGMENTS

The Numeracy at Scale research study team is grateful to the many organizations, experts, government officials, and teachers who were willing to share their experiences working to improve math outcomes at scale. We are most grateful to our partner programs and the individuals within those programs who did so much in support of this study, including the Akshara Foundation team implementing GKA in India; the team at Naandi Foundation and the Education Initiatives team, both working to implement Nanhi Kali using Mindspark in India; the ESMATE and Japan International Cooperation Agency teams in El Salvador; the RAMP team in Jordan; the R-Maths team and curriculum team in the Western Cape Education Department; and the TAFITA team in Madagascar.

Spanning six programs in five countries and three continents, this research was a formidable undertaking that would not have been possible without the expertise, long hours, and dedication of our in-country research teams, including Education NGO (Madagascar), Decipher Data (South Africa), Dajani Consulting (Jordan), and FEDISAL (El Salvador). We are grateful to the 100+ researchers across these countries who dedicated late hours and traveled long distances to ensure that quality observation and interview data were collected.

This research would never have taken place without the experienced, guiding hands of our in-country research experts: Arleen Oviedo, Paul Randrianina, Philip Browne, and Ankush Singh. Thank you for all your hard work and strategic problem solving.

We are grateful to the thought leaders, advisors, and experts who have helped guide this research study, including Benjamin Piper and Asyia Kazmi at the Bill and Melinda Gates Foundation; Justin Sandefur, Susannah Hares, Rita Perakis, Dave Evans, Christelle Saintis-Miller, and Ayesha Khan at the Center for Global Development; the Learning at Scale advisory committee members; and expert consultants Christine Van Keuren, Norma Evans, George Evans, Shirin Lutfeali, Dr. Minahil Asim, Jennifer Collette, Sheetal Digari, and Nicole Urbelis.

Lastly, we know that research has little impact unless it is communicated effectively. For this, we owe a great deal to Alice Cornish and the Better Purpose team for leading a thoughtful and targeted dissemination strategy, as well as to our resident wordsmith Morgan Stoffregen (editing), graphic artist Stefan Peterson, and formatting mastermind Felice Sinno-Lai.

CONFLICT OF INTEREST DISCLOSURE

The Learning at Scale research study is led by RTI International and is part of the Center for Global Development education research consortium, funded by the Bill and Melinda Gates Foundation. RTI International is also the lead implementer of one of the programs (RAMP in Jordan) selected for inclusion in the second phase of this study: Numeracy at Scale.

From the outset of this work, RTI took steps to mitigate against conflicts of interest. The first Learning at Scale deliverable for our client was a document outlining mechanisms to mitigate against conflicts of interest throughout the study.

RTI adopted a transparent process for developing research questions and methodologies and for selecting partner programs. All proposed questions and methodologies were reviewed by an independent advisory board. RTI presented two different methods for using selection criteria and presented the results to the advisory board (with programs anonymized) during the first phase of Learning at Scale. The final selection of the Learning at Scale programs was determined by this independent advisory committee. For the Numeracy at Scale phase, selection criteria were adjusted slightly and proposed to the Center for Global Development for review and feedback. RTI conducted the final selection process independently.

RTI also took steps to ensure that data from this study were not used for any purpose outside Learning at Scale research. We contracted a consultant to work on all cost data and analyses. The study team has also ensured that non-public data from the study are not accessible to RTI staff who are not part of the Learning at Scale research team. No business development staff or non-Learning at Scale team members have access to these data or files. We have also developed and are following guidelines for business development and proposal purposes, which lay out the personnel, documents, and data related to this research project that cannot be used for these purposes.

Where Learning at Scale reports and other outputs address the work of RTI programs, we have taken steps to counter subjectivity, including questioning positive perceptions of RTI programs and retaining data and findings that may show our programs in a negative light, making all analytic decisions based on data.

Ultimately, RTI as a whole has made a commitment to objectivity in the Learning at Scale study and to using data for their intended purposes only. We all take this commitment very seriously.

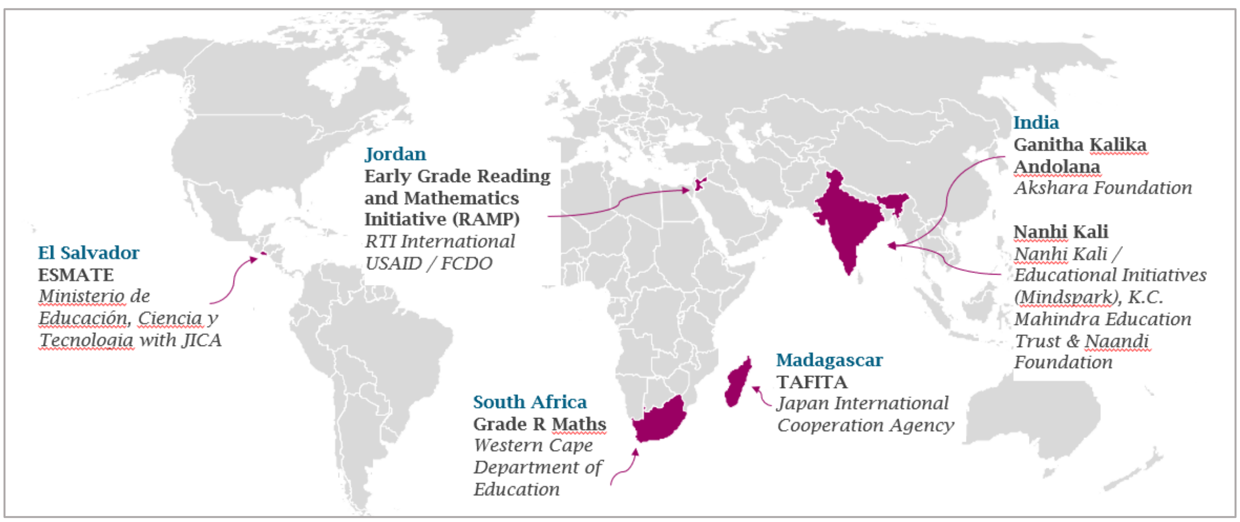
1 EXECUTIVE SUMMARY

1.1 Introduction

The Learning at Scale study was designed to explore programs that have a demonstrated impact on foundational learning outcomes at scale. The goal of this research is to identify and examine successful aspects of these programs to provide policy makers and development practitioners with evidence-based strategies for improving instruction and learning outcomes across contexts. The research is being led by RTI International and is part of the Center for Global Development education research consortium, funded by the Bill and Melinda Gates Foundation.

While the first phase of Learning at Scale focused on literacy, the second phase, Numeracy at Scale, is focused on (1) identifying instructional strategies that are essential for improving math outcomes at scale in low- and middle-income countries; and (2) learning about the characteristics of the education systems within which successful scaled-up numeracy programs operate. To this end, the study team identified and analyzed six programs across five countries that had rigorous evidence of impact on math learning outcomes and which were operating at scale or which showed the potential for scale in an entire region or country (see **Figure 1**).

Figure 1. Numeracy at Scale programs



1.2 Research Questions

The aim of the Numeracy at Scale study was to address the following research questions.

1. What classroom ingredients (e.g., teaching practices, classroom environment) lead to learning in programs that are effective at scale?
2. What methods of training and support lead to teachers adopting effective classroom practices?

3. What system support is required to deliver effective training and support to teachers and to promote effective classroom practices?

To address these questions, we developed a theory of change, generalizable across programs. The theory of change drew on causal relationships among actors in the education system or the program. For each causal relationship, we developed a set of hypotheses to test in data collection. We tested these hypotheses using quantitative and qualitative data that describe the nature of the causal relationship. In addition, we used a quasi-experimental design to identify the impacts of each program on teacher behavior and classroom activities. Our theory of change can be found in **Section 3** of the report.

1.3 Findings

1.3.1 Overview of programs

Three programs in this study have impact evidence demonstrating effectiveness at scale: GKA in India, RAMP in Jordan, and R-Maths in South Africa.

- Evidence on the impact of GKA on student outcomes draws from two studies. An internally commissioned longitudinal study (Vaijayanti et al., 2016) found a significant impact on the GKA implementation group after two years of implementation in grade 3 (ES 0.43), grade 4 (ES 0.27) and grade 5 (ES 0.34). An external experimental evaluation conducted over 15 months by the Abdul Latif Jameel Poverty Action Lab (deBarros et al., 2023) found a significant positive impact only on girls' math outcomes (0.18 SD) at endline. GKA has been rolled out, through the State Education Department, to all 45,000 schools in Karnataka State and to most of the government schools in Odisha State.
- Since RAMP was a national-scale program from the outset, operating in 2,970 primary schools, it was not possible to identify a control group for the program's impact evaluation. Therefore, evidence of impact was obtained from an internal evaluation using a pre-post design. The baseline sample consisted of control schools from a pre-RAMP intervention, with data collected in May 2014, and included 240 schools across all 12 governorates in the county. At the initial endline, the program had an estimated 9–10 percentage point impact in mathematics performance in grade 2 and grade 3 (i.e., a 100% increase in grade 2 and a 50% increase in grade 3). After a reduction in performance due to the COVID-19 pandemic, a second endline evaluation (in 2023) showed that scores increased even further, with an overall impact of 15.5 percentage points in grade 3 and 11.2 percentage points in grade 2.
- The impact of R-Maths on students' mathematical knowledge in the Western Cape of South Africa was assessed through a quasi-experimental evaluation comparing schools that had received the program with those waiting to receive the program. This evaluation was conducted with 622 grade R students in 101 urban schools and 51 rural schools. Among rural schools, students in the intervention group had improved mathematics scores compared to the control group, with an effect size of around 0.20. There was no significant difference between the intervention and control group in the urban schools. Implementers felt that the small effect size found in the study was encouraging for the first year of rollout, with the expectation that the impact would increase as the program continued to implement in all eight districts of the Western Cape, comprising nearly 1,000 schools.

Two other programs—TAFITA in Madagascar and ESMATE in El Salvador—detected evidence of impacts during a pilot phase and continue to implement in these schools, while scaling a similar program design to additional schools.

- The Japan International Cooperation Agency (JICA) undertook a randomized controlled trial within the first phase of the TAFITA program, within the Amoron'i Mania region, during the 2018–2019 school year (Maruyama & Igei, Forthcoming). Researchers randomly sampled 140 schools of the 1,002 public primary schools in the region. The baseline was conducted in November–December 2018, and the endline took place in September 2019. Endline math assessment scores showed that the program improved learning for all the targeted grades. The magnitude of impact is largest in grade 3 students (0.47 standard deviations), followed by grades 4 (0.38 standard deviations), and 5 (0.36 standard deviations). All impacts are statistically significant at a 99% confidence interval ($p < 0.01$). The math component of TAFITA is now being implemented at scale in two regions of Madagascar, reaching 2,725 public primary schools and 288,896 students in grades 2 through 5.
- From 2018 to 2019, ESMATE consisted of a pilot project, implemented with support from JICA, to evaluate the impact of the newly created student textbooks, paired with teacher training and ongoing classroom support by the ESMATE team. The RCT was conducted in 250 public schools in four departments. After two years of instruction, 2^{nd} grade students in year 1 in the treatment group showed improved scores on math outcomes (estimated at 0.49 standard deviations after year 1, and 0.13 standard deviations in year 2 compared to students in the control group who only received one year of exposure) (Maruyama & Kurosaki, Forthcoming). After these successful pilot results, in 2020 ESMATE was scaled up to all of the country's public schools—namely, 4,666 primary schools, 2,726 junior high schools, and 705 senior high schools.

Lastly, evidence of impact for Nanhi Kali is specific to the Mindspark application, developed by Education Initiatives to deliver tablet-based math instruction tailored to students. An impact evaluation of Mindspark was conducted by the Abdul Latif Poverty Action Lab in 2015 and 2016 to assess the application's impact on students' mathematical knowledge by comparing the baseline and endline scores of 619 students recruited from five public middle schools in Delhi. Students who received the Mindspark intervention scored 0.37 standard deviations higher than the control group at endline and improved their scores by more than double the control group over the intervention period. While the Mindspark experiment was conducted separately from Nanhi Kali, it is critical to understand the design of Nanhi Kali given that the program is a vehicle for delivering this software at scale—there are currently 6100 Nanhi Kali centers (all using Mindspark) across eight regions in India.

Although all six of the programs selected for inclusion in this study had evidence of improved math outcomes for early grade students, their designs and overall approaches were varied. For example, four of the programs focused on core instruction in public schools (ESMATE, GKA, R-Maths, and RAMP), while two focused on remediation or alternative learning opportunities (Nanhi Kali and TAFITA). The grades ranged from kindergarten only (R-Maths) to grades 4 and 5 only (GKA). Instruction in most programs was provided by teachers (ESMATE, GKA, R-Maths, RAMP, and TAFITA), but in Nanhi Kali the core instruction was provided by tablet-based software and was supported by community volunteers. Further, program funding came from a variety of sources, including bilateral donors, local foundations, and governments. Two of the programs of study (ESMATE and R-Maths) are now entirely led, implemented, and funded by the government.

1.3.2 High-level analyses

The availability of program design data, provided by each partner, allowed us to understand how the Numeracy at Scale interventions worked across the five domains of program characteristics that we identified: *Materials*, *Pedagogy*, *Training*, *Teacher Support*, and *Systems*. We found that the program design structures shared some elements and that all included aspects of these five areas. While each program was unique in what it saw as critical to its success, we found 14 elements that were determined to be key for three or more programs and suggest that future interventions consider these elements as essential for program impact. These "essential elements" emphasize meeting students where they are and using different models and representations—including while students practice with one another—to support conceptual understanding in math. The elements also emphasize close alignment with the government, especially around curriculum, monitoring, and coaching. As part of a minimum package, these elements also point to the value of in-person training over multiple points in time. The 14 essential program elements are as follows:

- Learning aids for students (e.g., counters, number cards, place value materials, etc.)
- Program materials aligned to government curriculum
- Structured teacher's guides (scripted lessons)
- Continuous and formative assessment
- Instruction targeted to student level (differentiated instruction)
- Focus on developing conceptual understanding
- Pair work or group work
- Using concrete materials and resources (manipulatives)
- Coaches who are government staff
- Initial face-to-face training
- Refresher face-to-face training
- Teacher training (lowest level of cascade) done by government officers
- Government staff responsible for conducting monitoring
- Program investment in capacity building at the decentralized level

1.3.3 Instructional practice findings

Analyses from qualitative and quantitative classroom observations point to a number of common instructional themes across program classrooms:

- **Teachers use multiple representations and models to support learning.** Across the four countries where both quantitative and qualitative observations were done, teachers were observed explicitly linking representations of a concept. In all programs, students used the concrete materials and other models themselves rather than simply observing the teacher doing so.

- **Instructional approaches include a specific focus on both conceptual understanding and procedural fluency.** In TAFITA, RAMP, and ESMATE, teachers asked at least some open questions where there was not just one correct answer. When students gave a wrong answer, teachers in three of the five programs where observations were done were more likely to help the student find the correct answer or discuss why the answer was incorrect. In the majority of programs, there was an emphasis on encouraging students to use multiple strategies in problem solving and, in some programs, to discuss their mathematical ideas. In addition, across all programs, teachers frequently made connections between math concepts and either the real world or students' experiences, which can support students' conceptual understanding as well as their ability to apply mathematics concepts to novel and real-world situations.
- **Various approaches are used to ensure active student engagement throughout lessons.** In virtually all observations, the majority of students remained engaged for the duration of the lesson, and often, all students stayed engaged throughout. Teachers across programs used a variety of questioning techniques. While opportunities for individual students to respond were prevalent across programs, teachers often called on a few of the same students to solve problems or answer questions, which reduces opportunities for engagement. In the majority of ESMATE and TAFITA lessons, however, teachers called on a variety of students throughout. While the level of engagement represented through questioning varied, all programs included some focus on having dedicated time for independent and group work. In all programs, a substantial amount of independent or group work time was spent in active learning or solving problems, which includes using manipulatives, playing a math game, measuring or cutting out and manipulating shapes, etc.
- **Teachers use assessment-informed instruction approaches to address differentiated needs.** One strategy that was used in all programs was for the teacher to monitor students while they worked during independent and group work. In qualitative observations, teachers in ESMATE, TAFITA, and RAMP were observed at some point helping students who appeared to be struggling. The use of formative assessments, whether formal or informal, was also found across programs, but the form and use of results varied. Some of the programs included a focus on assessment and use of results to teach students according to their level.

1.3.4 Instructional support findings

Our investigation of the instructional support provided to teachers identified the following themes across programs:

- **Teacher supports focus explicitly on math content and improving instruction.** While the design of teacher support models varied greatly across programs, all prioritized building teachers' pedagogical math knowledge. Math instruction and lesson plan development were deemed the most useful training content by teachers in four of the six programs.
- **Trainings emphasize modeling and practice over lecturing, providing teachers with opportunities to practice and discuss.** Teachers across all programs reported that training sessions used more modeling and demonstration, small-group practice, and

discussion than previous teacher training sessions they had attended. When asked which of these training methods was *most* useful, over one-half of teachers in four of the programs (GKA, RAMP, R-Maths, and TAFITA) said modeling and demonstration. The second most popular response was small-group practice, cited most often by teachers in ESMATE and Nanhi Kali.

- **Teacher and student materials provide explicit guidance for instruction.** Generally, teachers across all programs reported that teacher materials were organized and easy to follow; included ample activities and examples; and included engaging manipulatives. Compared with previous materials, teachers most often reported that the student materials they received under these programs were newer and more attractive and that the content presented was easier to follow and more clearly aligned to the curriculum and context.
- **Ongoing support emphasizes feedback, problem solving, and learning new content over inspection and evaluation.** There was some variation in how teachers were supported (including through teacher meetings, coaching, mentoring, and monitoring visits), but what this support focused on, and what teachers found to be most helpful, was similar: Teachers had opportunities to get feedback, solve problems, and learn new content. Additionally, the individuals who provided professional development were more supportive and friendlier compared to earlier programs.

1.3.5 Systems findings

Our analysis found all programs engaging at the systems level in various ways:

- **Programs actively collaborate with key stakeholders.** Collaboration among stakeholders (government officials, school staff, donors, external implementers, and school communities) was essential. This collaboration took place through formal channels (such as steering committees) and informal channels (such as WhatsApp groups). The roles and responsibilities of the various stakeholders were usually well understood.
- **Investments are made in resources to improve quality classroom instruction.** Programs invested heavily in the professional development of education actors such as teachers, instructional coaches, and school management committees. They also invested heavily in providing sufficient quantities of essential teaching and learning materials such as math kits, textbooks, tablets, and teacher’s guides.
- **Programs emphasize continuous monitoring and use of data for system improvement.** Programs emphasized the importance of collecting and using data for system improvement. They collected data on teacher practices and student math outcomes. The data were then used by various education actors to determine which students needed remedial or targeted support, content areas for teacher training, and where teaching and learning materials were needed, among other things.
- **Programs focus on systematically embedding and institutionalizing best practices, with an eye toward sustainability.** Sustainability was a priority for all programs. The programs were strategically designed with sustainability in mind, aligned their activities to the government’s goals and objectives, and advocated for new policies supportive to improving learning.

1.4 Recommendations and Considerations

1.4.1 Pedagogy

- There is no one prescribed way to improve math instruction.
- Involving all students in modeling and explanation is important.
- Students need time to practice.
- It's not enough to just have manipulatives.
- There must be a strong link between concrete materials, pictures or drawings, and abstract symbols.
- Assessing students is not enough; teachers need to know how to use that information to inform their instruction.

1.4.2 Instructional support

- Ensuring that classrooms have high-quality teaching and learning materials will make teachers' lives easier.
- Just as students need time to practice new skills in the classroom, teachers need time to practice new skills during training and teacher meetings.
- Developing teachers' mathematical knowledge for teaching (MKT) through training and support should be a constant focus.
- Ensuring that teacher supports are complementary can help overcome financial, logistical, and capacity limitations.
- Making coaches and mentors friendlier and more supportive is just the beginning.

1.4.3 Systems

- Data and evidence should be used to inform decisions.
- The education system should focus on professional development for education staff.
- Designing for scale and implementing through government systems are necessary steps for achieving improved learning outcomes at scale.
- Governments and external funders should invest strategically in resources to bolster quality instruction.

2 INTRODUCTION

2.1 Background

The Learning at Scale study was designed to explore programs that have a demonstrated impact on foundational learning outcomes at scale. The goal of this research is to identify and examine successful aspects of these programs to provide policy makers and development practitioners with evidence-based strategies for improving instruction and learning outcomes across contexts. The research is being led by RTI International and is part of the Center for Global Development education research consortium, funded by the Bill and Melinda Gates Foundation.

While the first phase of Learning at Scale focused on literacy, the second phase, Numeracy at Scale, is focused on (1) identifying instructional strategies that are essential for improving math outcomes at scale in low- and middle-income countries; and (2) learning about the characteristics of the education systems within which successful scaled-up numeracy programs operate. To this end, the study team identified and analyzed six programs across five countries that had rigorous evidence of impact on math learning outcomes and which were operating at scale or which showed the potential for scale in an entire region or country.

2.2 Report Outline

After this introductory section, the report is organized as follows: **Section 3** covers the research design, with attention to the development of new, math-specific instruments. **Section 4** describes how the research team prepared for and carried out data collection. **Section 5** discusses the findings, with subsections that elaborate on the features of each of the six selected programs; analyze the various characteristics found in each program; perform quantitative and qualitative analyses of cross-program instructional approaches; and offer qualitative analyses of how the education system is either implementing or supporting each program. **Section 6** ties together the findings and offers recommendations based on them. Throughout this report, we refer to approaches and tools developed under Phase 1 of the Learning at Scale study. The final report from this first phase, along with accompanying findings briefs, can be found at www.learningatscale.net.

2.3 Program Selection

Since program selection criteria were already developed under the first phase of Learning at Scale, the second-phase criteria required only slight modifications (see **Table 1**). The goal of the work was to identify four numeracy programs—led by either the government or a nongovernmental organization (NGO)—as well as two government-led programs (focused on any subject). While it was theoretically possible for there to be overlap, the intent was to identify six total programs for inclusion.

Table 1. Criteria for programs to be considered for inclusion in the Numeracy at Scale study

	Numeracy programs	Government-led programs
Unique criteria	▪ Improved effectiveness in teachers' math instruction, among other subjects	▪ Improved effectiveness in teachers' literacy or math instruction, among other subjects

	Numeracy programs	Government-led programs
	▪ Can be led by governments or NGOs	▪ Implementation is led directly by government entities
Combined criteria	▪ Implemented in the public sector, private sector, or civil society in low- and middle-income countries ▪ Active through at least 2022 ▪ Has local demand ▪ Covers at least 500 schools (and ideally operates in at least two administrative subdivisions) ▪ Has evidence of causal impact at scale *or* evidence of causal impact of a pilot study that has been effectively scaled	

The selection process for numeracy programs began with independently researching potential programs for inclusion in late 2020, followed by a call for programs on the Center for Global Development's website in early 2021. After the call for programs was posted, the Learning at Scale team began discussions with prospective programs. The team ultimately contacted more than 60 organizations, including NGOs, foundations, bilateral agencies, ministries of education, universities, and think-tanks.

For those programs that were interested in possible inclusion in the study, a structured call was set up with a Learning at Scale team member. The interview focused on gathering information about the program's approach, scale, impact, and integration into the education system in which it was functioning (see **Annex B** for interview questions).

Ultimately, initial and follow-up discussions led to a list of 28 programs that were considered for inclusion in this phase of Learning at Scale. Of these, eight were viable options to consider as numeracy programs, while three were viable options for government-led programs.

2.3.1 Final selection

The final selection of programs included four numeracy programs and two government-led initiatives (both of which also happened to be numeracy programs). Three programs had evidence of effectiveness at scale: RAMP in Jordan, GKA in India, and R-Maths in South Africa's Western Cape (government led). Two programs had evidence of effectiveness at a smaller scale but were continuing to implement the same approach in pilot schools and in scaled schools: TAFITA in Madagascar and ESMATE in El Salvador (government led). The final program, Nanhi Kali in India, was selected due in large part to its unique approach and the fact that scaling does not impact the instructional approach (since it is entirely tablet based). However, an additional caveat is necessary: the Nanhi Kali program (for which the Mindspark software is used as the instructional approach) does not have prior rigorous evidence of effectiveness. The final programs and their selection criteria are presented in **Annex C**.

2.4 Research Questions

The aim of the Numeracy at Scale study was to address three overarching research questions and two sub-questions. The first two overarching questions were focused on understanding the components of effective instruction, and the third was targeted toward understanding the system-level support that leads to effective instruction. The two sub-questions sought to gain a deeper understanding of what math instruction looks like in the classroom and how universal best practices are designed and implemented in different contexts.

1. What classroom ingredients (e.g., teaching practices, classroom environment) lead to learning in programs that are effective at scale?
2. What methods of training and support lead to teachers adopting effective classroom practices?
3. What system support is required to deliver effective training and support to teachers and to promote effective classroom practices?

   Sub-question 1: In what ways are materials or teachers emphasizing conceptual math learning? What strategies do they use to build conceptual understanding in mathematics? To what extent are they successful in building conceptual knowledge in mathematics?

   Sub-question 2: In what ways are materials or teachers emphasizing use of hands-on activities and appropriate representations and models? How do they support teachers in using them? How successful are they in the classroom?

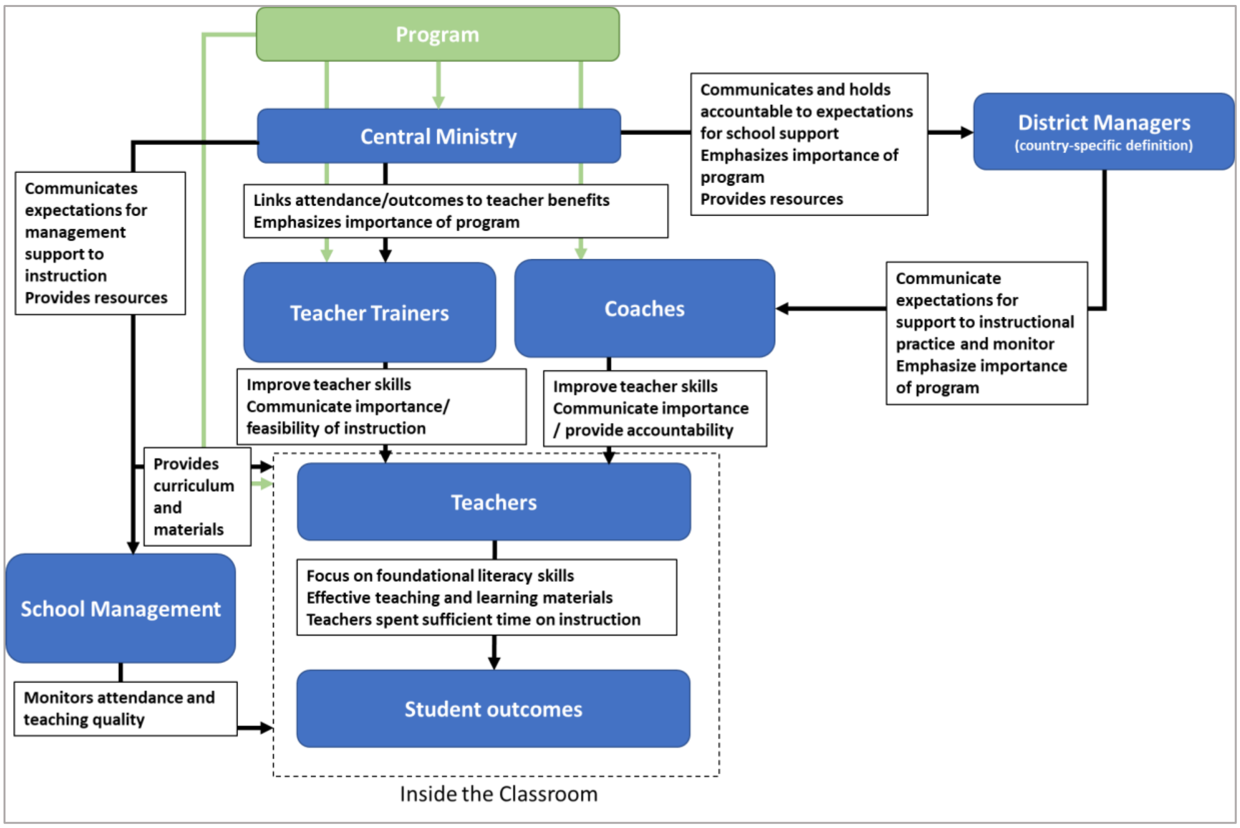
We addressed these research questions through primary data collection, including classroom observations and interviews with teachers, head teachers, trainers, coaches, teacher meeting facilitators, district officials, central Ministry of Education officials, and program staff.

3 RESEARCH DESIGN

3.1 Numeracy at Scale Research Methods

In order to address the main research questions, we returned to the theory of change developed under the first phase of the Learning at Scale study. **Figure 2** shows our generalized theory of change for how education systems achieve learning at scale. Each arrow represents a causal effect of one actor (in blue boxes) on another. The causal effects are labeled (in white boxes) with key mechanisms. The theory of change is designed to be universally applicable at a conceptual level but with country-specific instantiations.

Figure 2. Learning at Scale theory of change



Four of the programs we examined were funded by donors through NGOs, and two of the programs were initiated by the host government. To cater to these differences, the theory of change recognizes that each program may improve learning by working entirely within the education system or by operating in parallel with it. The green arrows represent ways in which programs implemented by NGOs may work in parallel with the education system rather than through it.

NOTE ON TERMINOLOGY

We recognize that in different contexts, "pupils," "learners," and other terms may be more accurate for describing children in early primary grades. For the sake of consistency, we use the term "student" throughout this report.

3.1.1 Hypotheses

The design of research methods is driven by a core set of hypotheses about how education systems achieve improved learning at scale. The hypotheses were first identified by the research team under Phase 1 of this study, based on recent literature on improving learning at a systems level. The instruction hypotheses were then revised to reflect the nuances of math instruction, and the system hypotheses were reviewed and refined based on initial findings from Phase 1. **Tables 2 through 5** set out the hypotheses that drove the Numeracy at Scale study at different levels, from classrooms to systems.

Table 2. Instruction hypotheses

1.	Instruction emphasizes conceptual knowledge and procedural fluency, which
	a. uses multiple representations to build understanding,
	b. encourages students' explanation and justification of concepts,
	c. follows clear trajectories of learning within domains, and
	d. creates real-life connections to enhance meaning.
2.	Teachers know the progress of more students because of increased interactions, informal monitoring, or assessments and adjust their instruction accordingly.
3.	Students have multiple opportunities to practice new skills and concepts.
4.	When manipulatives are used, they are in the hands of most or all students.
5.	Teachers provide clear models of new content.
6.	Teachers maximize instructional time.
7.	Teachers foster mathematical discussions.
8.	Teachers are motivated because they see how the instruction leads to positive student outcomes.

Table 3. Training hypotheses

1.	Practice: The training, its follow-up, and any peer support include time to practice the methods.
2.	Expectations: The goals of the training are clear and manageable.
3.	Collaboration: The interactions between trainer and teacher at the teacher training are positive and working toward a shared goal.
4.	Teachers are given clear directions (from training and materials) on how to do the lesson.
5.	Prioritization: A realistic amount is expected of teachers during training or coaching.
6.	The training improves teachers' procedural knowledge of effective instructional practices.
7.	The training improves teachers' self-efficacy in implementing effective instructional practices (that they may have already known).
8.	Teachers are inspired by the program.

Table 4. Teacher support hypotheses

1.	Instructional support staff improve or reinforce teachers' knowledge and skills.
2.	Instructional support staff provide support, motivation, and problem solving to teachers.
3.	Instructional support staff hold teachers accountable.
4.	Instructional support staff are provided basic resources (e.g., per diems, transport).

1. Instructional support staff improve or reinforce teachers' knowledge and skills.
5. The program or government gives incentives. Resources are structured in a way to incentivize better implementation of instructional support and more school visits.
6. The program or government makes instructional support staff accountable.
7. The job descriptions of instructional support staff reflect a focus on improving instruction and learning.
8. Instructional support staff believe in the program goals and purpose.
9. Instructional support staff are provided with training.
10. Instructional support staff are provided with tools to observe and give feedback.
11. Program or government provides support on how to more effectively provide feedback, guidance, motivation, and problem solving.

Table 5. System hypotheses

1. Key system actors are informed about the program.
2. System actors play substantive roles in implementation.
3. Expectations for system actors at all levels are specified.
4. System communicates expectations for districts, schools, teachers, and students.
5. System monitors performance relative to stated expectations.
6. Managers see subordinate parts of the system as having agency to solve or address problems.
7. Necessary inputs and resources are reliably made available.
8. System institutionalizes changes in policy, procedures, or practices as a result of the program.
9. System capacity in key technical areas is reinforced or developed by the program.
10. System actors can speak honestly about challenges faced in implementation.

3.1.2 Research methods

We used a variety of research methods, with different approaches to identifying causality, in addressing the research questions and investigating the hypotheses.

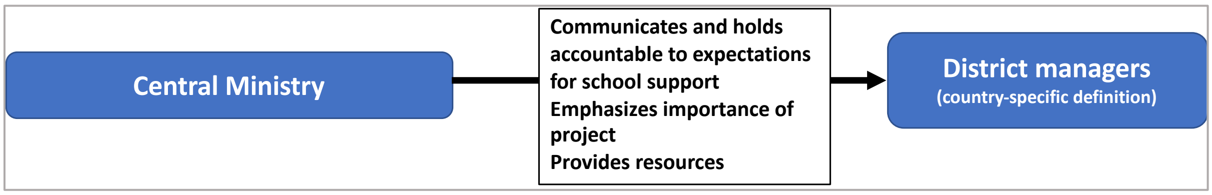
Quantitative methods to estimate the counterfactual

We first analyzed the impact evaluations reported by selected programs (see **Section 5.1** for an overview of these impacts). Where possible, we then used counterfactual reasoning to understand the causal effects of programs. In Madagascar, we used quasi-experimental methods to select a matching control group and compared school-level data between treatment and control groups. In El Salvador, where pilot and scaled schools were available, we selected samples of schools from both groups to determine if pilot schools were implementing with a greater level of fidelity than scaled schools.

Theory-based methods

Since the majority of programs did not have control groups available, the study also relied on theory-based methods to investigate the causal links between actors described in the theory of change. Our approach was based on the method of process tracing (Fairfield & Charman, 2017, 2019). Taking one causal link from the theory of change as an example (see **Figure 3**), we wanted to know whether—and how—the actions of the central ministry led to district managers performing their duties with greater fidelity (i.e., monitoring coaches).

Figure 3. Example of a causal link from the theory of change



We investigated this link by interviewing officials in the central ministry about the actions they took to communicate with district managers, hold them accountable, support them, and provide them with resources. We also interviewed district managers to inquire about their interactions with the central ministry. In both sets of interviews, we asked respondents not just about their actions but also about the impact of their actions. With this approach, we aimed to understand the actions of the ministry and their effects on the district, triangulated from two sources (i.e., both ends of the causal arrow). After interviews and collection of documentation, we assessed the strength of the evidence in support of the key mechanisms by which the central ministry influences the behavior of district managers.

For all hypotheses, we sought evidence both for and against the hypotheses. Some hypotheses were set up as alternate explanations for the same mechanism. In some cases, the competing explanations were compatible. For example, we had three hypotheses about the impact that coaching would have on teacher performance: improving or reinforcing teachers' knowledge and skills (coaching hypothesis #1 in **Table 4**), offering support and encouragement (hypothesis #2), and providing accountability to teachers (hypothesis #3). It is possible that coaches improved teacher performance by one or more of these mechanisms. In other cases, competing hypotheses were simply the inverse of stated hypotheses. For example, we had a hypothesis about how expectations were communicated to districts, schools, teachers, and students by the education system (hypothesis #4 in **Table 5**). Alternatively, the expectations may have been communicated by the program. In this case, stronger evidence for one hypothesis implied weaker evidence for the alternative.

System hypotheses were addressed using data from qualitative interviews and information from reports and documents. Hypotheses at the school level—relating to instruction, coaching, and training—were informed by both qualitative interviews and quantitative surveys.

Cross-case comparison

Lastly, we examined findings across programs. While all programs were studied independently, similar tools were used in order to ensure that comparability would be possible. If findings were repeated consistently across programs, we were able to draw stronger generalizable conclusions. For qualitative data, findings were mostly restricted to those supported by several programs.

3.1.3 Data collection and analyses

The above methods resulted in several types of data collection, with corresponding methods for analysis.

Quantitative school-level instructional data collection

The sample for this activity included 60 treatment schools and 30 comparison or pilot schools (where available). The instruments used for this activity included (where applicable) classroom observation, coaching observation, teacher meeting observation, MKT surveys, and stakeholder interviews (teacher, head teacher, coach, meeting facilitator, trainer). Through this activity, we aimed to gather information about the relationship between the implementation of successful numeracy program inputs at scale and improved instructional practice.

Program design mapping and interviews

A strength of this study is its replication across countries. Initial program design mapping and interviews showed patterns across programs' design, expected instructional outcomes, and definitions of key concepts. This enabled us to detect differences and similarities between programs' actual implementation vis-à-vis design, leading to conclusions about essential elements for effective math interventions. The instruments utilized for this activity consisted of program interviews and document reviews.

Qualitative school-level instructional data collection

Math interventions commonly use terms such as "conceptual understanding," "hands-on practice," and "doing math with understanding." However, there is little rigorous evidence on effective instructional models for math instruction and their implementation in low- and middle-income countries, and it is not clear what these terms look like in practice. By including multiple, consecutive instructional observations (where applicable), we sought to more clearly define and identify implementation challenges in an "essential early grades numeracy package." The sample for this activity consisted of ten schools per program, where applicable. The instruments used for this activity were cognitive interviews with students, qualitative lesson observation (used during three consecutive visits), and qualitative teacher interviews.

Qualitative systems-level interviews

Key actors from the education system provided information on programs' approaches to government engagement, communications, monitoring, capacity building, expectation management, and teacher support in order to achieve more effective math instruction at scale. This activity included a sample of 10–25 key informants for each program. The instruments used for this activity were central ministry interviews (with high-level policy makers, managers, inspectors, teacher education officials, and curriculum and materials development officers) and district- and subnational-level interviews (managers, inspectors, teacher education officials, and curriculum and materials development officers).

3.2 Sample

For each program, quantitative instructional data were collected from between 80 and 130 schools, qualitative instructional data were collected from between 9 and 24 schools, and qualitative systems-level interviews were conducted (for five programs) with between 6 and 22 key informants. Samples varied depending on how the study design was tailored to program implementation in each country.

For the quantitative school-level data collection, the study team drew a stratified multistage sample. First, with inputs from each program, we purposively selected between two and four subnational administrative units, which were deemed by the program to be strong implementation units. Within each administrative unit, we randomly selected between 10 and 25 schools, depending on the final sample design.

For the qualitative school-level data collection, the study team randomly selected ten schools from the quantitative sample.

For the systems-level data collection, the study team worked with program staff and in-country research consultants to identify key informants at the central level and subnational level of education administration.

The subsections below present sampling details for each program. **Table 22** at the end of this section presents final sample counts for each instrument.

3.2.1 ESMATE – El Salvador

Table 6. Quantitative sample achieved in El Salvador (ESMATE)

90 schools					
La Libertad		**San Miguel**		**Santa Ana**	
La Libertad	**Santa Tecla**	**Chirilagua**	**San Miguel**	**Chalchuapa**	**Santa Ana**
15 schools	15 schools	15 schools	15 schools	15 schools	15 schools

Table 7. Qualitative sample achieved in El Salvador (ESMATE)

10 schools
10 schools subsampled from La Libertad Department (2), San Miguel Department (4), and Santa Ana Department (4)

Table 8. System-level interviews in El Salvador (ESMATE)

Level	**Department or position**
Central ministry	▪ National director of curriculum ▪ Head of Mathematics Department ▪ ESMATE technical lead
District 1	▪ Pedagogical coaches (2)
Donor (JICA)	▪ Senior advisor for ESMATE

3.2.2 R-Maths – South Africa

Table 9. Quantitative sample achieved in South Africa (R-Maths)

80 schools			
Eden and Central Karoo	**Metro Central**	**Metro East**	**West Coast**
19 schools	21 schools	20 schools	20 schools

Table 10. Qualitative sample achieved in South Africa (R-Maths)

24 schools
Teachers in 24 schools received qualitative questions in addition to a quantitative interview

Table 11. System-level interviews in South Africa (R-Maths)

Level	Department or position
Provincial ministry	▪ Chief education specialist for grade R ▪ Deputy chief education specialist for grade R ▪ Former deputy chief education specialist for foundation phase mathematics ▪ Director of research and evaluation at Zenex Foundation (funders)
District 1	▪ Foundation phase coordinator ▪ Head of curriculum support in grade R
District 2	▪ Foundation phase coordinator
Cross-district	▪ Grade R subject advisor (supporting 4 districts)

3.2.3 GKA – India

Table 12. Quantitative sample achieved in Karnataka, India (GKA)

78 schools					
Bengaluru Rural	**Chamarajana gara**	**Chikkaballap ura**	**Chitradurga**	**Dharwad**	**Gadag**
17 schools	16 schools	6 schools	15 schools	12 school	12 schools

Table 13. Qualitative sample achieved in Karnataka, India (GKA)

11 schools		
4 schools subsampled from Dharwad District	**3 schools subsampled from Chitradurga District**	**4 schools subsampled from Gadag District**

11 schools		
1 grade 4 teacher, 3 classroom observations, and 5 grade 4 students (random sample) per school	1 grade 4 teacher, 3 classroom observations, and 5 grade 4 students (random sample) per school	1 grade 4 teacher, 3 classroom observations, and 5 grade 4 students (random sample) per school

Table 14. System-level interviews in Karnataka, India (GKA)

Level	Department or position
State ministry	▪ State project director, Samagra Shikshana Karnataka ▪ Director, Department of State Educational Research and Training ▪ Senior assistant director of public instruction, Department of State Educational Research and Training ▪ Program officer, Samagra Shikshana Karnataka
Bengaluru Rural	▪ Principal- District Institute of Education and Training ▪ GKA nodal person, District Institute of Education and Training ▪ Senior lecturer, District Institute of Education and Training ▪ Lecturer, District Institute of Education and Training
Chikballapur	▪ Principal, District Institute of Education and Training ▪ GKA nodal officer, District Institute of Education and Training ▪ Lecturer, District Institute of Education and Training
Akshara Foundation staff	▪ Chairperson ▪ Director of research and evaluation ▪ Head of mathematics program ▪ Pedagogy expert

3.2.4 Nanhi Kali – India

Table 15. School sample achieved in Andhra Pradesh, India (Nanhi Kali)

80 academic support centers			
	Anakapalli	**Visakhapatnam**	**Alluri Seetharama Raju**
Overall sample	15 centers	25 centers	40 centers
Classroom observation	10 centers	20 centers	30 centers
Cognitive interviews with students	5 centers	5 centers	10 centers

3.2.5 RAMP – Jordan

Table 16. Quantitative sample achieved in Jordan (RAMP)

80 schools			
الجنوب / Junoob – South	**الشمال/ Shamal – North**	**الوسط / Wasat – Central**	
الكرك / Karak	اربد / Irbid	الزرقاء / Zarqa	عاصمة عمان / Amman
24 schools	16 schools	20 schools	20 schools

Table 17. Qualitative sample achieved in Jordan (RAMP)

9 schools
9 schools subsampled from Marka field directorate in Amman Governorate (6 schools) and Zarqa first field directorate in Zarqa Governorate (3 schools)
1 grade 2 teacher per school (total of 9 teachers interviewed) 5 grade 2 students per school (total of 45 students interviewed)

Table 18. System-level interviews in Jordan (RAMP)

Level	**Department or position**
Central ministry	▪ Head of textbooks division (Curriculum Department) ▪ Math specialist (Curriculum Department) ▪ Head of supervision division (Education Training Center) ▪ Head of educational policies division (Education Training Center) ▪ Head of monitoring and evaluation (Education Training Center) ▪ Head of school and directorate development section (Education Training Center) ▪ Head of planning and research division ▪ Head of testing division (Examinations Department) ▪ Executive director of National Center for Curriculum Development ▪ Deputy director of National Center for Curriculum Development ▪ Technical Assistance Program project director ▪ [Former] Queen Rania Teacher Academy math specialist
District 1	▪ Head of field directorate 1 ▪ Supervisor 1 ▪ Head of supervision 1 ▪ Head of supervision 2

3.2.6 TAFITA – Madagascar

Table 19. Quantitative sample achieved in Madagascar (TAFITA)

Intervention schools (80 schools)				Comparison (50 schools)	
Amoron'i mania		Analamanga		Alaotra-mangoro	
Ambositra	Fandriana	Ankazobe	Manjakandriana	Ambatondrazaka	Moramanga
20 schools	20 schools	20 schools	20 schools	25 schools	25 schools

Table 20. Qualitative sample achieved in Madagascar (TAFITA)

Intervention 5 schools – Amoron'i Mania	Comparison 5 schools subsampled from Alaotra-Mangoro
1 grade 2 teacher/classroom per school 5 grade 2 students (random sample)*	1 grade 2 teacher/classroom per school 5 grade 2 students (random sample)*

** The qualitative sample included 48 students (two grade 2 students were missing).*

Table 21. System-level interviews in Madagascar (TAFITA)

Level	Department or position
Central ministry	▪ General secretary of the ministry ▪ Director of inspection, National Inspection Directorate ▪ Director, National Institute for Pedagogical Training ▪ Deputy director, National Institute for Pedagogical Training ▪ Head of training, National Institute for Pedagogical Training ▪ TAFITA focal point, National Office of Pedagogy ▪ Former head of service, Basic Education Service
Funding Agency (JICA)	▪ Social development officer
Implementing partners	▪ Education advisor, TAFITA office ▪ Two technical consultants, TAFITA office ▪ Executive director, SOFIASIVE ▪ Financial and operations director, SOFIASIVE
Regional Education Office, Anamalanga (DREN)	▪ Regional director ▪ Head of school services ▪ School management committee point of contact
District Education Office, Manjakandriana (DREN)	▪ District education head ▪ Head of pedagogical support division ▪ School management committee point of contact ▪ Head of a pedagogical support zone
District Education Office, Ankazobe (DREN)	▪ District education head

Level	**Department or position**
	▪ Head of school services ▪ Head of pedagogical support division ▪ School management committee point of contact ▪ 14 heads of a pedagogical support zone
Regional Teacher Training Center, Analamanga	▪ Center director ▪ Center pedagogical coordinator

Table 22. Final instrument count, by program

	Instrument	GKA (India)	Nanhi Kali (India)	ESMATE (El Salvador)		RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)		
				Original treatment group	Scaled-up program group			Control	Comparison (normal class)	Treatment (TAFITA class)
Quantitative, school level	Head teacher interview	79	79	29	60	80	80	50	n/a	77
	Teacher interview	79	79	48	108	79	79	51	n/a	80
	Lesson observation	78	57	48	108	80	80	50	78	70
	MKT survey	80	n/a	48	108	79	n/a	50	n/a	79
	Trainer interview	2	n/a	n/a	n/a	15	4	47	n/a	50
	Coach interview	n/a	5	n/a	n/a	15	4	n/a	n/a	50
	Meeting facilitator interview	n/a	5	n/a	n/a	26	n/a	n/a	n/a	n/a
	Teacher meeting observation	n/a	6	n/a	n/a	13	n/a	n/a	n/a	n/a
	Coach/mentoring observation	n/a	n/a	n/a	n/a	14	n/a	n/a	n/a	n/a
Qualitative, school level	Teacher interview	9	n/a	4	6	9		5	5	n/a
	Lesson observation (3 days each)	9	n/a	4	6	9	n/a	5	5	n/a
	Student interview	45	44	19	29	45	n/a	21	23	n/a
Qualitative, systems level	Central/regional ministry interviews	12	n/a	n/a	3	12	3	n/a	n/a	11
	District/local ministry interviews	4	n/a	n/a	2	4	2	n/a	n/a	22
	Donor/program interviews	9	n/a	n/a	1	7	1	n/a	n/a	5

3.3 Instruments

The majority of instruments used for this study were adapted from Phase 1 of the Learning at Scale study. The development of these tools is detailed in Section 3 of the *Learning at Scale Report* (www.learningatscale.net). These instruments include the following:

Quantitative Instruction Interviews: Teachers and Teacher Support Providers

- Teacher interview[1]
- Head teacher interview
- Coach interview (where applicable)
- Trainer interview
- Teacher meeting facilitator interview (where applicable)

Quantitative Instruction Observations: Teachers and Teacher Support Providers

- Coaching observation (where applicable)
- Teacher meeting observation (where applicable)

Qualitative Systems-Level Interviews

- Central ministry or high-level education official interview
- District- or county-level education official interview
- Program staff interview
- Donor staff interview

These instruments were reviewed and revised to include (1) specific mathematics-related elements, including math-related materials; (2) questions related to COVID-19 school closures and remediation efforts; and (3) more explicit investigation of counterfactuals (similar programs that were not effective).

3.3.1 Mathematics instruments

For this phase of the study, the research team developed, tested, and finalized mathematics-specific quantitative and qualitative classroom observation and teacher interview tools, a student cognitive interview tool, and the MKT Survey.

3.3.2 Tool development

Quantitative classroom observation tool

The research team began by mapping the observation tool using observable components of effective instructional practices from the *Framework for Teaching* by the Danielson Group.[2] From these components, we listed the subskills that we felt were most relevant for an observation, again pulling from the *Framework for Teaching*. We then mapped the

[1] Given variations in program design and implementation, we targeted teachers or facilitators of varying grades and learning levels (for after-school programs). All interviews and observations, however, were conducted in early primary (grade R/kindergarten through grade 4).

[2] The *Framework for Teaching*, developed by Charlotte Danielson (2007), is a detailed roadmap for effective instruction, covering four domains: planning and preparation, classroom environment, instruction, and professional responsibilities.

observable components and subskills to evidence-based high-leverage instructional practices and strategies. We incorporated any key practices and strategies that were not already represented.[3] We then added any components and subskills that were math specific and were not present in the *Framework for Teaching*. Finally, we listed behaviors in the classroom aligned to these components that we believed would be observable.

This mapping of frameworks became the core of the quantitative observation tool. We took the observable behaviors and categorized them according to when they might occur (during whole-class instruction, independent work, or both) and then organized them in a format that we believed would be understandable to data collectors.

Qualitative classroom observation tool

The qualitative observation methodology was closely tied to the quantitative tool. It was designed to provide insight on elements of the quantitative tool. For example, the quantitative tool let us know if the teacher asked an open-ended question; meanwhile, the qualitative tool told us what that question was, how it was framed, who answered it, and any discussion that might have followed it. The methodology started with a running record of notes of the lessons, guided by four "buckets" of topics and subtopics:

1. Clarity and effectiveness of math/model/explanation/use of representation, including questions, incorrect responses, math errors, and use of explaining why incorrect is incorrect
2. Appropriateness/difficulty and sequencing of content
3. Task types: appropriateness, difficulty level, amount, and alignment
4. Student engagement, including questions, group work, and individual work

Qualitative researchers (selected for in-country expertise in math education and research) were given these topics, plus examples and guiding questions, on a laminated sheet of paper that they used to guide their notetaking. A data-entry matrix was also developed according to these topics, including prompts to help researchers separate their thoughts and impressions from the actual evidence in their notes.

Quantitative teacher interview

For Numeracy at Scale, we edited the original Learning at Scale (Phase 1) teacher interview tool to focus on mathematics—rather than literacy—instructional and learning components. A teacher attitudes scale was also added. Otherwise, the tool remained consistent with the quantitative teacher interview used in the first phase of Learning at Scale.

---

[3] The University of Michigan has detailed high-leverage teaching practices for elementary mathematics instruction, including leading a discussion, explaining and modeling content, interpreting student work and providing feedback, and coordinating and adjusting instruction (University of Michigan Teaching & Learning Exploratory, n.d.).
Experts from several organizations working in low- and middle-income contexts compiled the research-based instructional strategies document, which explains four key strategies in early math instruction: using developmental progressions, connecting formal and informal math, explanation and justification, and using multiple models and representations.
https://shared.rti.org/content/instructional-strategies-mathematics-early-grades

Qualitative teacher interview

The qualitative teacher interview, conducted following the third observation, was designed to query the teacher on specific aspects of the lesson as observed, to better understand why the teacher made the decisions they did.

Mathematical Knowledge for Teaching survey

MKT is the knowledge of students, instruction, and content that teachers need in order to be effective mathematics teachers (Lowenburg-Ball et al., 2008). Building from other surveys, Learning at Scale mathematics experts Dr. Yasmin Sitabkhan and Dr. Wendi Ralaingita developed the MKT survey in 2019–2020 in the Kyrgyz Republic and Nepal using RTI internal research and development funding.[4] The specific aim of this survey is to have an open-source, adaptable survey that is targeted to low- and middle-income countries. The survey builds on existing work and consists of multiple-choice problems that target four domains of knowledge (number, operations, geometry, and measurement). There are three problem types:

- Developmental progressions, which aim to diagnose teachers' knowledge of how students learn math in the early grades.
- Scaffolding, which aims to diagnose teachers' understanding of common misconceptions that students have, and how best to address them.
- Content, which aims to diagnose teachers' knowledge of math content in the primary grades.

We adapted the MKT survey for use in each country by working with local mathematics education experts in advance of data collection.

Cognitive interviews with students

The cognitive interviews with students were intended to provide insight into students' development of higher-order skills and conceptual understanding of mathematics. The cognitive interview tools asked students to solve a problem, with the data collector noting how they solved the problem. Students were then asked how they solved the problem, and data collectors noted their clarity and ability to justify their solution. Finally, data collectors provided a "countersuggestion," a common technique in mathematics education whereby the data collector tests the robustness of students' answers and their ability to reason through alternate solutions.

3.3.3 Tool piloting

Quantitative and qualitative tools were piloted in Nairobi, Kenya, in Alternative Provision of Basic Education and Training low-cost private schools by a team of three trainers and five highly skilled data collectors with prior experience conducting classroom observations under Learning at Scale. The pilot activity ran for two weeks, with training and initial testing taking place during the first week and quantitative tool piloting taking place during the second week.

Following an initial two-day training on the observation and interview tools, the full team visited one to two schools per day. The general approach to piloting was to conduct

[4] For more information, see Ralaingita et al. (2023).

classroom observations, teacher interviews, and student interviews in the morning, and then debrief and revise the instruments as needed in the afternoon. Key decisions were made in this way.

Once revisions to the quantitative observation tool were finalized at the end of week 1, the instrument was rendered for tablet-based data collection using RTI's Tangerine platform.

To pilot test the quantitative classroom observation tool, two teams of two data collectors visited two schools per day (one school per team), conducting a total of 50 observations of math lessons in grades 1–4. The pilot team leader alternated between the teams, conducting inter-rater reliability (IRR) tests with data collectors on a rotating basis.

Analysis of the eight IRR tests conducted during piloting demonstrated two instances of a kappa value between 0.61 and 0.80, indicating substantial agreement (based on the criteria set by McHugh (2012)). The other six IRR instances saw a kappa value between 0.81 and 1, indicating almost perfect agreement (**Table 23**).

Table 23. IRR results during piloting

id	n_enumerators	kappa	obs_prop	exp_prop	se_kappa	z_kappa
1	2	0.716905	0.828282828	0.393429	0.064252	11.15771
2	2	0.911035	0.949494949	0.432303	0.067831	13.43103
3	2	0.982478	0.98989899	0.423528	0.068634	14.31471
4	2	0.850953	0.909090909	0.390062	0.0656	12.97178
5	2	0.930914	0.95959596	0.415162	0.067812	13.72784
6	2	0.862524	0.919191919	0.412203	0.066533	12.96383
7	2	0.786062	0.878787879	0.433425	0.067901	11.57667
8	2	0.945951	0.96969697	0.439343	0.069715	13.56892

The qualitative and quantitative classroom observation tools developed specifically for this phase of the study can be found in **Annex E**.

3.3.4 Instrument translation and rendering

After finalization, we hired translators to translate all tools to be administered by trained data collectors (**Table 24**). Following translation, these tools were reviewed by our in-country education consultants and data collection firms to ensure that appropriate technical language was used where necessary. Quantitative tools were then rendered in RTI's tablet-based data collection platform, Tangerine. Qualitative tools were formatted for paper-based data collection. All tools were again refined during data collector training and piloting.

Table 24. Translation needs, by program

	ESMATE	GKA	Grade R	Nanhi Kali	RAMP	TAFITA
Training materials	Spanish	Kannada	English	Telegu	Arabic	French
Quantitative interviews	Spanish	Kannada	Afrikaans, IsiXhosa	Telegu	Arabic	French, Malagasy

	ESMATE	GKA	Grade R	Nanhi Kali	RAMP	TAFITA
Qualitative interviews (instruction)	Spanish	Kannada	Afrikaans, IsiXhosa	Telegu	Arabic	French, Malagasy

3.4 Study Limitations

When interpreting study findings, it is important to keep the following limitations in mind:

- Instruments developed for use across six programs and five countries lose some attention to each program's unique design and operating context. While teams field-tested and adapted response options and added a few context-specific items during data collection preparations, the focus of each interview and observation item had to be consistent across programs in order to ensure cross-country applicability. As a result, some of the nuances of the programs may not be discussed in the findings.
- Depending on the design and implementation of each program, some interviews were not conducted during data collection. Therefore, the total sample and instrument count varies in each data collection. As noted in the section above, the samples drawn in each country are not statistically representative of the population receiving these intervention programs.
- Due to differences between programs, we were not able to collect student- or class-level data from a single grade across all programs. While we did adapt observation tools (as much as possible to maintain comparability) to cater for differences between grades, this should be kept in mind, especially when interpreting MKT results and classroom observation findings.
- Language is always a key consideration for conducting any research, and it played an important role in every country of study. Data collector trainings were conducted in six different languages, interviews and observations were conducted in nine different languages, and data were collected in nine languages. While all translations underwent an additional round of review, it is highly likely that meaning was lost in translation in some cases.

4 DATA COLLECTION

4.1 Program Entry Visits

Once the six programs were selected for inclusion in the Numeracy at Scale study, we organized program entry visits to each country. This step served as our first pre-data-collection activity. The purpose of these visits was to begin in-depth conversations with programs regarding the mutual expectations of the Numeracy at Scale research team and the in-country program team. The timing of these visits is shown in **Table 25**.

Table 25. Program entry visits, in chronological order

Program	Country	Date of visit
ESMATE	EL Salvador	October 4, 2022
TAFITA	Madagascar	October 24,2022
R-Maths	South Africa	October 24, 2022
RAMP	Jordan	November 7, 2022
Nanhi Kali	India	November 14, 2022
GKA	India	December 5, 2022

These visits proved invaluable for our team to better understand each program's structure and intervention approach, as well as to work jointly with each program's team on planning for our large-scale data collections.

4.2 Data Collector Training

In preparation for the school-level data collection activity for each program, the Numeracy at Scale team led a five-day, in-person training of 18 to 26 data collector candidates. The overall objective was to train the candidates to become fully capable of accurately and reliably administering all Numeracy at Scale instruments and protocols to ensure that consistent, high-quality data would be collected across countries and programs.

Each quantitative training was led by two RTI staff with expertise experience in leading trainings on administering interviews and classroom observation tools. The qualitative training sessions were led by an RTI staff member or an RTI consultant, all of whom possess significant knowledge in mathematics instruction, research, and qualitative research methods.

Before the first in-country training, RTI developed a manual that detailed step-by-step daily procedures. The documented training approach incorporated experiential learning and skills demonstrations (e.g., classroom practice videos, role play and peer practice, and situated learning opportunities in real settings) and self-reflection and discussion sessions. All quantitative and qualitative trainers took part in a remote master training, led by our numeracy research director, Dr. Yasmin Sitabkhan, to further enhance their proficiency of the training manual and ensure congruency among country-level trainings.

The Numeracy at Scale team created training materials in easily accessible electronic formats, which facilitated sharing, updating, adapting, and reusing them across all trainings. All instructional materials were cross-referenced to the training manual. They included presentation slide decks, video clips, peer-evaluation checklists, quick-reference guides, discussion prompts, mini quizzes, sample lesson logs, and descriptions of possible assessment scenarios.

Local data collection firms selected and hired the training participants based on the candidates' prior experience and on recommendations from sector partners in-country. Although 24 data collectors would be required to cover all data collection components in each country, 26 candidates participated in each training, allowing for selection of the 24 trainees who scored highest on the accuracy and reliability measures administered throughout the training.

4.2.1 Ensuring data reliability and accuracy

To measure the accuracy and reliability of the candidates' tool administration, over the course of the training, the quantitative training team administered three assessor accuracy measures[5] (AAMs):

- **One interview AAM (day 2)**: All the data collector trainees observed a mock in-person interview conducted by the Numeracy at Scale in-country consultant and a data collection supervisor (both of whom were fluent in the language of data collection). Their results served as a gold standard for comparison to the trainees' results.
- **Two classroom observation AAMs (days 3 and 4**): The participants watched a video of a lesson being taught and recorded their observations. Their scoring was then compared to a gold standard pre-scored by the master trainers.

Table 26 provides the average AAM scores on three measures taken during the training week across all six programs. These scores represent all data collector candidates who participated in the training. As previously noted, the lowest-performing trainees were not engaged after the week of training; only the top 24 candidates with the best scores proceeded to the actual data collection stage.

Table 26. Results of trainee accuracy and reliability measures

Measure	Average percentage scores, by program					
	GKA	Nanhi Kali	ESMATE	RAMP	R-Maths	TAFITA
Interview AAM	82	86	85.00%	91	87% (teacher interview) 89% (head teacher interview)	87
Classroom observation, AAM 1	79	87*	88.80%	87	83%	81

[5] The "assessor accuracy measure" approach was developed by RTI to ensure reliability across data collectors, and it was conducted with all data collectors.

Measure	Average percentage scores, by program					
	GKA	Nanhi Kali	ESMATE	RAMP	R-Maths	TAFITA
Classroom observation, AAM 2	88	n/a**	n/a**	88	93%***	88

Boldfaced values represent final reliability estimates for each measure.
** Due to the unique nature of the observation tool, the AAM did not include observation items. Data collectors were tested on post-observation student interview items only.*
*** Because the scores on the first classroom observation AAM were sufficiently high, a second AAM was not conducted. However, after each AAM, the trainer went over the items with the highest discrepancy with the group and provided further clarification and practice in those areas.*

**** R-Maths data collector training conducted a classroom observation AAM 3 for which the average agreement score was 96%.*

4.3 Data Collection

4.3.1 School-level data collection: Quantitative

Every team was composed of four data collectors, one of whom assumed the role of team leader and maintained communication with the data collection firm to report daily targets. Each team was divided into two pairs of data collectors, with each pair assigned to visit a single school. By organizing into pairs of data collectors, each team could visit two schools per day. The data collectors contacted the school approximately two days prior to their arrival to notify them about the visit. Additionally, the teams or local subcontractor staff reached out to coaches, trainers, and facilitators of teacher meetings ahead of time to schedule interviews and observations either during, before, or after the school visits.

Depending on the instrument, data were collected directly on tablets using Tangerine software or on a paper copy first and then inputted into Tangerine. Teams synced their data to the cloud as soon as possible following completion of the day's work. The data collection firm reported sample counts for each day to RTI statisticians and copied the program team. In collaboration with lead trainers, RTI statisticians provided daily feedback from data quality checks.

4.3.2 School-level data collection: Qualitative

Each qualitative researcher visited one sampled teacher over three consecutive days. Each day, the researcher observed one math lesson. A post-lesson debrief occurred every day following the lesson and the day's instruction. On the third day, the researcher conducted a longer teacher interview. At some point over the three days while visiting a school, the researcher would also conduct five cognitive interviews with students in the observed teacher's class. All data were collected on paper and subsequently entered electronically.

4.3.3 Systems-lead data collection: Qualitative

Systems leads finalized target interview lists prior to travel, with the guidance of research consultants and in consultation with country leads. Research consultants prepared instructions and arranged for interviews. For all interviews with system participants, either the central government interview tool or the local/subnational government interview was used. The wording of these questions was sometimes adapted slightly, depending on the role of the respondent (e.g., focus on inspection, management, training, curriculum, materials, etc.). Some questions were omitted if they were not relevant to the program of study or the respondent. In addition to taking interview notes, leads were strongly encouraged to capture illustrative quotations. All qualitative notes were entered into Dedoose software, with systems leads aligning evidence to hypotheses first and then sending the data for analysis and review.

The research team set a minimum benchmark of 75% for data collectors to be allowed to administer the classroom observation instrument, in comparison to 80% for the teacher interview. The scores obtained by the prospective classroom observers were in line with those achieved in other contexts, given the level of training, the complexity of the instruments, and the variability in classroom activity (Brown et al., 2010; Wolf et al., 2018).

The classroom observation tool used for Nanhi Kali was different than the classroom observation tool used for other programs. Therefore, the Nanhi Kali data collector training did not include an AAM on the observation tool. Instead, data collectors participated in an AAM on the post-observation student interview questions.

During data collection, RTI statisticians conducted daily data quality checks and provided feedback and follow-up on any issues that arose.

4.4 Program Details

Tables 27 through 32 provide a summary of all completed data collection activities.

Table 27. TAFITA (Madagascar)

	Quantitative	Qualitative
Training dates	February 6–10, 2023	January 30–February 3, 2023
No. of data collectors trained	39	3
Trainers	Julianne Norman, Norma Evans	Norma Evans
Data collection dates	February 13–21, 2023	February 13–March 22, 2023
Data collection firm	Education NGO Network	Education NGO Network

Table 28. GKA (India)

	Quantitative	Qualitative
Training dates	February 6–10, 2023	February 6–10, 2023
No. of data collectors trained	24	6
Trainers	Cosmus Gatuyu, Kellie Betts	Wendi Ralaingita
Data collection dates	February 13–27, 2023	February 13–27, 2023
Data collection firm	SIGMA	SIGMA

Table 29. Nanhi Kali (India)

	Quantitative and qualitative (combined)
Training dates	February 20–24, 2023
No. of data collectors trained	30*
Trainers	Kellie Betts, Robert Momanyi, Mitch Rakusin
Data collection dates	February 27–March 21, 2023
Data collection firm	SIGMA

** Data collectors were divided into two groups during training: (1) group 1 data collectors were trained on the community associate interview, headmaster interview, program officer/staff interview, school monitoring visit observation, and monthly review meeting observation; (2) group 2 data collectors were trained on classroom observation, student qualitative interview, and program officer/staff interview.*

Table 30. RAMP (Jordan)

	Quantitative	Qualitative
Training dates	February 26–March 2, 2023	March 7–9, 2023
No. of data collectors trained	26	3
Trainers	Laiba Bahrawar, Karon Harden	Shirin Lutfeali
Data collection dates	March 5–19, 2023	March 12–20, 2023
Data collection firm	Dajani Consulting	Dajani Consulting

Table 31. ESMATE (El Salvador)

	Quantitative	Qualitative
Training dates	April 17–21, 2023	April 17–21, 2023
No. of data collectors trained	23	2
Trainers	Jessica Mejia, Karon Harden	Yasmin Sitabkhan
Data collection dates	April 23–May 5, 2023	April 24–May 12, 2023
Data collection firm	FEDISAL	FEDISAL

Table 32. R-Maths (South Africa)

	Quantitative	Qualitative
Training dates	May 15–19, 2023	May 15–19, 2023
No. of data collectors trained	28 (25 data collectors + 3 field supervisors)	12
Trainers	Kellie Betts, Robert Momanyi	Kellie Betts
Data collection dates	May 29–June 9, 2023	May 29–June 9, 2023
Data collection firm	Decipher Data	Decipher Data

5 FINDINGS

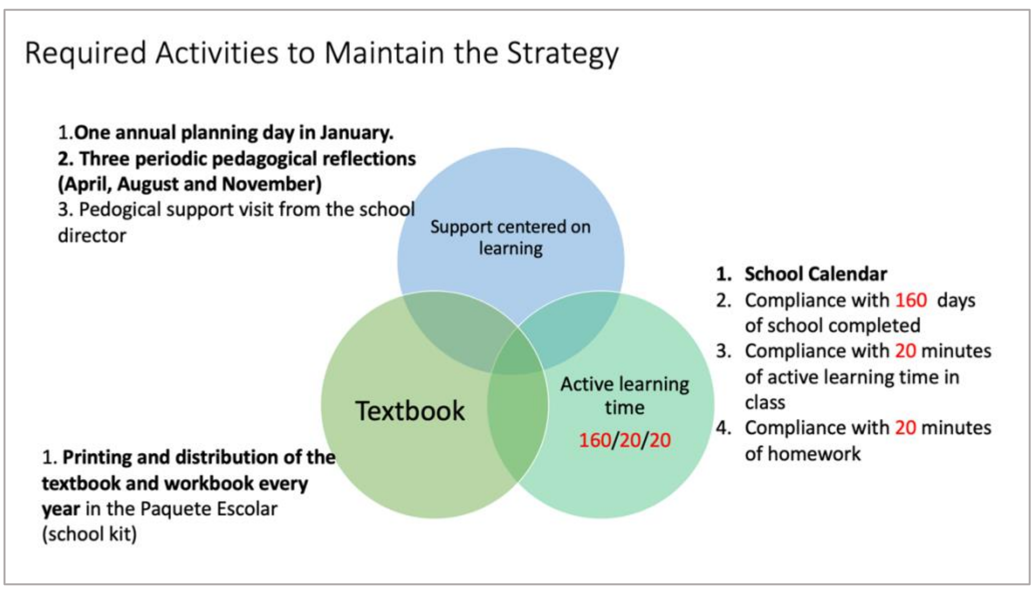
5.1 Program Overviews

Although all six of the programs selected for inclusion in this study had evidence of improved math outcomes for early grade students, their designs and overall approaches varied. For example, four of the programs focused on core instruction in public schools (ESMATE, GKA, R-Maths, and RAMP), while two focused on remediation or alternative learning opportunities (Nanhi Kali and TAFITA). The grades ranged from kindergarten only (R-Maths) to grades 4 and 5 only (GKA). Instruction in most programs was provided by teachers (ESMATE, GKA, R-Maths, RAMP, and TAFITA), but in Nanhi Kali the core instruction was provided by tablet-based software and was supported by community volunteers. Moreover, program funding came from a variety of sources, including bilateral donors, local foundations, and governments.

5.1.1 ESMATE

The ESMATE program is implemented by El Salvador's Ministry of Education. The program supports mathematics instruction and learning in grades 1–11 for all schools in the country, with technical support provided by JICA. ESMATE's theory of change is centered around three elements: (1) the provision of high-quality textbooks for every student, provided every year; (2) active time on task, with students working independently; and (3) teacher support for student learning, including one annual planning day, three annual teacher reflection days, and pedagogical support from school directors (see **Figure 4**).

Figure 4. ESMATE's theory of change



Source: JICA

Evaluation

From 2018 to 2019, ESMATE consisted of a pilot project, implemented with support from JICA, to evaluate the impact of the newly created student textbooks, paired with teacher training and ongoing classroom support by the ESMATE team. The pilot used a randomized controlled trial and tracked the same students through two years of schooling; the treatment group received the intervention for two years, and the control group received the intervention for only one year. The RCT was conducted in 250 public schools in four departments (Cabañas, La Union, San Miguel, and San Vicente). After two years of instruction, 2nd grade students in year 1 in the treatment group showed improved scores on math outcomes (estimated at 0.49 standard deviations after year 1, and 0.13 standard deviations in year 2 compared to students in the control group who only received one year of exposure (Maruyama & Kurosaki, Forthcoming).

After these successful pilot results, in 2020 ESMATE was scaled up to all of the country's public schools—namely, 4,666 primary schools, 2,726 junior high schools, and 705 senior high schools. Given the scale, certain elements of the pilot needed to be adapted; for example, training and intensive support were not provided to the scaled-up regions, and the program instead relied on existing system-level supports that were already embedded in teachers' daily lives.

Program model

Pedagogical Approach

ESMATE's approach centers on high-quality student textbooks that are provided to all students every year. Classroom materials are structured to provide ample opportunities for students to practice problems and gain an understanding of the content, both during class and at home. The program's pedagogical approach emphasizes a problem-solving approach with clear models and time to practice, as seen in **Figure 5**. These five steps are embedded into every lesson in the student textbooks, as well as the teacher's methodological guide.

Figure 5. ESMATE's problem-solving approach

Structured materials: All lessons promote student engagement and mathematical thinking using a consistent structure:

- *Analiza (analyze)*: Problems for students to solve independently
- *Soluciona (solution)*: The solution to the problem(s) presented in Analiza
- *Comprende (understand)*: Explanation and model of the key concept of the lesson
- *Resuelve (solve):* Problems to solve in class
- *Resuelve en casa (homework):* Problems to solve at home

Teacher Training and Teacher Support

Given that organizing large-scale teacher trainings is not financially sustainable, ESMATE relies on preexisting structures within the system to train and support all teachers. In particular, the ESMATE team takes advantage of the teacher pedagogical reflection sessions that are run three times a year, organizing topics of discussion focused around the ESMATE approach and math content. In addition, at these sessions, teachers review content that is upcoming and plan for future instruction.

The central ESMATE team also provides support as needed to regional office staff who are involved with pedagogical support (*gestores pedagógicos*). Lastly, school directors are charged with providing teacher support when needed.

Institutionalization

In this regard, ESMATE has two unique features:

1. The government of El Salvador budgets every year for a "paquete escolar" that is provided to all students at the start of the school year. The ESMATE team was able to include the printing of textbooks in this budget, ensuring that all students in the country receive a new textbook every year.
2. The ESMATE team is based in the Ministry of Education, with each member being responsible for communications with departmental staff for a particular region. The team is well staffed and focuses only on ESMATE.

5.1.2 R-Maths

R-Maths (formally titled the Grade R Mathematics Programme, accompanied by a Grade R language program called ELIT) was initiated in 2017 in the Western Cape, one of South Africa's nine provinces, by the Western Cape Education Department (WCED)—part of the provincial government—with funding from the Zenex Foundation and the Maitri Trust and technical support from the Schools Development Unit at the University of Cape Town.

The goal of the program is to improve the conceptual understanding and mathematics skills of grade R students in the Western Cape. It has three specific objectives:

1. To develop an effective program (including resources and training materials) that supports a conceptual approach to grade R mathematics instruction and learning.
2. To capacitate grade R and foundation phase subject advisors in the Western Cape to train and support grade R teachers and practitioners effectively and to serve as a knowledgeable resource for grade R mathematics content and instruction methods.
3. To capacitate grade R teachers and practitioners in foundational mathematics conceptual knowledge and instruction skills and improve the quality of grade R mathematics instruction.

For the program's initial rollout, the province was divided into two groups that adopted the program sequentially. Training for the first phase of the program took place with one group from January to June 2017. The second group received training from October 2017 to June 2018. During this time, the program reached 79 foundation phase subject advisors and other WCED officials and approximately 3,000 grade R teachers and practitioners. In 2019, the focus turned to the training of novice teachers by lead teachers and to raising awareness among department heads about R-Maths. The period 2019–2020 was considered a "consolidation phase," where the program continued without donor and technical support. Since 2019, R-Maths has been funded and run by WCED alone.

Evaluation

An evaluation of the program was conducted in 2018 by JET Education Services and Kelello Consulting. The impact of the program on students' mathematical knowledge was assessed through a quasi-experimental evaluation comparing schools that had received the program with those waiting to receive the program. This evaluation was conducted with 622 students in 101 urban schools and 51 rural schools. Among rural schools, students in the intervention group had improved mathematics scores compared to the control group, with an effect size of around 0.20. There was no significant difference between the intervention and control group in the urban schools. Implementers felt that the small effect size found in the study was encouraging for the first year of rollout, with the expectation that the impact would increase as the program continued to implement.

The evaluation also assessed subject advisors—subject specialists based in district offices who support teachers in schools—for their grade R pedagogical knowledge, subject matter knowledge, and pedagogical content knowledge. Their overall performances on these assessments rose from 48.6% before training to 66.1% after training. Teachers' scores on the same test improved from 51.4% before training to 68.1% after training. Improvements for both groups were statistically significant.

A subsequent evaluation of the 2019–2020 consolidation phase took place (Roberts and Mawoyo, 2020). This evaluation focused on the mechanisms used to embed the R-Maths program into provincial-, district-, and school-level functioning. It identified several factors that promoted R-Maths' sustainability, including provincial buy-in and ownership that was assured from the start, with deep engagement by provincial leaders in the design of materials and the program as a whole; a budgetary commitment to the continuation of the program, with training extended to incorporate new teachers; and the integration of monitoring and evaluation processes into provincial reporting. The evaluation found that although the program faced numerous challenges—such as challenges related to the establishment of effective governance structures, a rollout that was considered too fast, and questions about the effectiveness of professional learning communities in supporting teachers—R-Maths demonstrated an ability to adapt over time, which was critical to ensuring success. For example, program leaders adjusted the dosage and structure of training over time so that it was more "demand led," and they adapted trainings to include foundation phase department heads as beneficiaries and local "lead teachers" as trainers.

Program model

The program was designed to align with the national Curriculum Assessment Policy Statement (CAPS), which details the learning goals of South Africa's education system. Upon its rollout, the program restructured the province's math curriculum to ensure an increased focus on children's understanding of mathematical concepts and ability to carry out mathematical procedures.

Teacher Training

The training portion of R-Maths follows a two-stage cascade model. The first stage involves training subject advisors over a five-day period. After this training, subject advisors conduct "dry runs" in which they model a lesson and get peer feedback. In the second stage, subject advisors are then responsible for training grade R teachers.

Teacher Support

The main support to grade R teachers comes from foundation phase subject advisors. "Foundation phase" refers to grade R to grade 3. Each subject advisor is responsible for a "circuit" consisting of approximately 10–25 schools. Subject advisors visit these schools on average around once a term, although the frequency of their visits depends on the needs of each school.

Grade R teachers are also supported by a professional learning community that meets termly to discuss a particular topic each time. There is also a WhatsApp group for the professional learning community. Lastly, grade R teachers receive support from district-level foundation phase heads.

Pedagogical Approach

The curriculum materials consist of a concept guide—which maps out the scope and sequencing of the curriculum—and activity guides with ideas for how to translate the concepts into practical activities, organized by term and week. The activity guides are intended to give teachers ideas, without being prescriptive, for things to do in class.

The format of each lesson in the R-Maths program is whole-class instruction mixed with small-group work. Children are divided into groups, with each group working at one of five different workstations in the classroom each day. One of these workstations involves work with the teacher, who guides children through tasks and asks them questions to gauge their level of understanding. In this way, the station—known as "the eye"—allows for continuous informal assessment.

The program materials are available in all three of the dominant Western Cape languages (Afrikaans, isiXhosa, and English) and include a facilitator's guide, a teacher's manual, a concept guide, activity guides, posters, and materials for students.

System

The program—which was initiated and is owned by WCED—is supported at the district level by subject advisors and district foundation phase heads. Districts, in turn, are supported by a small team based at WCED. Although WCED sought technical support from external organizations at the outset, the parameters for engagement were clearly set by WCED. Thereafter, the program's evolution from a multi-partner endeavor to a wholly WCED program was relatively seamless. Today, WCED is responsible for the recurrent costs of training—namely, training for new grade R teachers and subject advisors, as well as continuous professional development for existing teachers. WCED has appointed a new post to ensure the sustainability of R-Maths. This person is committed to ensuring that all grade R teachers receive training and support and that monitoring by subject advisors continues.

5.1.3 GKA

GKA (formally titled Ganitha Kalika Andolana) is a multi-stakeholder initiative designed by the Akshara Foundation for improving mathematics learning outcomes for primary students attending government schools in three Indian states: Karnataka, Odisha, and Andhra Pradesh. While GKA's current reach is considerable, for this study we examined GKA only in its origin state, Karnataka.

The roots of Akshara Foundation can be traced back to 2000, when they responded to a government request with programming focused on health and nutrition in low-income preschools. In 2007, this effort evolved into a structured curriculum, incorporating child development principles across 3,000 schools. By 2010–2011, external evaluations from universities highlighted gaps in child outcomes. Partially in response to these findings, in 2013, India introduced an early childhood policy, which prompted the Akshara Foundation to create a daily guide for teachers. The foundation's emphasis then transitioned to language-related interventions, utilizing technology, and the "I can read program" for grades 4, 5, and 6, eventually expanding to create print-rich environments through collaboration with other foundations.

The National Position Paper on Teaching of Mathematics in India, published in 2007, prioritized a number of curriculular goals: learning to enjoy math, using the techniques of math, considering math as a medium to communicate and work together, relating content to students' life experiences, and using abstract models to understand relationships and structures. However, data from primary school surveys showed that improvements in students' math learning had been difficult to achieve.

It was against this context that GKA was developed in 2011. The crux of the program's approach was to provide hands-on experience in mathematics teaching and learning with the aid of tactile and concrete teaching-learning materials (TLMs) to provide primary grade math teachers with multiple strategies and to help children enjoy math. The Akshara Foundation implemented GKA as a pilot in more than 500 government primary schools in Karnataka between 2011 and 2014. Encouraged by the pilot study findings, in 2014 the Government of Karnataka allocated budget to implement GKA in grades 4 and 5, in all 45,000 government schools. The Government of Karnataka earmarked budget for a phased roll out, procuring teaching and learning aids and building the capacity of teachers and school leaders.

Evaluation

Evidence on the impact of GKA on student outcomes draws from two studies: an internally commissioned longitudinal study (Vaijayanti et al., 2016) and an external experimental evaluation conducted by the Abdul Latif Jameel Poverty Action Lab (deBarros et al., 2023).

The internal, non-representative longitudinal study tracked 615 students (367 treatment, 248 control) in 21 schools across three years of schooling in Bengaluru Rural District, Karnataka. The school sample comprised two educational blocks, Hoskote (treatment block) and Devanahalli (control block). The student sample was broken down into three cohorts: cohort 1 followed students from Grade 1 in 2012 to Grade 3 in 2015; cohort 2 followed students from Grade 2 in 2012 to Grade 4 in 2015; cohort 3 followed students from Grade 3 in 2012 to Grade 5 in 2015. Each year, Competency-based pen and paper tests were administered to all the students in the sample at three timepoints, culminating in nine assessments through the duration of the study. Findings from a comparison of pre- and post-tests, by treatment group, are organized by cohort: cohort 1 saw a significant impact at Grade 3 endline (2015) with an effect size of 0.43; cohort 2 saw a significant impact at Grade 4 endline (2015) with an effect size of 0.27; cohort 3 significant impact at Grade 5 endline (2015) with an effect size of 0.34.

The external, randomized evaluation compared treatment and control groups comprising 98 Gram Panchayats and 294 schools in two districts of Karnataka: Bijapur and Tumkur. Data were collected from grade 4 students in sampled schools at three timepoints: November 2018 (Baseline), September 2019 (Midline) and February 2020 (Endline). While no significant positive impacts were detected for Gram Panchayat 'contests' or on student math outcomes overall, GKA was found to have a positive impact on girls' math outcomes only (0.18SD).

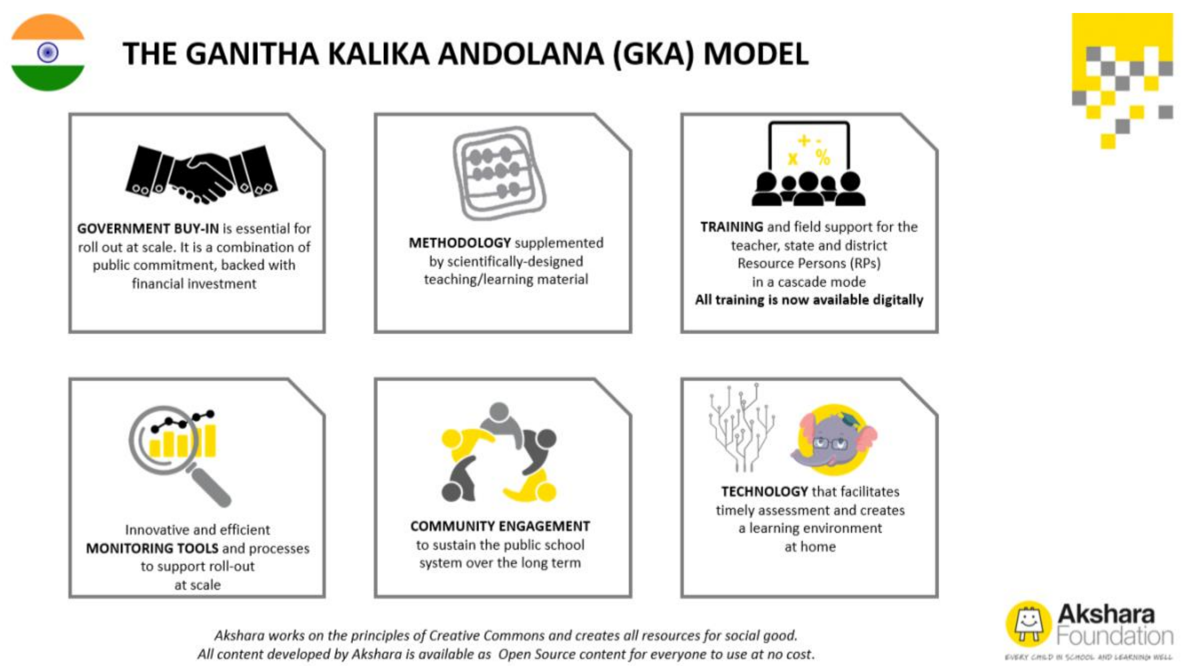
Program Model

The Akshara Foundation developed the GKA model to work at scale from the onset and funded the pilot. Now the state government funds the trainings and math kits and communicates expectations to districts and schools. The foundation continues to train the trainers and monitors the program. **Figure 6** presents the program's theory of change and program model.

Figure 6. GKA's theory of change and program model

Government buy-in	Monitoring		➤ scale	(e.g., multiple states; millions of texts; tech appropriate)
Teaching and Learning Materials	Assessment	Technology	➤ sustain	(e.g., volunteers; state investment)
Training	Community engagement		➤demand generation	(i.e., with scale and sustainability, it creates demand)
Building Blocks				

Source: Recreated from an image drawn by Akshara Foundation Chairman Ashok Kamath



The alignment between the program's content and government policies is one reason for the program's success. Another is the community's engagement in monitoring the use of math kits and student learning.

Materials

Each school receives a math kit that comes in a sturdy plastic box with handles and a lid. The supplies are intended to be used across the grades (grades 1–3 or 4–5). Each box includes the following materials: abacus, base ten blocks, clock, clothes clips, coins, decimal place value strips, decimal set, dice, fraction shapes, fraction strips, geo board, geosolids with nets, math concept cards, measuring tape, number line, place value mat, place value strips, play money, protractor and angle measure, square counters, tangram, and weighing balances. The cost per box is 3,500 rupees (about US$42) and is paid for by the state government.

Teacher Training

GKA's approach to teacher trainings has evolved to respond to current realities and improved technology. GKA designs the trainings, and the state government funds them. The three-level cascade model for in-person trainings begins with handpicked math practitioners to serve as key resource people who train a second level of resource people who then train the teachers. GKA ensures that at least one teacher per school is trained and that head teachers are educated about the math kits in order to support and monitor teachers' use of the materials. In addition, monthly teacher meetings are held at the cluster level and are organized by the government, not the Akshara Foundation. The monthly podcast *Pratibimba*—designed by teachers, for teachers—is another means to provide training to teachers on best practices.

During the COVID-19 pandemic, training was provided virtually. One method was via an online symposium attended by 5,000 teachers. Another was the sharing of videos and information through WhatsApp. Currently, a 25-hour GKA -online teacher training course is being implemented to build the capacity of teachers on GKA pedagogical approach and usage of kit. The online course is implemented through the government's technology platform (DIKSHA).

Teacher Support

GKA uses various methods to indirectly support teachers. One method consists of monitoring aimed at ensuring effective implementation of the program. Block and Cluster level government officers conduct routine school visits. Trained education volunteers visit school periodically, using an app to capture what is happening. The app has five, action-oriented binary questions about teachers' adherence to the class schedule, teacher presence and training, teachers' usage of instruction and learning materials, and group work activities. These visits help shed light on any differences in teachers' performance and guide future trainings and programming.

Pedagogical Approach

GKA's approach is based on the activity-based experiential learning approach recommended by India's National Curriculum Framework. The GKA model posits that learning is active, dynamic, and social and happens when students have hands-on experiences. Group collaboration among five to six students of mix-ability is a key feature of the program and encourages peer learning. The teacher handbook addresses all of the concepts in the government syllabus and provides guidance so that teachers can act more as facilitators than instructors during the group work.

Systems

Following the initial pilot, the Akshara Foundation worked hard to gain state government buy-in for the program. The GKA program is institutionalized within the education system in several ways: It has dedicated funding by the government for the procurement of math kits, it is aligned with the government's curriculum, and training takes place through the District Institute for Education and Training systems.

At the district level, GKA works through the block resource persons and cluster resource persons who visit the schools to monitor teachers. They hold regular cluster sharing meetings, which are seen as useful for flagging issues and finding solutions. Akshara also developed a mobile-based app that system actors are encouraged to use when they go out to schools.

Education officials at the state and district levels are responsible for procuring and distributing the GKA math kits, cascading training down to the teachers, and monitoring the use of kits in the classroom. Each district has also assigned one staff, the GKA nodal person, who is responsible for coordinating activities with GKA and in schools.

Currently, the government and Akshara Foundation are in the process of developing and piloting the math kits for grades 6–8 (GKA 2.0), all with government funds. Another method of monitoring aimed at understanding and supporting implementation is through math contests at the community level. Held at the lowest administrative unit (which consists of about seven to nine schools), children in grades 4, 5, and 6 convene at one school to take a written competency-based math test. Students are arranged according to their school, while teachers and community members serve as onlookers. Over a four-hour period, the assessment is given and scored, and the results are shared and discussed by the teachers and community. Trained education volunteers serve as mediators. The Akshara Foundation creates the assessment, and the day is funded by the local community.

5.1.4 Nanhi Kali

Nanhi Kali, funded by the Mahindra & Mahindra group of companies and other individual donors, through the K. C. Mahindra Education Trust and other donors, started in 1996 as an after-school education program for girls in primary grades. In 2005, the Naandi Foundation was brought on to jointly manage the program, and implementation expanded to include girls in secondary school up to grade 10. The program provides education support in three academic subjects (vernacular language, mathematics, and English as a second language) in grades 1 through 10, as well as science in grades 6 through 10. The program also includes a sports curriculum and provides each participant with a school supply kit annually. Over the years, the program has expanded across India and has reached more than 500,000 girls to date. Nanhi Kali currently operates in eight states and has 160,000 girls enrolled. Since 2019, the program has partnered with Education Initiatives to deliver personalized instruction through an adaptive learning software called Mindspark that is pre-loaded onto tablets.

Evaluation

Of particular interest to this study is the Mindspark software from Education Initiatives. An impact evaluation of Mindspark was conducted by the Abdul Latif Poverty Action Lab in 2015 and 2016, prior to the adoption of Nanhi Kali's adoption of the software (Muralidharan et al., 2019). The study assessed Mindspark's impact on students' mathematical knowledge through an experimental evaluation comparing the baseline and endline scores of 619 students recruited from five public middle schools located near Mindspark centers in Delhi. Students were randomly assigned to either the Mindspark intervention or to a control group. Students assigned to the Mindspark group received 90 minutes of additional instruction six days per week, culminating in 540 minutes of additional instruction in math, Hindi, and English per week over a period of about four and a half months. Baseline and endline assessments were conducted at the beginning and end of this period with all students.

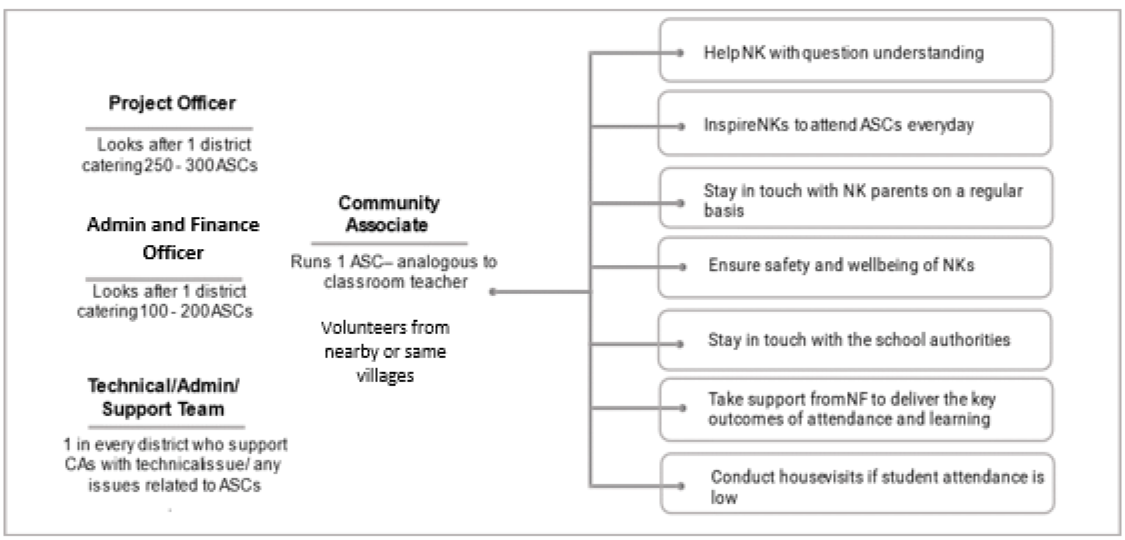
Students who received the Mindspark intervention scored 0.37 standard deviations higher than the control group at endline and improved their scores by more than double the control group over the intervention period.

While the Mindspark experiment was conducted separately from Nanhi Kali, we are interested both in the scalability of this software and in the Nanhi Kali program as a vehicle for doing so.

Program model

Nanhi Kali has approximately 20 program offices. Each program office oversees a catchment area, typically a district, and is staffed by three to four people, including a program officer, a digital learning officer and an admin and finance officer. More than half the offices also a learning and development trainer, who may provide training to CA's across multiple field offices, depending on their language capabilities.

Figure 7. Key actors in Nanhi Kali's implementation

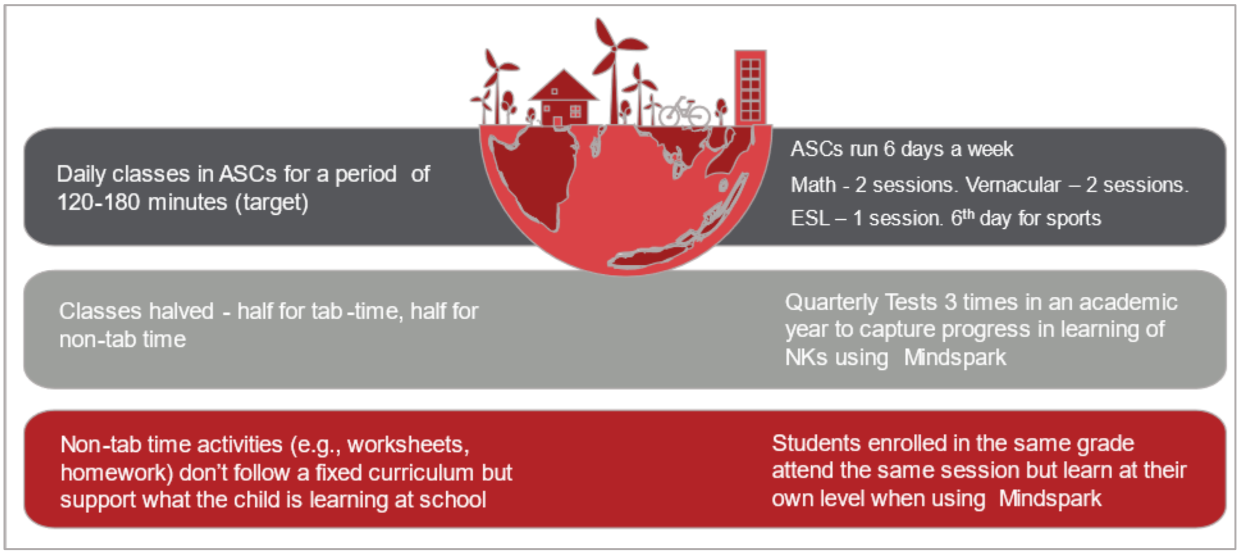


To a large degree, Nanhi Kali is supplemental to and implemented outside of the public education system[6] (i.e., it does not use government teachers and does not occur during regular school hours). The program does, however, take place in academic support centers (ASCs) that are set up within government schools. ASCs tend to be multigrade and have a maximum of 30 girls. Depending on the size of the school, one school may have several ASCs. The program runs approximately two hours a day, six days a week. For girls in

[6] The Mindspark curriculum and non-tablet based instructional time are aligned to the government curriculum. This content is delivered according to the learning level of the child, and not her official grade level.

grades 1 through 10, the program focuses on teaching three subjects—vernacular languages (two days a week), mathematics (two days per week), and English as a second language (one day per week). Sports and fitness are taught one day a week. As part of the design, the program employs "community associates"—women from local villages who speak the local languages and dialect—to oversee the ASCs.

Figure 8. Operation of academic support centers[7]



Community Associate Training and Support

Community associate training. Naandi Foundation and Education Initiatives staff provide orientation training for new community associates. Training lasts three days and includes sessions on general program operations, as well as hands-on activities using tablets and other technologies, the Mindspark application and content, and data and dashboards. In addition, new community associates spend time shadowing another ASC or community associate during orientation.

Twice-monthly review meetings. Program office staff lead block-level twice-monthly review meetings with community associates.[8] Meetings are typically held at a central school or the program office and last approximately six hours. Agendas are prepared and shared in advance through WhatsApp. Typically, one meeting a month is dedicated to training community associates on content and lesson plans, while the other meeting is dedicated to reviewing performance data, dashboards, and other administrative tasks. During the review-focused meetings, program officers and technical officers do the following:

- Show and discuss performance data, including showcasing the best ASCs, and allow community associates to share their experiences and best practices

[7] The non-tab time curriculum is designed as a bridge between the girls' learning level and her grade level. The community associate is provided lesson plans, activities, and worksheets that she uses to teach girls of similar grade levels during non-tab time.

[8] Districts are divided into blocks—typically five to eight blocks per district and 20 to 30 community associates per block.

- Provide incentives and competitions using data and dashboards to motivate community associates and improve performance among community associates and ASCs
- Allow community associates to discuss challenges and share solutions among the group
- Check and monitor ASC and community associate documents such as registers from parent-teacher association meetings, home visit reports, and participant attendance and dropout reports
- Share creative ideas and examples from different community associates and ASCs, such as unique learning activities or craft activities that aim to enhance the learning environment
- Provide reminders and tips to community associates about the program's objectives and values and about the role of community associates (especially important when there are new community associates)
- Train community associates on lesson plans and content for the upcoming month (often done by "master" community associates or community associates with previous instruction experience)

ASC site monitoring visits. Program office staff conduct site visits to monitor and support ASCs and community associates. Ideally, each ASC is visited two to three times per year. New ASCs and community associates—in addition to ASCs that are identified as "critical"—are prioritized and may be visited more frequently. During their visit, the program office staff use a digital monitoring form loaded on tablets to observe a community associate and collect data on other indicators such as average daily tablet time and student attendance. If a school has more than one ASC, the program office staff can observe more than one community associate during their visit. At the end of the visit, the program office staff provide feedback to the community associates.

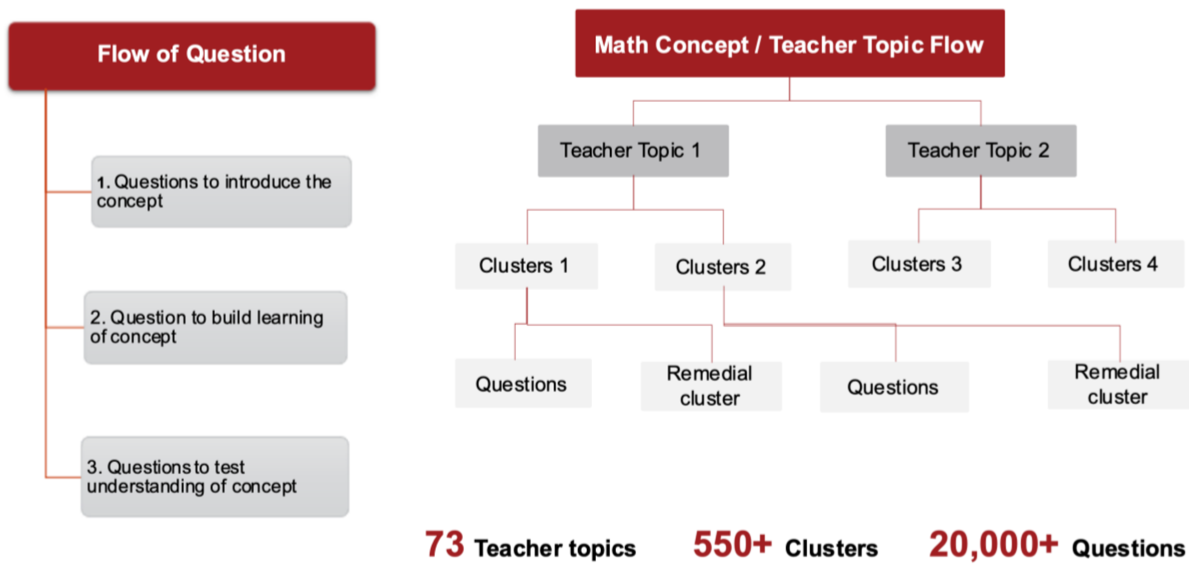
Weekly remote monitoring and communication with community associates. Program office staff monitor ASCs and community associates via weekly dashboard reports. These weekly reports are aggregated by block and shared with community associates in each block's WhatsApp group. One of the main indicators used to monitor ASCs is students' average time spent per day on tablets/Mindspark. The program staff highlight the top ten performing ASCs each week. The dashboard also identifies "critical" ASCs, which are those that have an average tab time of less than one hour per week. Program staff hold Zoom calls with community associates working at critical ASCs.

Pedagogical Approach

Math is taught twice a week for approximately two hours each session. Each session is split into dedicated tablet-based instruction (i.e., tab time) and non-tablet-based instruction (i.e., non-tab time). While half the students are working on tablets, the other half are engaged in non-tab time, which doesn't follow a fixed curriculum but includes subject-related activities such as homework, worksheets, and practice problems generally aligned with what children are learning in school. During tab time, instruction is delivered via a personalized adaptive learning software called Mindspark. Mindspark's primary method of instruction and learning is through questioning rather than delivering instruction through lecturing or modeling. Ultimately, the instructional approach focuses on practice and the application of concepts to help students move toward learning with understanding. The software doesn't teach content according to a child's grade level, so meeting curriculum standards or outcomes is not the driving force behind what children are learning. Rather, children learn at their own level and at their own pace. The software has a built-in adaptive flow that uses a child's response to decide if the child needs additional practice on a given topic or subtopic.

Mindspark covers 23 math concepts, which are broken down into "teacher topics" and "clusters." The software uses a variety of question types—including multiple-choice, fill-in-the-blank, dropdown, and interactive questions—to introduce each concept, build learning, and test understanding.

Figure 9. Mindspark math structure (grades 1–9)



Quarterly tests are held three times during the academic year to capture students' learning progress. These tests help ensure that students receive content that challenges and motivates them to learn.

Systems support

While Nanhi Kali largely operates outside of the formal education system, the program is in communication with district officials and head teachers, who give permission for the use of schools as ASCs and provide some administrative oversight.

5.1.5 RAMP

RTI International, in coordination with the Ministry of Education and its nine technical assistance partners, implemented RAMP (formally titled the Early Grade Reading and Mathematics Initiative) in Jordan, with funding from USAID and the United Kingdom's Foreign, Commonwealth and Development Office. The program ran from January 2015 through July 2023, which included two program extensions.

RAMP was a nationwide program designed to improve the reading and mathematics skills of students in Jordan from kindergarten through grade 3 (K2–G3). More specifically, the program worked with the Ministry of Education to (1) develop and distribute improved learning materials to every K2–G3 classroom in Jordan; (2) train teachers, principals, supervisors, and field directorate and Ministry of Education administrators on how to provide more effective instruction; (3) promote community participation in reading and mathematics education; and (4) support the nationwide adoption of early grade reading and mathematics policies, standards, curricula, and assessments.

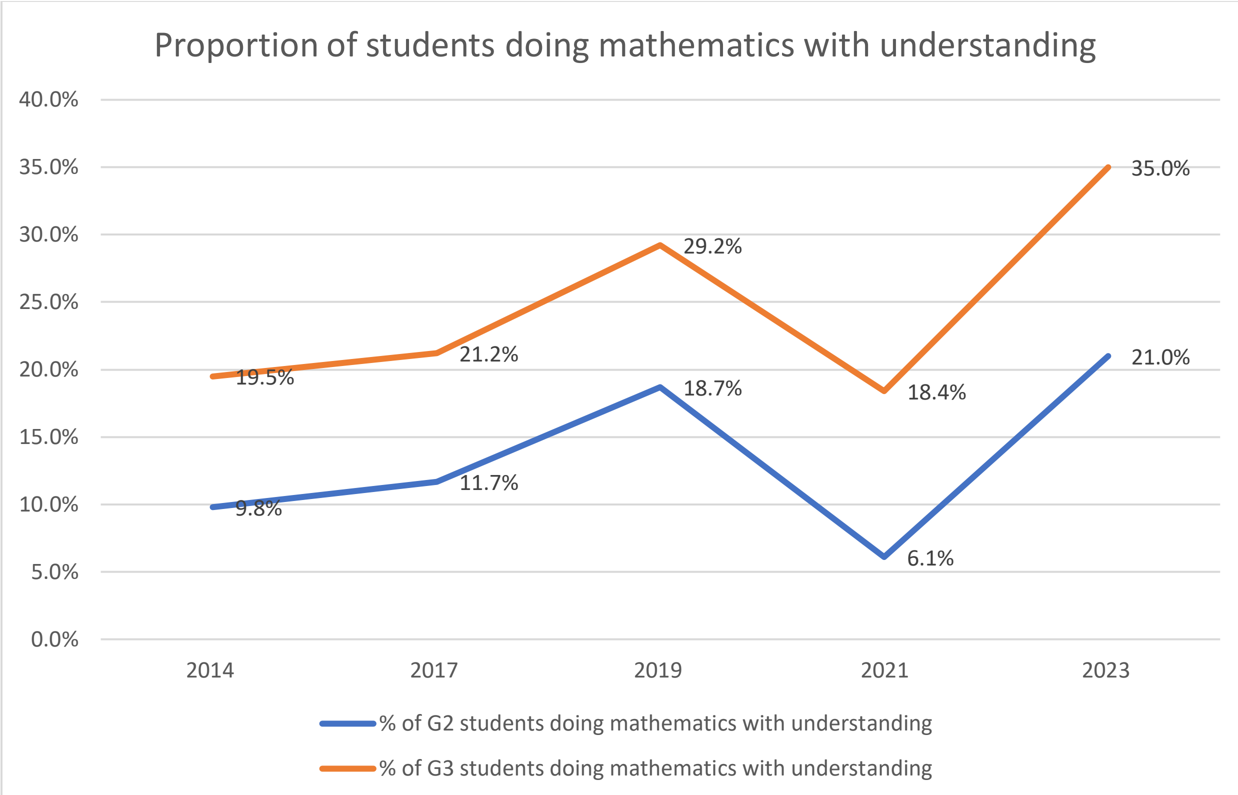
Evaluation

Since RAMP was a national-scale program from the outset, it was not possible to identify a control group for the program's impact evaluation. Therefore, evidence of impact was obtained from an internal evaluation using a pre-post design. The baseline sample consisted of control schools from a pre-RAMP intervention, with data collected in May 2014. The initial endline for RAMP (prior to the two extensions) occurred in May 2019 and included a sample of 240 schools across all 12 governorates in the county (RTI International, 2019). The evaluation included a range of early grade reading assessment (EGRA) and early grade mathematics assessment (EGMA) subtasks, while also producing a key reporting indicator on math outcomes for USAID: the proportion of students who, by the end of two grades of primary schooling, demonstrate that they can do grade-level mathematics with understanding (i.e., scoring at least at least 80% on the level-two addition and subtraction tasks combined, as well as at least 70% on missing number on EGMA). An identical indicator was also defined for grade 3.

RAMP was selected for inclusion in Numeracy at Scale based on the 2019 endline evaluation. As shown in **Figure 10**, the program had an estimated 9–10 percentage point impact in mathematics performance in grade 2 and grade 3 (i.e., a 100% increase in grade 2 and a 50% increase in grade 3). After a reduction in performance due to the COVID-19 pandemic, the second endline evaluation (in 2023)[9] showed that scores increased even further, with an overall impact of 15.5 percentage points in grade 3 and 11.2 percentage points in grade 2.

[9] Results have yet to be published.

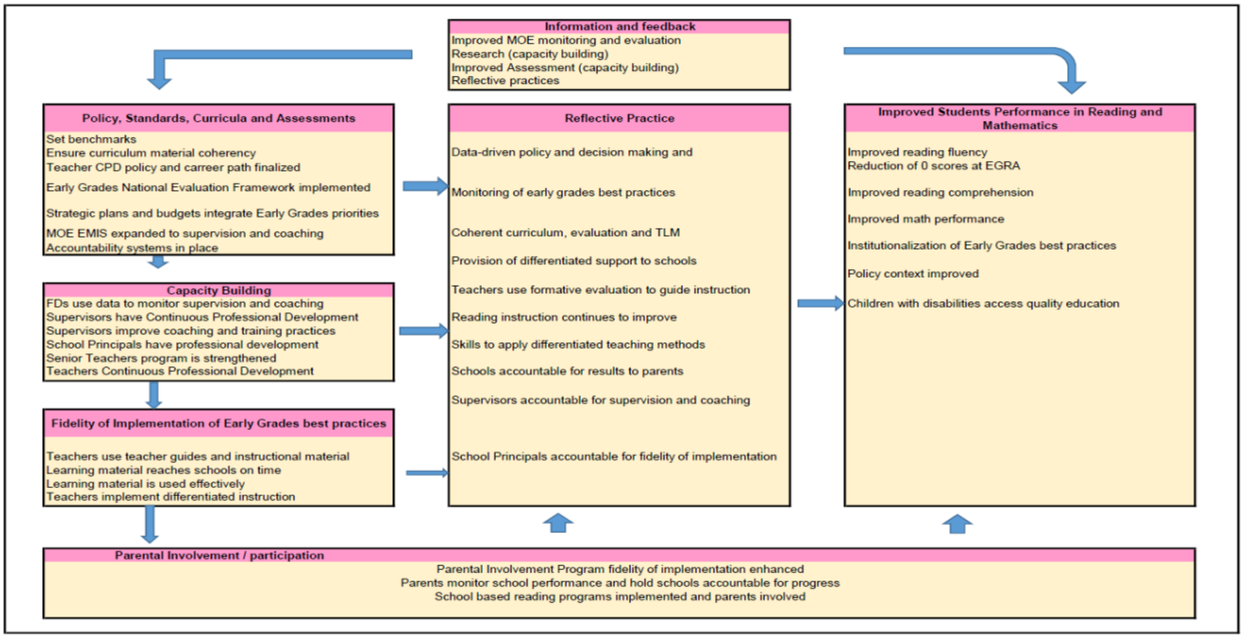
Figure 10. RAMP internal evaluation results



Program model (for mathematics)

RAMP's baseline results showed that students were performing well on procedural tasks but markedly less so on items that required both an understanding and an application of their procedural knowledge. These results aligned with the historical norm of teachers in Jordan teaching mathematics with a procedural focus, as opposed to providing students with instruction and guidance on a conceptual understanding of mathematics. Therefore, the goal of RAMP was to develop a program focusing on problem-solving skills, where students would learn to resolve complex problems requiring critical thinking, reasoning, and the application of mathematical concepts in real-life situations. This required a shift in the mindsets of teachers and supervisors. The overall theory of change for RAMP is displayed in **Figure 11**.

Figure 11. RAMP's theory of change



<u>Materials</u>

At the start of the program, RAMP focused on training and did not provide any materials beyond what teachers received from the Ministry of Education (e.g., a math textbook and blank notebook). After poor midline results on students' mathematics performance, RAMP developed structured teacher's guides, student workbooks, formative assessments, and remedial worksheets—none of which previously existed for early grade math in Jordan. All of these materials were designed to incorporate more discussion around the conceptual learning of mathematics in the classroom and to provide teachers with guidance on how to use differentiated instruction approaches based on continuous assessment. More recently, all classes have been provided with learning kits, which include counters, number cards, and number lines, among other things.

<u>Teacher Training</u>

At the outset of the program, RAMP was designed to provide a single five-day training on math instruction to all teachers at the start of the first semester. This was conducted via a cascade model, with RAMP providing a "training of trainers" to supervisors, who then trained teachers. There was then a shift in the training approach in response to a lack of improvement in students' mathematics outcomes from baseline to midline. Beginning in 2017, RAMP introduced a mathematics booster training, which exposed teachers to a new instructional approach that embraced (1) a conceptual understanding of mathematics; (2) learning progressions; (3) an initial focus on foundational skills; and (4) targeted remediation. These math booster trainings also provided a shift in training approach from lecture based to practice based.

Teacher Support

Coaching was an integral component of the RAMP model. At the start of the program, it was determined that each teacher would receive 12 coaching visits per year, provided by external coaches and government supervisors. However, math instruction was not improving as much as expected, so the program shifted to a differentiated instructional support model. Schools were targeted for additional support based on monitoring and evaluation data, including a 12-item effective instruction indicator that was co-developed with the Ministry of Education. Those items were integrated into the ministry's supervision tool, such that all early grade supervisors were tasked with collecting the same data on teacher performance. During school visits, supervisors met with teachers, observed lessons, and randomly selected a small number of students for a brief math assessment. Supervisors then engaged with teachers to discuss improvement plans and next steps. Supervisors' job descriptions were also officially changed as a result of the program. RAMP also introduced a "senior teacher" model into the system. Senior teachers were released from a few hours of instructional time in order to provide instructional support to other teachers in their school (receiving credit toward becoming a supervisor). This approach lessened the burden on supervisors to provide coaching.

Pedagogical Approach

The introduction of a focus on foundational skills was at the heart of RAMP's pedagogical approach. Prior to the program, teachers did not follow clear learning progressions that ensured mastery of foundational skills before moving on to higher-order skills. The RAMP approach introduced this progression through a range of methods, including the use of manipulatives and concrete materials prior to moving toward abstract concepts. This approach was based on teaching both procedural and conceptual problem solving for math, which was a change for teachers. The program also introduced "math talk" into the lessons, which had teachers placing a problem in front of students and asking them how to solve it, following by a discussion around student solutions and different problem-solving approaches. This was incorporated into the teacher's guides and ultimately made its way into the new Ministry of Education math curriculum. Finally, teachers were expected to employ a differentiated instruction approach, based on student needs (from continuous assessments).

Systems

Recognizing from the outset that teachers saw new instructional approaches and activities as "extra" work, the program worked alongside the Ministry of Education to ensure that the RAMP approach was integrated into the ministry's official curriculum and materials. First and foremost, RAMP introduced a reflective approach to Ministry of Education decision-making, which centered around the analysis and application of high-quality data. This collaborative, data-driven decision-making approach was consistently reported by ministry officials as one of the major factors behind RAMP's success. This approach included the use of regularly administered formative and summative assessments, as well as the introduction of diagnostic tools in the early grades. The program even established a community of practice for ministry leaders so that ministry leadership could learn and apply these approaches to other programs and discussions.

Furthermore, the Ministry of Education noted that educational materials produced by donors or NGOs do not typically have National Centre for Curriculum Development (NCCD) approval and that such materials are therefore not aligned with ministry priorities. However, RAMP engaged NCCD throughout its material development process, and when NCCD was updating the ministry textbooks, it held meetings with and sent all drafts to the RAMP team for feedback.

Lastly, RAMP focused on institutionalization from the outset, which led to high levels of sustainability. In addition to planning its program activities in conjunction with the Ministry of Education, RAMP sought to ensure that program elements were integrated into Jordan's education system to the extent possible. For example, all assessment approaches under RAMP have been adopted by the Ministry of Education and have become official policy. EGRAs, EGMAs, and lot quality assurance sampling were initially administered by RAMP but have since been taken over by the ministry. Further, diagnostic assessment tools (such as RAMP's "coarse-grain tool") are being integrated into ministry-issued textbooks. Moreover, based on RAMP's approach, the Curriculum Department and NCCD have developed a new framework for the early grades. One major component of this new framework is the integration of basic skills, which was a key focus of RAMP. The new materials also incorporate many best practices from RAMP (including "math talk" and other activities to improve students' conceptual understanding of mathematics). In addition, the ministry's national literacy strategy includes a remedial program from RAMP that extends to 2025, while the School District Development Plans also include RAMP activities. Lastly, RAMP's continuous professional development model is now official ministry policy.

5.1.6 TAFITA

The TAFITA ("**Ta**ntsoroka ho an'ny **Fita**ntananany sekoly" in Malagasy) program, funded by JICA, started in 2016. The program focuses on strengthening the capacity of school management committees to lead extracurricular remedial activities using Pratham's Teaching at the Right Level (TaRL) approach. The program includes two main parts: (1) a series of trainings to strengthen the capacity of school management committees (Farimbon'Ezaka ho Fahombiazan'ny Fanabeazana eny Ifotonyy, or FEFFIs) to develop and carry out school action plans, including plans for the TaRL remediation activities, and (2) trainings and ongoing support to local actors to implement the TaRL remediation intervention in reading and mathematics for children in grades 2–5.

TAFITA is being carried out in two phases. Under the first phase (2016–2020), the program was implemented in two regions of Madagascar—Anamalanga and Amoron'i Mania—including 2,725 schools and 288,896 students in grades 2 to 5. For the second phase (2020–2024), the program has been expanded to nine additional regions, beginning with school management support and TaRL activities.

Evaluation

JICA undertook a randomized controlled trial within the first phase of the program, within the Amoron'i Mania region, during the 2018–2019 school year (Maruyama & Igei, Forthcoming). Researchers randomly sampled 140 schools of the 1,002 public primary schools in the region (after disqualifying 96 schools because they were in highly insecure or inaccessible areas). Half of these 140 schools were assigned to the treatment group and the other half to a control group through a stratified assignment based on district, rural/urban, and school size. The baseline was conducted in November–December 2018, and the endline took place in September 2019. The series of TAFITA trainings on FEFFI formation in treatment schools began in January 2019, while the training on TaRL for reading and math was held in March 2019.

The study included reading and math assessments, as well as survey instruments. To assess math performance, the researchers utilized a math test focused on basic skills, including number identification and the four basic operations, with difficulty level ranging from single-digit addition to two-digit multiplication and division, as well as word problems (given in Malagasy). All students in grade 3 (2,687), grade 4 (2,687), and grade 5 (2,260) were tested at baseline and endline. Ending math assessment scores showed that the program improved learning for all the targeted grades. The magnitude of impact is largest in grade 3 students (0.47 standard deviations), followed by grades 4 (0.38 standard deviations), and 5 (0.36 standard deviations). All impacts are statistically significant at a 99% confidence interval ($p < 0.01$).

Subsequent to the endline, the researchers also collected data from all schools on dropout and repetition rates. These data showed that treatment schools had lower dropout rates in grades 3 and 4 and lower repetition rates for grade 5.

Program model

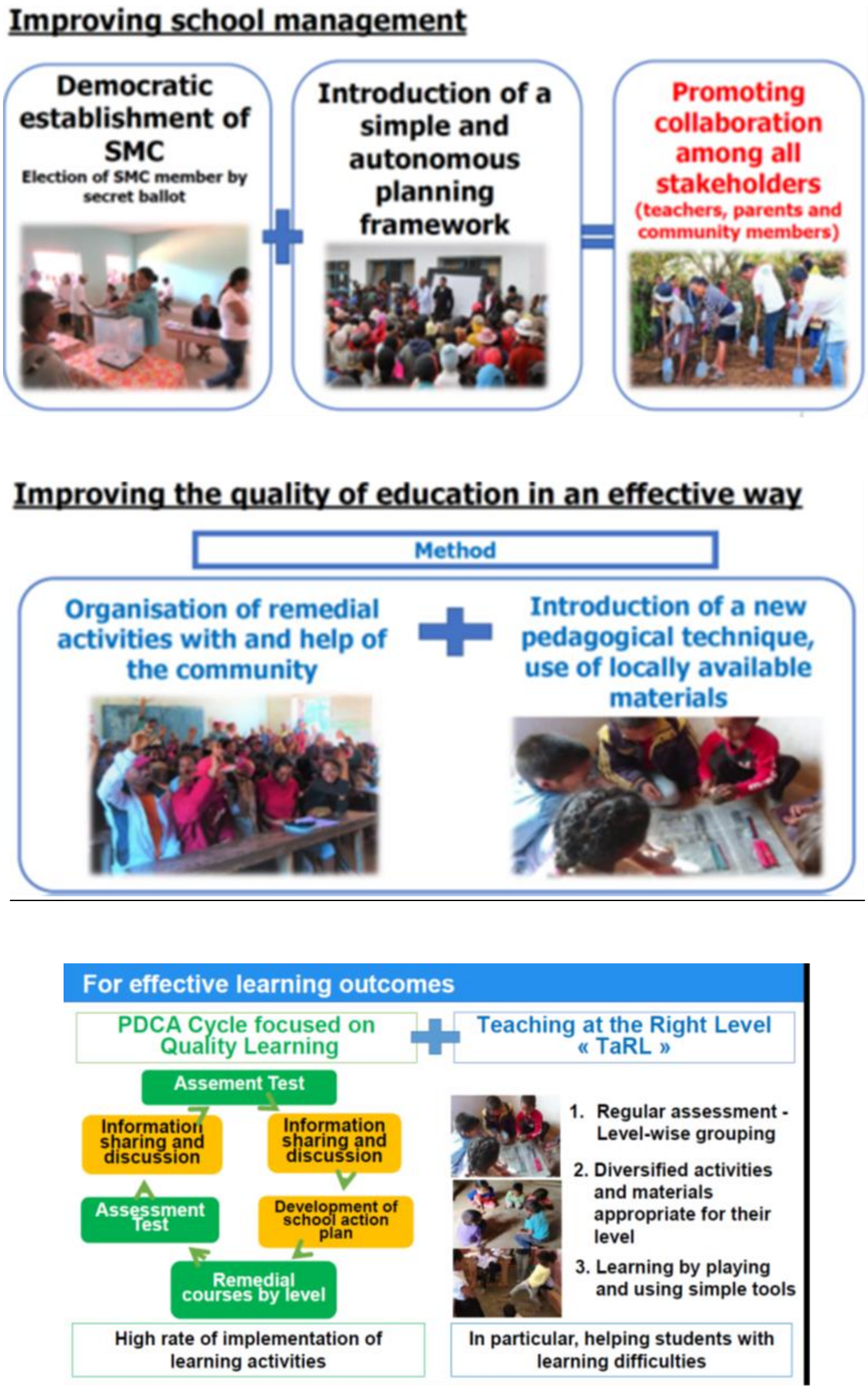
TAFITA's model is designed to work through government systems and to cover all schools in a region (aside from the experimental design period in Amoron'i Mania). In this way, ministry officials at each level are trained in both aspects of the program—strengthening of school management committees and the remediation intervention—which also provides capacity building for their core roles and responsibilities. Because many education programs in Madagascar work in select schools within a district or education zone, they often utilize parallel structures rather than system structures for implementation.

The focus on school management committees is in line with ministry policy, which states that all schools should have democratically elected FEFFIs. Having an established FEFFI and creating a school improvement plan is tied to the receipt of school grant funds. The TAFITA program was thus able to respond to the need to help ensure that FEFFIs were indeed formed using democratic processes and to strengthen their ability to develop plans that would support foundational literacy and numeracy.

The FEFFI-strengthening portion of the TAFITA program happens through a series of trainings of local-level officials, who then train school directors to carry out a democratic process for electing FEFFI officers. Once elected, FEFFI officers receive training on holding community events and developing school action plans—including preparing for, and subsequently supporting and monitoring, the after-school remediation intervention.

Figure 12 shows how the two parts of the program are intended to lead to improved quality, where collaboration among stakeholders, via FEFFIs, is strengthened, which in turn provides the foundation for implementing the remediation intervention to improve learning.

Figure 12. TAFITA's approach



Materials

TAFITA provides training materials and, most significantly, a teacher's guide that includes explicit instructions for teachers to follow when delivering the remediation program. While TAFITA does not provide teaching aids, such as posters or manipulatives, the training and teacher's guides include guidance on how to develop key aids using locally sourced materials. Examples include place value charts and bundling sticks to be used for developing students' conceptual understanding of place value and basic operations.

Teacher Training

One of the fundamental principles of the TAFITA program is that it is carried out with and through the education system. Initially, a small group of TAFITA master trainers received training directly from the Pratham TaRL team. These master trainers have since trained trainers within the Ministry of Education. The ministry provides training for district- and local-level officials—specifically, pedagogical counselors and the heads of zones (chefs ZAP)—who, in turn, train school directors and teachers on the implementation of the after-school program. This training focuses on helping school personnel be able to undertake the steps necessary to run the program, including (1) administering the simple ASER-based assessment instrument; (2) using the results to group children into four learning levels; (3) implementing the activities included in the teacher's guide, according to the given timetable; and (4) reassessing students after each two-week session and then regrouping them as they progress from one level to the next. In order to ensure that teachers are able to follow the activities in the teacher's guide, the training incorporates ample opportunities for modeling and practice.

While the training of trainers is external to the normal training program, and supported by TAFITA, the teacher training is incorporated into the professional development days (*journées pédagogiques*), which are organized by districts three times per year.

Teacher Support

While TAFITA does not provide direct coaching to teachers, the pedagogical counselors and chefs ZAP are responsible for providing ongoing support to schools and teachers, which is in line with their regular duties. TAFITA provides guidance for these officials to monitor the remediation implementation and to support school directors and teachers, as well as FEFFIs. TAFITA also supports meetings of FEFFI clusters, during which both FEFFI progress and the remediation implementation are discussed.

Pedagogical Approach

As noted above, the TaRL remediation intervention is intended to be an after-school program, consisting of ten sessions over the course of two weeks, with each session lasting one and a half to two hours. At the end of each two-week period, students are retested and regrouped, and then the ten-session cycle is repeated. Typically, schools hold four of these two-week cycles—in other words, eight weeks of remediation activities. School communities have shifted the timing of the intervention according to their local context.

The teacher's guide includes explicit instructions for carrying out the after-school sessions. Generally, sessions include time for a presentation by the teacher of a target concept and modeling utilizing the key teaching aids—such as showing how to regroup in addition using a place value chart and sticks and bundles. Then students have time to practice solving problems similarly to how the teacher has demonstrated, usually working in groups. The teacher then leads students in playing games that give them an opportunity to practice the skills they are working on.

Systems

As noted earlier, the TAFITA program is intended to be fully integrated into the education system, with system actors carrying out all key roles in its implementation. In addition to the school- and community-level FEFFI initiative, actors at each level of the system support TAFITA's implementation through clearly defined roles. Key parts of this infrastructure include regional education offices, district education offices, and the heads of pedagogical zones (subdistrict clusters of schools).

Moreover, TAFITA-related training and support functions are fully integrated in the Ministry of Education, with central, regional, and local ministry staff involved in training and ongoing support. Chefs ZAP and pedagogical counselors are trained to train teachers and to carry out school visits to ensure proper implementation of the TaRL approach.

5.2 High-Level Analysis: Program Matrix

5.2.1 Introduction

This section presents a descriptive analysis of the various elements of the Numeracy at Scale programs presented in detail in **Annex D**. We wanted to know what program design elements were evident in each of the six interventions and which of these elements were seen by programs as key to their success. These data were collected from program documents, program visits, and interviews with program teams. After we collected the descriptive data, the programs confirmed the elements presented here. We organized the program elements into five domains:

1. **Materials**—This domain describes the type of teaching and learning materials implemented in the program and the characteristics of those materials. (13 elements) Example: Program provided supplementary materials.
2. **Pedagogy**—This domain specifies the instructional approach and pedagogical methods used in the program. (10 elements) Example: Program used phonics-based instruction.
3. **Training**—This domain describes the particular types of training utilized and the elements of training design and training implementation. (13 elements) Example: Program used face-to-face initial training.
4. **Teacher Support**—This domain describes the particular coaching support structures and communities-of-practice meetings used to support teachers implementing the program. (15 elements) Example: Coaches have structured coaching tools.
5. **Systems**—This domain examines how the program works within or alongside government systems and how it seeks to change government behavior at all levels. (20 elements) Example: Program has staff at the regional level.

The six programs were asked to identify which elements were key to their success. To simplify the analysis, programs could note up to three key elements per domain.

We first describe the most frequently implemented elements in each program and then share which elements the programs designated as key.

5.2.2 Most frequently implemented program elements

We found a wide variety of elements implemented across the six programs, a reflection of both the varied design of programs and the multiple paths to effective implementation. Of those elements, 18 were reported by at least five of the programs, as presented in **Table 33**. Understanding these elements is useful for other programs hoping to implement effective large-scale programs. This section describes these program elements by domain.

Table 33. Most-implemented elements across domains

Domain	Element
Materials	Local-language materials
Materials	Learning aids for students (e.g., counters, number cards, place value materials, etc.)
Pedagogy	Focus on developing conceptual understanding
Pedagogy	Focus on developing procedural knowledge
Pedagogy	Continuous or formative assessment
Pedagogy	Instruction targeted to student level (differentiated instruction)
Pedagogy	Pair work or group work
Pedagogy	Using concrete materials and resources (manipulatives)
Pedagogy	Supporting student discussion or explanation of math concepts
Pedagogy	Using multiple models and representations
Training	Initial face-to-face training
Training	Refresher face-to-face training
Training	Structured training manuals
Training	Training for head teachers
Teacher Support	Coaches meet in groups or with supervisors
Teacher Support	External-to-school coaching
Systems	Government staff responsible for conducting monitoring
Systems	Government responsible for monitoring frequency of coaching visits

Only two *Materials* elements were frequently included. While the programs showed variation in terms of the types of math materials they provided to classrooms, almost all developed materials in the locally spoken language and invested in some type of manipulatives to support teachers' use of concrete objects to demonstrate abstract math concepts.

The *Pedagogy* domain had the most shared elements across successful programs, with eight elements reported by at least five of the programs. All programs supported teachers in using more than one representation of a concept and in using models (including manipulatives) to demonstrate or explain a concept or procedure. The programs also included elements of ongoing assessment to check for understanding and differentiated instruction. The latter was most pronounced in the TaRL design of TAFITA and in Nanhi Kali, where the Mindspark software automatically tailored math problems and explanations based on student performance. Collaborative learning in groups was another main feature across programs' design, as was enabling students' discussion of math concepts (though the latter was less evident in classroom observation data).

Several elements of *Training* were used consistently across at least five of the six programs. Program trainings were implemented as multiple, shorter trainings to avoid overloading teachers with too much content at one time. Trainings were also provided to head teachers, not only to create a sense of ownership but also to build school leaders' knowledge in the same way as teachers' knowledge. Programs used structured training manuals to increase fidelity to the key training approaches.

Only two elements of *Teacher Support* were shared by at least five programs: having coaches external to the school setting and ensuring that those coaches met with one another or school leadership to share feedback from school visits.

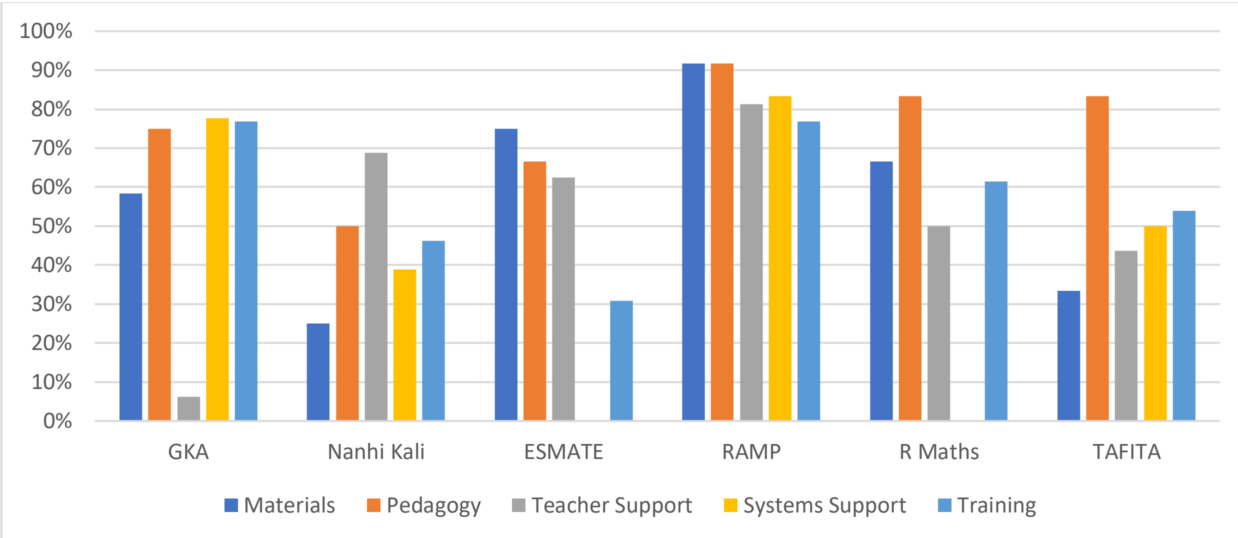
The *Systems* domain was more difficult to map, given that two of the programs were being implemented entirely by the government, while one was working almost completely outside of the education system. Of the (five) programs working entirely or partially through the system, all reported that monitoring program implementation and teacher uptake was entirely under the purview of government staff—both to conduct monitoring visits to schools and to ensure that coaches and other support staff were doing the same.

5.2.3 Program design

We used the basic information derived from each program to develop the results shown in **Figure 13**. This figure presents how many of the possible elements occurred in each program, by domain (where more elements does not inherently mean better implementation). In the subsections that follow, we indicate the elements that the programs determined as key to their success; but in this subsection, we simply present data on all the possible elements.

This figure illustrates the considerable differences in design and approach across the six programs. It also suggests what decisions were made around targeting resources and attention. The *Systems* domain is not included for ESMATE or R-Maths given their government-led nature. All of the programs were found to include at least half of the elements in the *Pedagogy* domain. ESMATE and R-Maths invested considerably in pedagogy and materials, with ESMATE focusing relatively less on training. TAFITA reported over 80% of the elements in the *Pedagogy* domain, as well as one-half of the elements in the *Systems* and *Training* domains, with relatively less focus on *Materials*. GKA focused heavily on *Materials*, *Pedagogy*, *Systems*, and *Training*—reporting over 50% of the elements for all four—with almost no inclusion of *Teacher Support* elements. Conversely, Nanhi Kali invested the most in *Teacher Support*. RAMP included over 70% of the elements in all five domains.

Figure 13. Numeracy at Scale domains, by program



5.2.4 Key elements of Numeracy at Scale programs

The previous subsection focused on the various elements found in the programs. It is also important, however, to know what elements the implementers themselves identified as key to their success. **Table 34** presents the top three elements (shaded in blue) that implementers named, by domain. This breakdown allows us to see similarities or differences in the elements that effective programs identified as being most impactful.

Table 34. Key program elements named by implementers, by domain

Category	GKA	Nanhi Kali	ESMATE	RAMP	R-Maths	TAFITA
Materials	Learning aids for students	Structured teacher's guides (scripted lessons)	Structured teacher's guides (scripted lessons)	Program materials aligned to government curriculum	Learning aids for students	Learning aids for students
Materials	Program materials aligned to government curriculum	Consumable student books	Textbook taken home	Consumable student books	Program materials aligned to government curriculum	Lesson plans
Materials	Local-language materials	Local-language materials	Student books (textbooks) for all students (1:1)	Materials developed with the government	Structured teacher's guides	-
Pedagogy	Focus on developing conceptual understanding	Continuous and formative assessment	Continuous and formative assessment	Focus on developing conceptual understanding	Continuous and formative assessment	Continuous and formative assessment
Pedagogy	Using concrete materials and resources	Instruction targeted to student level (differentiated instruction)	Instruction targeted to student level (differentiated instruction)	Using concrete materials and resources	Using concrete materials and resources	Instruction targeted to student level (differentiated instruction)
Pedagogy	Pair work or group work	Focus on developing conceptual understanding	Pair work or group work	Instruction targeted to student level (differentiated instruction)	Pair work or group work	Teacher model or explanation followed by student practice

Category	GKA	Nanhi Kali	ESMATE	RAMP	R-Maths	TAFITA
Training	Refresher face-to-face training	Initial face-to-face training	Initial face-to-face training	Teacher training (lowest level of cascade) done by government officers	Initial face-to-face training	Teacher training (lowest level of cascade) done by government officers
Training	Teacher training (lowest level of cascade) done by government officers	Refresher face-to-face training	Refresher face-to-face training	Teacher training emphasizes modeling and practice	Structured training manuals	Structured training manuals
Training	Training for head teachers	Training of trainers done by program staff	Training for head teachers	School-based training	Teacher training emphasizes modeling and practice	Residential teacher training
Teacher Support	Coaches are government staff	Caregiver and community involvement	Coaches are government staff	Coaches have structured tools	Community-of-practice meetings (across schools)	Coaches are government staff
Teacher Support	Coaches have structured tools	Community-of-practice meetings (across schools)	Coaches meet in groups and with supervisors	Coaches use tablets or other devices	Program oversight and support for coaches during school visits	Coaches meet in groups and with supervisors
Teacher Support	Program oversight and support for coaches during school visits	Virtual communities of practice (WhatsApp, SMS, etc.)	Caregiver involvement in student learning at home	Internal-to-school coaching and mentoring	Coaches are provided with program and teacher materials	Caregiver and community involvement in school management
Systems	Government staff responsible for conducting monitoring	Program uses monitoring data to make decisions about implementation	n/a	Program invested in capacity building at the decentralized level	n/a	Government staff responsible for conducting monitoring
Systems	Government uses	Program has regional staff	n/a	Program uses dashboard for	n/a	Program invested in

Category	GKA	Nanhi Kali	ESMATE	RAMP	R-Maths	TAFITA
	monitoring data to make decisions about implementation			result and data sharing		capacity building at the decentralized level
Systems	Program uses dashboard for result and data sharing		n/a	Government responsible for monitoring frequency of coaching visits	n/a	Government uses monitoring data to make decisions about implementation

Table 35 shows the 14 program elements cited by at least three programs as being key to their success. (The full analysis of program elements and key issues is provided in **Annex D**.)

Table 35. Key program elements as described by Learning at Scale programs (minimum three programs each)

Items	GKA	Nanhi Kali	ESMATE	RAMP	R- Maths	TAFITA	TOTAL
Materials	1				1	1	3
Program materials aligned to government curriculum	1			1	1		3
Structured teacher's guides (scripted lessons)		1	1		1		3
Pedagogy							
Continuous and formative assessment		1	1		1	1	4
Instruction targeted to student level (differentiated instruction)		1	1	1		1	4
Focus on developing conceptual understanding	1	1		1			3
Pair work or group work	1		1		1		3
Using concrete materials and resources (manipulatives)	1			1	1		3
Teacher Support							
Coaches are government staff	1		1			1	3
Training							
Initial face-to-face training		1	1		1		3
Refresher face-to-face training	1	1	1				3
Teacher training (lowest level of cascade) done by government officers	1			1		1	3
Systems							
Government staff responsible for conducting monitoring	1				1	1	3
Program invested in capacity building at the decentralized level				1	1	1	3

5.3 School-Level Analysis

In this section, findings are first presented for research questions 1 and 2, according to common themes and drawing on evidence from qualitative and quantitative analyses. To be considered a theme, a trend must have been observed in at least half of the programs studied. Integrated into these results are findings from the cognitive interviews with students.

Following the research questions, we present findings from a survey of teachers' mathematical content knowledge.

Because of the unique nature of the Nanhi Kali program, the research team developed a mixed-methods custom observation tool, oriented toward the facilitation and use of Mindspark. For this reason, classroom observation data from Nanhi Kali are minimal under research question 1 and are found in more detail in **Section 5.3.4**. This custom tool allowed data collectors to first focus on the ASC as a whole, then on a small group of students, and finally on two individual students (using both observation and interviews).

Additional findings are presented in **Annex A.**

5.3.1 What classroom ingredients lead to learning in programs that are effective at scale? (research question 1)

To answer research question 1, the study team analyzed data collected using the qualitative and quantitative classroom observation tools, as well as teacher interviews, described in **Section 3.3.2**. These tools are anchored in current evidence on best practices in math instruction and designed to be applicable across the early grades. Due to variations in program design and implementation, observations were conducted in grades R (kindergarten) as well as grades 2, 3, 4, and, only when necessary, grade 5. The observation tools were also used in mixed-level after-school programs in two programs. Data were collected at different points in the school year across contexts. We expect that certain practices will be more pervasive at some grade levels than others, or during certain points on the learning progression. Therefore, findings from these observations focus on identifying common instructional practices shared by multiple successful programs, rather than articulating a singular mold that all classrooms should fit into. This section presents four common themes observed across the programs.

Theme 1: Teachers use multiple representations and models to support learning.

Across all of the programs where classroom observations were done, teachers frequently used more than one representation of a concept in their instruction and used models, such as manipulatives, to demonstrate or explain a concept or procedure. During whole-class teacher explanations, materials were used during 64% of ESMATE lessons and 97% of R-Maths lessons, with other programs falling within that range. Pictorial representations were also seen in the Mindspark program used by Nanhi Kali, where models were used in problems given to students and in explanations of answers. **Table 36** shows some of the models that were used across programs during independent work time—though other models, such as number lines and pictorial images drawn on the board or in books, were also used.

Table 36. Observation: Which materials were used during independent work?

	GKA (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Counters	50%	36%	35%	63%	41%
Number cards	35%	6%	12%	21%	-
Shapes	35%	2%	4%	66%	-
Dice	21%	1%	-	5%	-
Measurement tools	21%	-	-	16%	
Place value materials	19%	9%	-	15%	5%
Number chart (1–100)	19%	-	-	-	5%
Objects for sorting and making patterns	10%	-	8%	42%	-
Objects for fractions	10%	-	2%	5%	-

An analysis of the quantitative and qualitative observation data revealed important details about how these representations were used. Across the four countries where observations were done, teachers were observed **explicitly linking representations** of a concept. **Figure 14** shows an example of this from a TAFITA session, where the teacher explicitly showed how the concrete and pictorial models for addition link to the abstract symbols.

In RAMP, ESMATE, and GKA, teachers were observed showing students how these different models could be used to solve the same problem and encouraged them to select from these models in their own problem solving. **Figure 15** shows different models used to support the development of the place value concept used by a teacher observed in ESMATE.

This connects to another aspect of the use of multiple representations found throughout the programs. In all programs, **students used the concrete materials and other models themselves**, rather than simply observing the teacher doing so. Qualitative interviews from South Africa revealed that teachers highlighted students' use of manipulatives as essential to their instruction. Teachers said that using simple, concrete objects can help students play, do practical activities, and learn math concepts.

Concrete materials were used during students' independent work during 57% of ESMATE lessons and 98% of R-Maths lessons, with other programs falling within that range. In ESMATE, RAMP, and R-Maths, materials used during independent practice were most often in the hands of all students (**Figure 16**). In TAFITA, materials were most often provided to a subset of students, which may be explained in part by the resource-poor environment in which students shared materials while working in groups. While materials were most often in the hands of teachers in GKA, in two-thirds of lessons, materials were also used by all students.

Figure 14. Example of TAFITA session

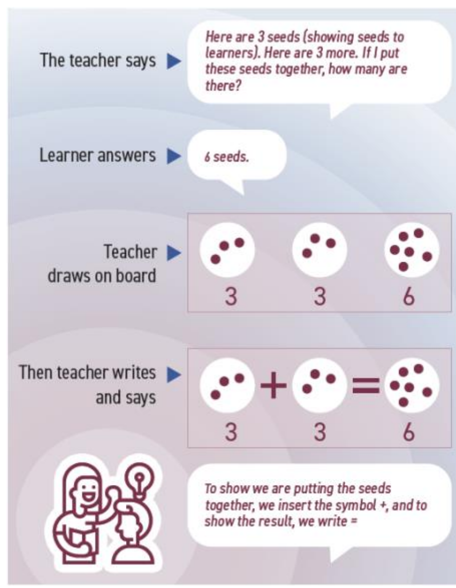


Figure 15. Models supporting development of ESMATE place value concept

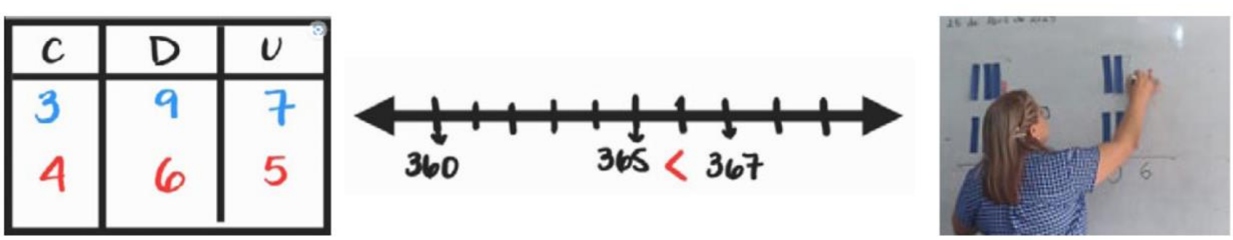


Figure 16. Observation: Who used materials during independent work?

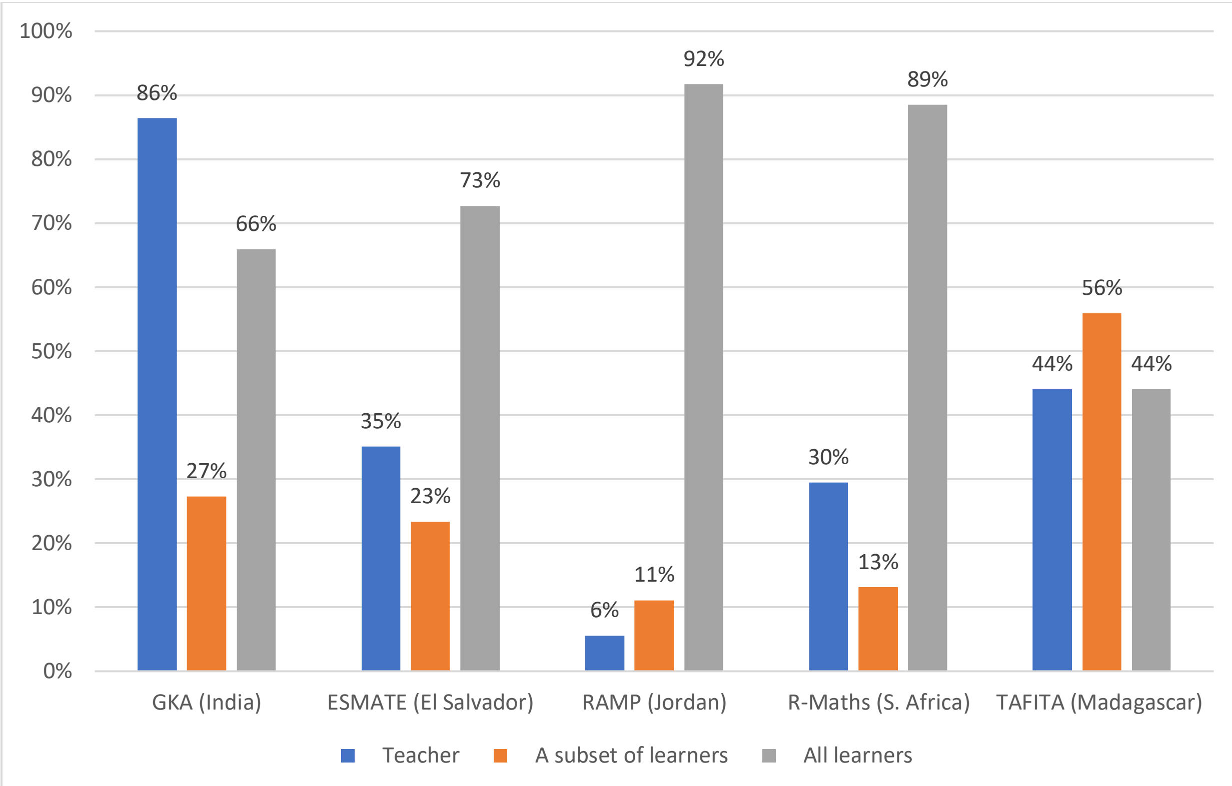


Figure 17 shows students in the RAMP program using a variety of materials and strategies to solve problems.

Figure 17. Materials and strategies used in RAMP

This emphasis on students' use of models, and its impact on students' ability to learn problem solving, is also reflected in the cognitive interviews with students. Across programs, students were observed selecting from and using materials to aid in their problem solving. **Table 37** shows the percent of students who used representations (either counters, drawings, or fingers) to solve simple and complex operations problems. Students used these most often for simple addition and subtraction problems, and a majority of students were highly successful with this strategy. For simple addition, 93% of students answered correctly in RAMP, 73% in ESMATE, and 66% in TAFITA.

Table 37. Percentage of students who used counters, drawings, or fingers to solve problems

	ESMATE (El Salvador)	**RAMP (Jordan)**	**TAFITA (Madagascar)**
Simple addition (13 + 6)	76%	47%	64%
Simple subtraction (16 – 4)	59%	42%	68%
Complex addition (38 + 26)	32%	9%	59%
Complex subtraction (52 – 37)	18%	4%	52%

For the more complex operations problems, fewer students used representations, which is expected given that these problems require a different, more abstract strategy. Fewer than half of students in ESMATE and TAFITA solved these problems correctly; in Jordan, just about half of students answered the complex addition problem correctly, and less than half answered the complex subtraction problem correctly. It is important to keep in mind that all students were interviewed at different points in their school year, and the complex addition and subtraction topics may not yet have been covered. Nevertheless, the results point to the sustained use of representations and models to solve problems and suggest that students were in the process of moving to more abstract strategies.

Theme 2: Instructional approaches include a specific focus on both conceptual understanding and procedural fluency.

Data from classroom observations across programs indicate an emphasis on ensuring that students develop a conceptual understanding beyond just completing the same problems modeled by the teacher. This theme manifested in a number of ways, with some variation across programs.

One way in which teachers supported conceptual understanding was in their approach to questioning and the way in which they responded to incorrect answers. In TAFITA, RAMP, and ESMATE, teachers asked at least some **open questions where there was more than one correct answer**. When students gave a wrong answer, teachers in three of the five programs where observations were done were more likely to **help the student find the correct answer or discuss why the answer was incorrect** (see **Table 38**). While this evidence is promising, it is also important to note that many teachers also responded, at different times, by telling the student to try again (without help) or by simply giving the correct response. While using a myriad of strategies to respond to incorrect responses is expected, it may be that teachers in some programs, such as GKA and ESMATE, were still working on improving their ability to provide feedback to students.

Table 38. Teachers' responses to incorrect answers during whole-class instruction

	GKA (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Helped student(s) solve through questions, modeling, or clarifying questions	48%	45%	62%	70%	61%
Discussed why answer was incorrect	60%	33%	66%	21%	33%
Asked another student to answer	45%	43%	47%	41%	37%
Asked student(s) to solve the same problem again without any additional help	79%	23%	48%	25%	43%
Gave correct response	88%	53%	34%	25%	41%
Ignored incorrect response and moved on to a new question	5%	4%	0%	0%	2%

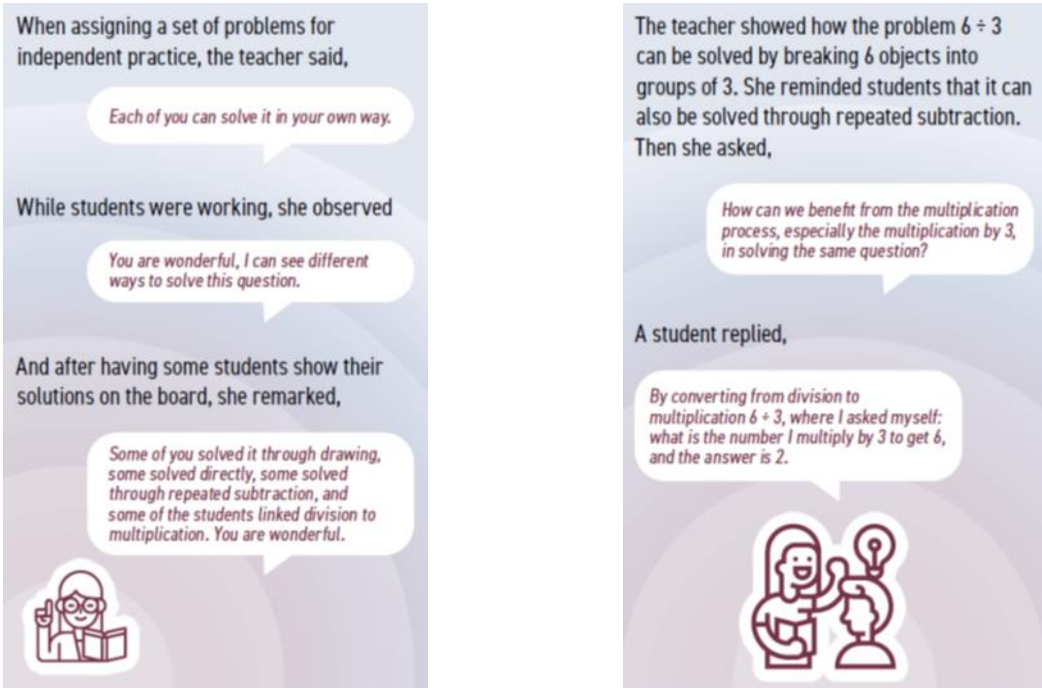
In the majority of programs, there was an emphasis on encouraging students to **use multiple strategies in problem solving**, and in some programs to **discuss their mathematical ideas**. Between one-half and three-quarters of teachers in four of the programs said that they focus more on using multiple strategies to explain and solve problems since the programs started (**Table 39**).

Table 39. Teacher interview: How has your regular class instruction changed since you started working with this program?

	GKA (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
More focus on using multiple strategies	76%	52%	64%	61%	46%
New methodology or instructional approach	48%	26%	60%	36%	29%

Lessons observed in the RAMP program perhaps exemplify this approach most clearly. In all lessons observed in RAMP, teachers explicitly encouraged students to use multiple strategies when problem solving and also asked students to discuss how they approached the problems. **Figure 18** provides some examples.

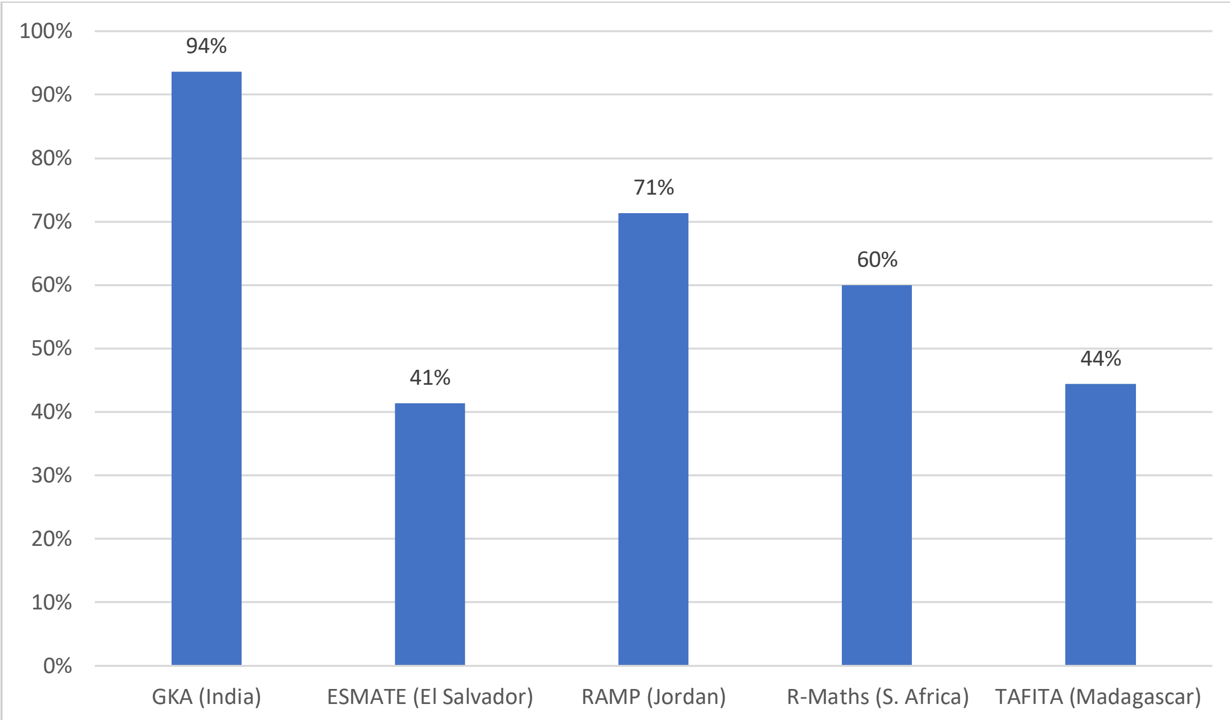
Figure 18. Examples of teacher encouragement of multiple strategies and discussion of a problem



The cognitive interviews with students reflect the approaches that teachers used for modeling, explaining, and discussing mathematical ideas. A majority of the students interviewed were able to describe how they solved problems. Across the different problems, students used a variety of strategies. For example, a majority of students in the RAMP program, while solving double-digit addition, applied place value concepts, while some used decomposition (breaking the numbers apart to make them easier to add) and did not use concrete materials—whereas for world problems, the majority used counters.

In addition, across all programs, teachers frequently made **connections between math concepts and either the real world or students' experiences**, which can support student's conceptual understanding and their ability to apply mathematics concepts to novel and real-world situations. In GKA, RAMP, and R-Maths, these connections were observed in over 60% of lessons (**Figure 19**).

Figure 19. Percent of lessons in which the teacher connected math concepts to real-life examples or the lives of students, by program



Theme 3: Various approaches are used to ensure active student engagement throughout lessons.

All of the programs had strong student engagement, as shown in **Figure 20**. In virtually all observations, the majority of students remained engaged for the duration of the lesson, and, often, all students stayed engaged throughout.

Figure 20. Percentage of lessons in which all, most, or few students were engaged for the whole time, by program

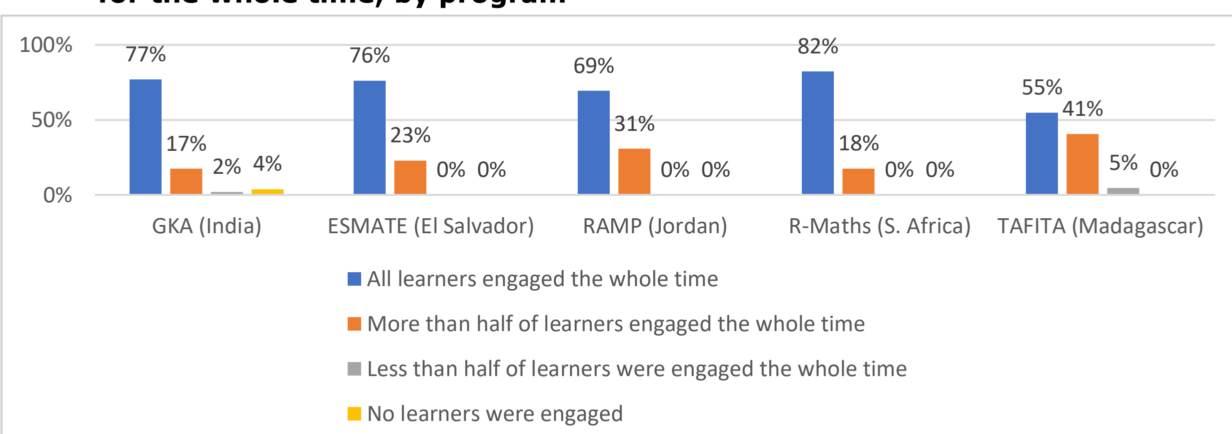
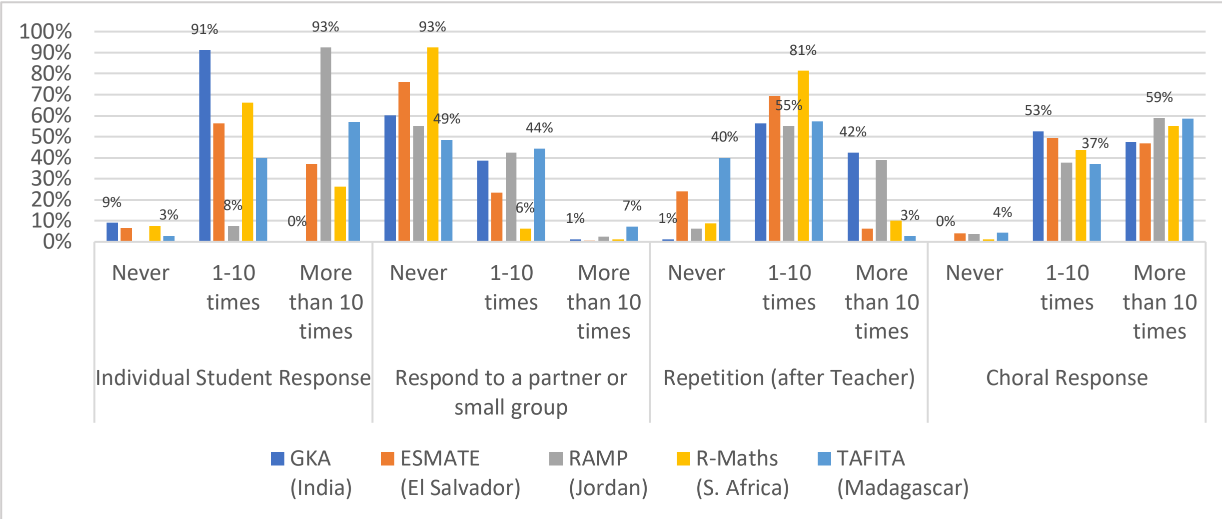


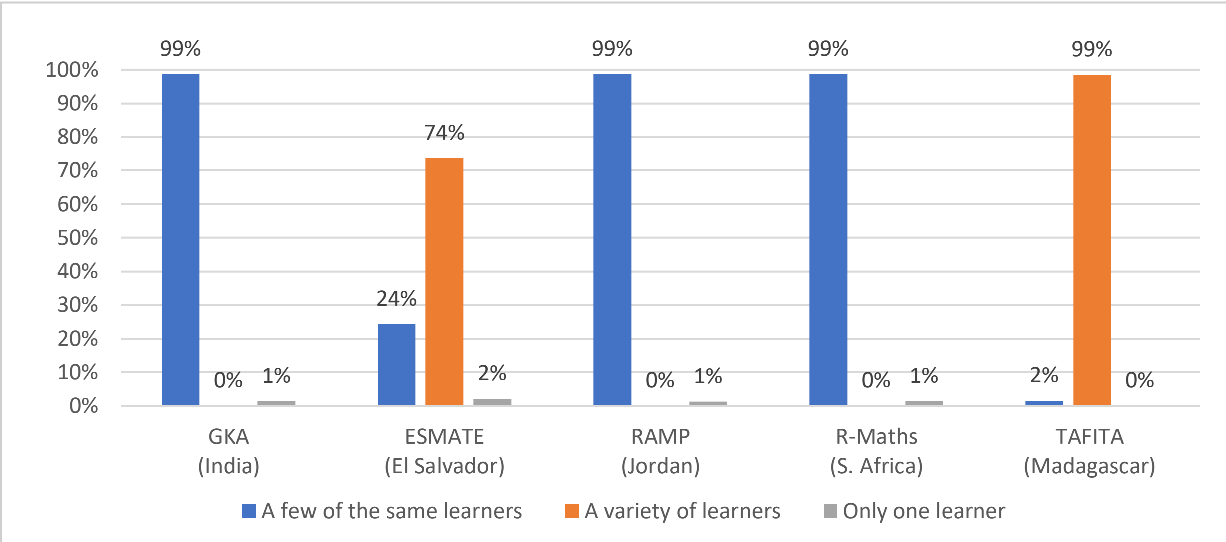
Figure 21. Percentage of lessons in which each response type was observed: never, 1–10 times, or more than 10 times



The approaches used for ensuring the engagement of all students varied across programs. **Figure 21** shows how teachers across programs used **a variety of questioning techniques**. This figure shows different types of question and response techniques observed during the whole-class portion of math lessons, by program. "Individual student response" means that a student was called on for an answer or asked to come to the board and solve a problem. "Respond to a partner or small group" includes opportunities for "think-pair-share." "Repetition" refers to any time the teacher asked students to repeat after her as she explained or modeled a math concept. Repetition involves the lowest level of cognitive engagement, followed by choral response. Across programs, almost all lessons included some opportunities for individual students to provide a response during the whole-class portion of the lesson. In particular, 93% of RAMP lessons included more than ten opportunities for an individual student to give an answer. While there was far less utilization of partner or small-group response, for three programs (GKA, RAMP, and TAFITA) these opportunities were present in over one-third of lessons.

Although opportunities for individual students to respond were prevalent across programs, teachers often called on a few of the same students to solve problems or answer questions (**Figure 22**), which reduces opportunities for engagement. In the majority of ESMATE lessons and virtually all TAFITA lessons, however, teachers called on a variety of students throughout.

Figure 22. Percent of lessons in which the teacher called on one student, a few of the same students, or a variety of students, by program



While the level of engagement represented through questioning varied, all programs included some focus on having **dedicated time for independent and group work**, whether independently or in groups. As **Table 40** shows, all programs dedicated at least 25% of the lesson time to independent or group work in over half the lessons observed. In ESMATE and R-Maths, this reached almost 50% of the time in the majority of lessons.

Table 40. Independent work, by program

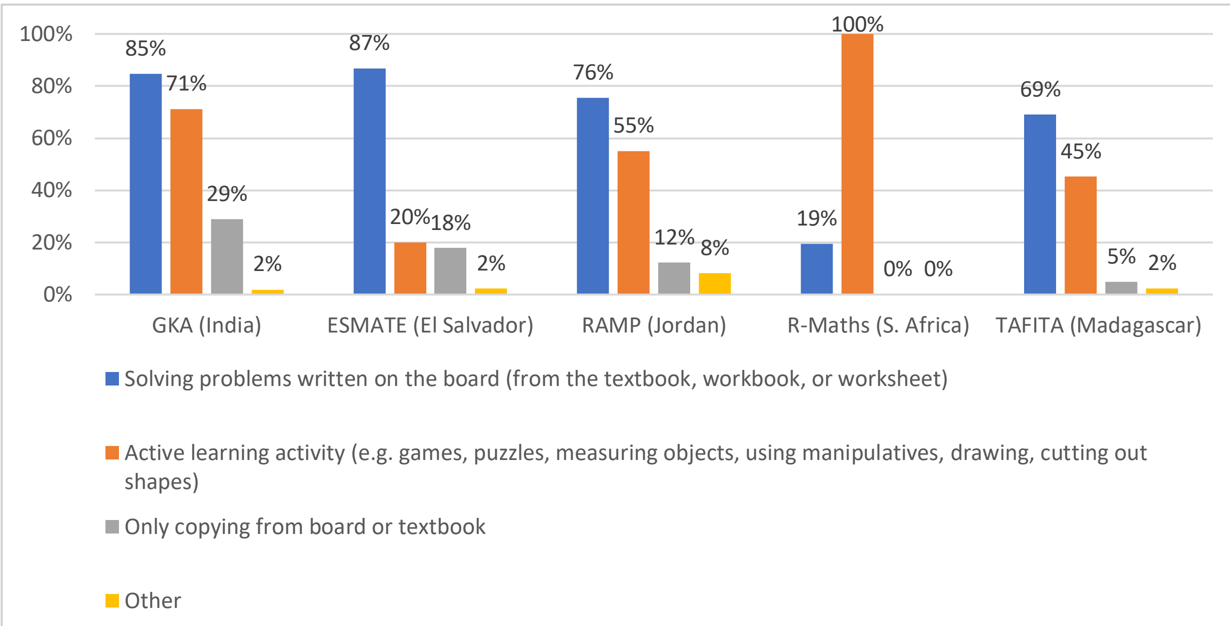
		GKA (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA[10] (Madagascar)
Independent or group work: Students working on their own	Percent of lessons (frequency)	58%	86%	61%	78%	57%
	Average time (duration, in minutes)	10	19	10	18	11
	Average % of total class time	25%	43%	29%	40%	27%

In most programs, a substantial amount of **independent or group work time was spent in active learning**, which could include using manipulatives, playing a math game, and measuring or cutting out and manipulating shapes, among other things, as shown in **Figure**

[10] The TAFITA findings presented under research question 1 are from observations of the after-school program only. There was little evidence in the quantitative observation data of spillover from the after-school sessions into the classroom.

23. Most of the programs also used this time for students to practice solving problems, typically similar to those that teachers had modeled or explained earlier in the lesson. In the RAMP program, for example, qualitative observations showed that practice time was generally used for students to work through problems together, and then the teacher provided worksheets with additional similar problems for them to gain further practice.

Figure 23. Tasks assigned during independent time, shown in percent of lessons where each task was observed, by program



In teacher interviews, teachers across programs noted that this emphasis on active learning and group work was something that had changed since the program began. Over one-half of teachers interviewed in GKA, ESMATE, and R-Maths noted an increased focus on student-centered exploration and problem solving. Between 29% and 49% of teachers in all five programs responded that their instruction now included more active learning and peer learning (**Table 41**).

Table 41. Teacher interview: How has your regular class instruction changed since you started working with this program?

	GKA (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
More focus on having students explore and solve problems	79%	57%	40%	54%	33%
More active learning less lecture	29%	47%	33%	43%	49%
More pair and group work	29%	47%	33%	43%	49%

Additionally, qualitative observations across programs noted that the use of interactive, fun activities, such as games and songs, garnered enthusiasm among students. Examples of active learning in TAFITA lessons are shown in **Figures 24 and 25**.

Figure 24. A group solves problems using sticks, bundles, and place value chart

Figure 25. Teacher shows students how to play a number game and then ensures that all students get a turn

Theme 4: Teachers use assessment-informed instruction approaches to address differentiated needs.

In all programs, teachers were observed using assessment-informed instruction—that is, monitoring students' progress and changing instruction or providing additional support to students who struggle (or more advanced tasks for students who excel). The strategies teachers used to do this varied somewhat across programs.

One strategy that was used in all programs was for the teacher to **monitor students while they worked** during independent and group work. This was observed in both qualitative and quantitative observations, with teachers monitoring students while they worked between 88% (TAFITA) and 98% (R-Maths) of the time. In qualitative observations, teachers in ESMATE, TAFITA, and RAMP were observed at some point **helping students who appeared to be struggling**. This was also approximated in quantitative observations by recording when teachers spent more time with one student or group and appeared to be helping them; this practice was noted in around 40% of lessons in TAFITA, ESMATE, and R-Maths.

The use of **formative assessment**, whether formal or informal, was also found across programs, but the form and use of results varied. Some of the programs **included a focus on assessment and use of results to teach students according to their level**. In TAFITA and Nanhi Kali, formative assessment is a key feature of each program, but the way it used is very different. In TAFITA, as noted in **Section 5.1.6**, students are periodically given a formal assessment, and then students are grouped based on the results. Each section of students is then taught according to their level. In Nanhi Kali, the Mindspark app applies assessment-driven instruction in a different way. In this app, each question posed is driven by the student's performance on the previous question—providing guided practice at the student's individual level. Under RAMP, qualitative observations noted that some teachers also grouped students and then gave them different tasks; in post-observation interviews, these teachers noted that they had used assessments in class to guide the selection of tasks.

Teachers monitored students while they worked between 88% (TAFITA) and 98% (R-Maths) of the time. Under three programs—ESMATE, R-Maths, and TAFITA—over 40% of teachers observed spent more time with one student or group, helping them solve a math problem or complete an activity. For R-Maths and TAFITA, this finding—which suggests differentiation of the lesson, as well as an increased focus on individual students—aligns with the instructional design of the programs (as described in **Section 5**).

In four of the programs, when students finished activities early, teachers most often gave them a different math activity to work on or more of the same problems to solve. Few teachers gave students nothing to do in these situations.

Table 42. Teachers' responses to students who finished an activity early, by program

When students finished early, the teacher gave them…	GKA (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
a different math activity or problem	39%	34%	52%	9%	78%

When students finished early, the teacher gave them…	GKA (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
more of the same problems to solve	32%	48%	21%	34%	4%
other non-math-related work	27%	6%	24%	28%	0%
nothing	2%	1%	0%	19%	0%
other	0%	10%	2%	11%	19%

When we asked teachers what part of their instruction has had the biggest impact on student learning (**Table 43**), responses were varied, but some commonalities emerged. The most popular responses among teachers in GKA, ESMATE, and R-Maths was an increased focus on having students explore and solve problems (**theme 1**). Roughly one-third of teachers in GKA and RAMP said that more focus on multiple strategies had the greatest impact (**theme 2**). Other teachers saw the greatest impact coming from more student-centered instruction, more pair or group work, and the use of more materials (**theme 3**).

Table 43. Teacher interview: What part of your instruction has had the biggest impact on student learning?

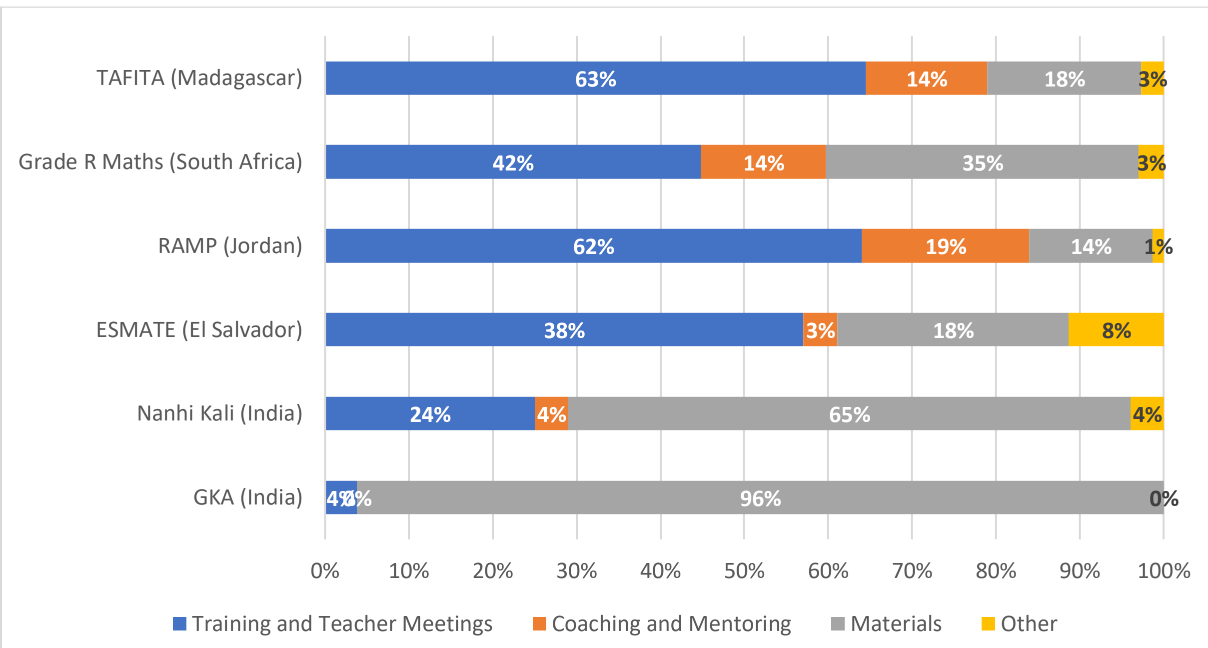
	GKA (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
More focus on having students explore and solve problems	35%	43%	22%	27%	7%
More focus on using multiple strategies	32%	17%	39%	19%	5%
More student-centered, less lecture	8%	8%	14%	22%	24%
Involves more materials or activities	9%	9%	8%	6%	**32%**
New methodology or instructional approach	5%	13%	10%	8%	25%
More pair or group work	11%	9%	8%	13%	5%
Other	0%	1%	0%	5%	3%

5.3.2 *What methods of training and support lead to teachers adopting effective classroom practices? (research question 2)*

> **NOTE ON TERMINOLOGY**
>
> Throughout this findings' section, we use the term "teachers" to refer to the main actors who delivered instruction under each program. As noted above, teachers spanned grade R (R-Maths) to grades 4 and 5 (GKA). Additionally, Nanhi Kali engages women from local communities—called "community associates"—to deliver instruction. In this section, these community actors are also included under the umbrella term of teachers.

Each of these programs provided teachers with a range of supports. When asked what the single most useful support was for improving their math instruction, teachers in four of the programs most often cited training (and, in some cases, teacher meetings) (**Figure 26**). In GKA and Nanhi Kali, the majority of teachers saw teacher and student materials as being the most useful support they received from the program. For GKA, this included the teaching and learning materials kit, and for Nanhi Kali, it included the tablets with Mindspark software. Materials were also frequently cited by teachers in R-Maths (35%), ESMATE (18%), and TAFITA (18%). It is worth noting that teachers may see materials and training as linked—in other words, that one may not work without the other. Coaching was seen as the single most useful support by 19% of teachers in RAMP, 14% in R-Maths, and 14% in TAFITA.

Figure 26. Most useful program supports reported by teachers, by program

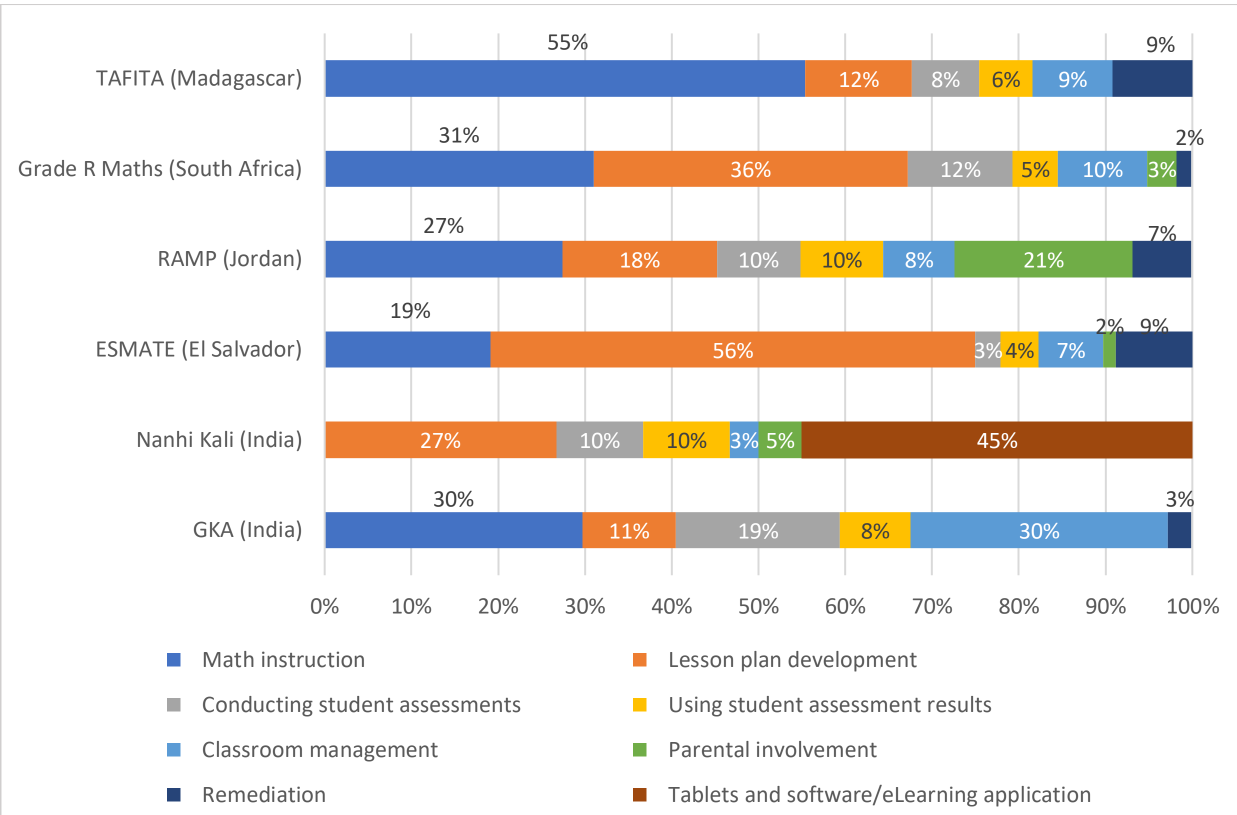


Theme 1: Teacher supports focus explicitly on math content and improving instruction.

While the design of teacher support models varied greatly across programs, all prioritized building teachers' pedagogical math knowledge. Between 46% and 92% of teachers across programs cited that they received training on math instruction. Relatedly, between 37% and 86% of teachers also reported a focus on lesson plan development (see full table in **Annex A**).

Teachers were then asked what training content was most useful to them (**Figure 27**). Consistent with the findings above, math instruction and lesson plan development were deemed the most useful training content by teachers in four of the six programs. (Tablets and software/eLearning applications were reported as the most useful training content area by 45% of community associates in Nanhi Kali, due to the tablet-based approach employed by this program.) GKA teachers saw math instruction and classroom management to be the most helpful part of the training content, while 21% of teachers in RAMP also reported parental involvement as a useful content area.

Figure 27. Most useful training content areas, as reported by teachers, by program



Teachers were then asked what they believed to be the most important overall differences between these program training sessions and other teacher training sessions they had attended in the past. The most commonly reported factors are presented in **Table 44**. In addition to overall agreement across programs that trainings were better organized in their current programs, teachers in four programs frequently cited a greater focus on specific math skills (ranging from 28% to 51% of teachers). Teachers in three programs also noted that trainers in current programs were more knowledgeable (ranging from 26% to 57%). Roughly one-half of GKA, TAFITA, and R-Maths teachers also reported having more time to practice (individually or in pair groups) and that training materials were more relevant and helpful.

Table 44. Most important difference between program trainings and others, as reported by teachers, by program

Overall, what do you see as the most important differences between this program's training and other teacher trainings you have attended?	GKA (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
More time to practice (individually, in pairs, or in groups)	54%	27%	14%	43%	52%
Training is better organized	46%	46%	52%	36%	34%
More focus on specific math skills	51%	28%	32%	31%	32%
Trainers are better prepared or more knowledgeable	57%	32%	26%	33%	39%
Materials are more relevant or helpful	62%	20%	22%	50%	48%
More time for discussion	49%	24%	27%	47%	29%
Trainings are more frequent	46%	13%	25%	17%	3%
Less lecture	8%	7%	23%	3%	40%
Better allowances	11%	2%	11%	9%	2%
Expectations are clear	5%	27%	4%	26%	32%

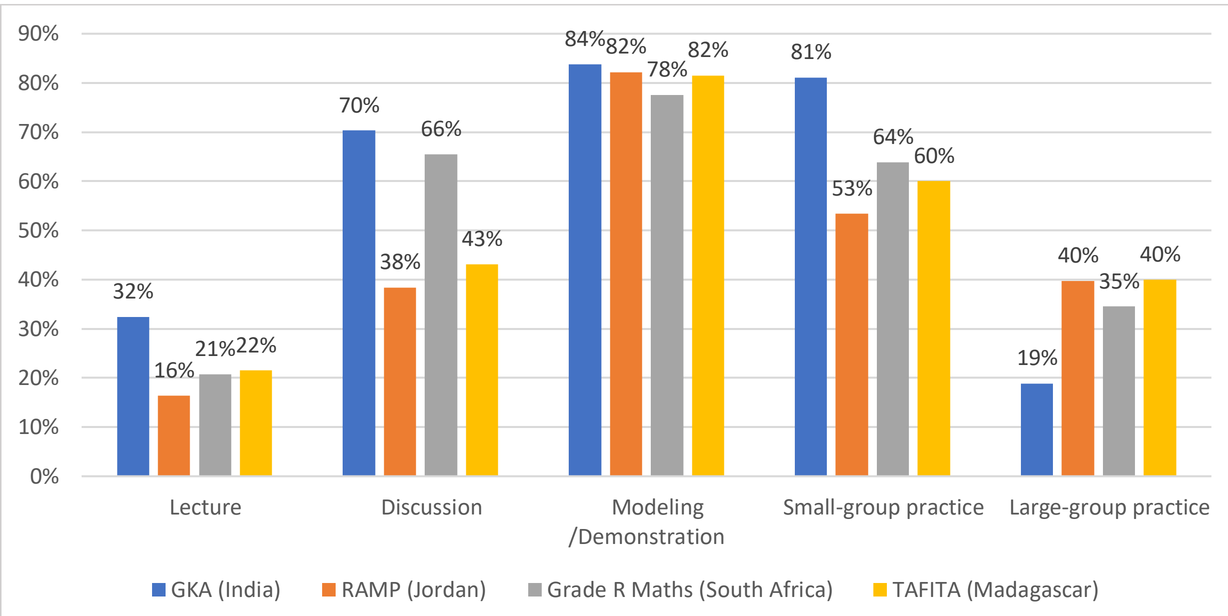
Gray cells indicate differences cited by more than one-third of teachers interviewed.

Theme 2: Trainings emphasize modeling and practice over lecturing, providing teachers with opportunities to practice and discuss.

Teachers across all programs reported that training sessions in their current programs used more modeling and demonstration (ranging from 77.6% to 83.8%), small-group practice (ranging from 53.4% to 81.1%), and discussion (ranging from 38.4% to 70.3%) than previous teacher training sessions they had attended. By comparison, far fewer teachers noted an increase in lecturing (**Figure 28**).

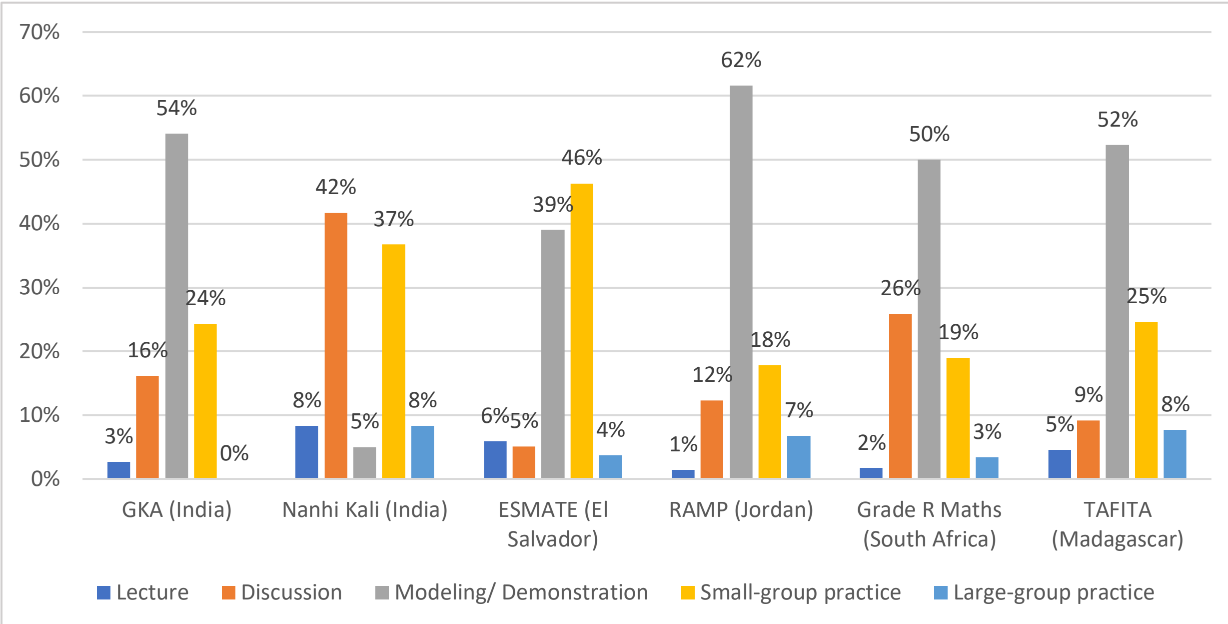
As we interpret these findings, it is important to note that the term "training" is used broadly here and may look quite different across programs. Some trainings may be held at the school, district, or department level, while others are larger and more centralized. In some cases, we believe that teachers interpreted teacher meetings as trainings when responding to the survey questions presented here.

Figure 28. Percentages of teachers reporting on the training methods that the program trainings employed more of compared to previous trainings



When asked which of these training methods was most useful, over one-half of teachers in four of the programs (GKA, RAMP, R-Maths, and TAFITA) cited modeling and demonstration. The second most popular response was small-group practice, cited most often by teachers in ESMATE and Nanhi Kali. Teachers in R-Maths and Nanhi Kali also frequently cited discussion as being most useful (**Figure 29**).

Figure 29. Most useful training methods, as reported by teachers, by program



Theme 3: Teacher and student materials provide explicit guidance for instruction.

We asked teachers how the teaching and learning materials they received under the six programs compared with those materials used previously.

Figure 30 maps teachers' perspective on the teaching materials they were given, by program. Generally, teachers across all programs reported that teacher materials, compared to past programs, were better organized and easier to follow; included better activities and examples; and were more engaging. Teachers in ESMATE and R-Maths also emphasized the step-by-step instructions, while teachers in GKA emphasized the materials' alignment with the curriculum.

Figure 30. Comparison of program teacher materials with prior materials, as reported by teachers, by program

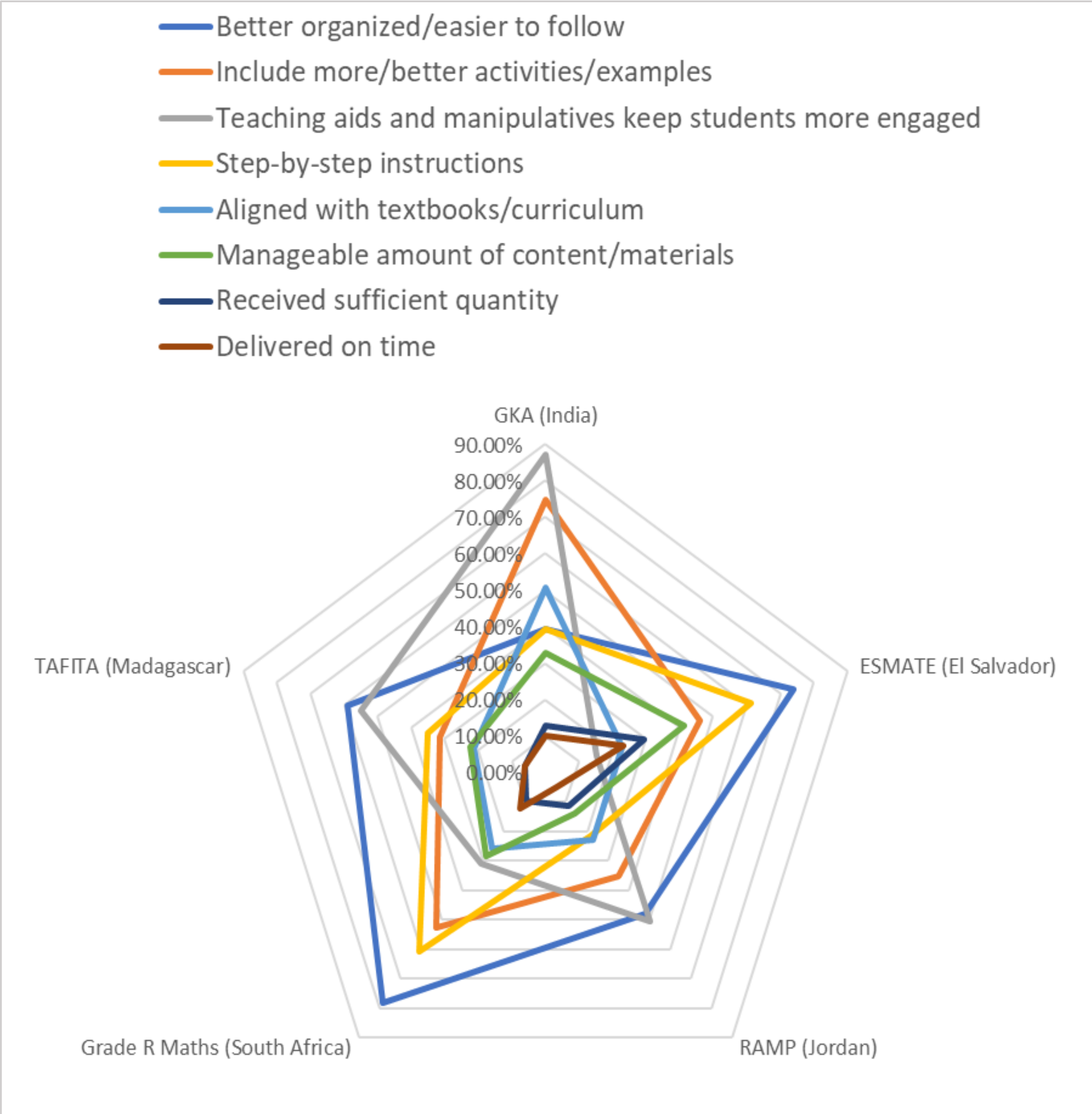
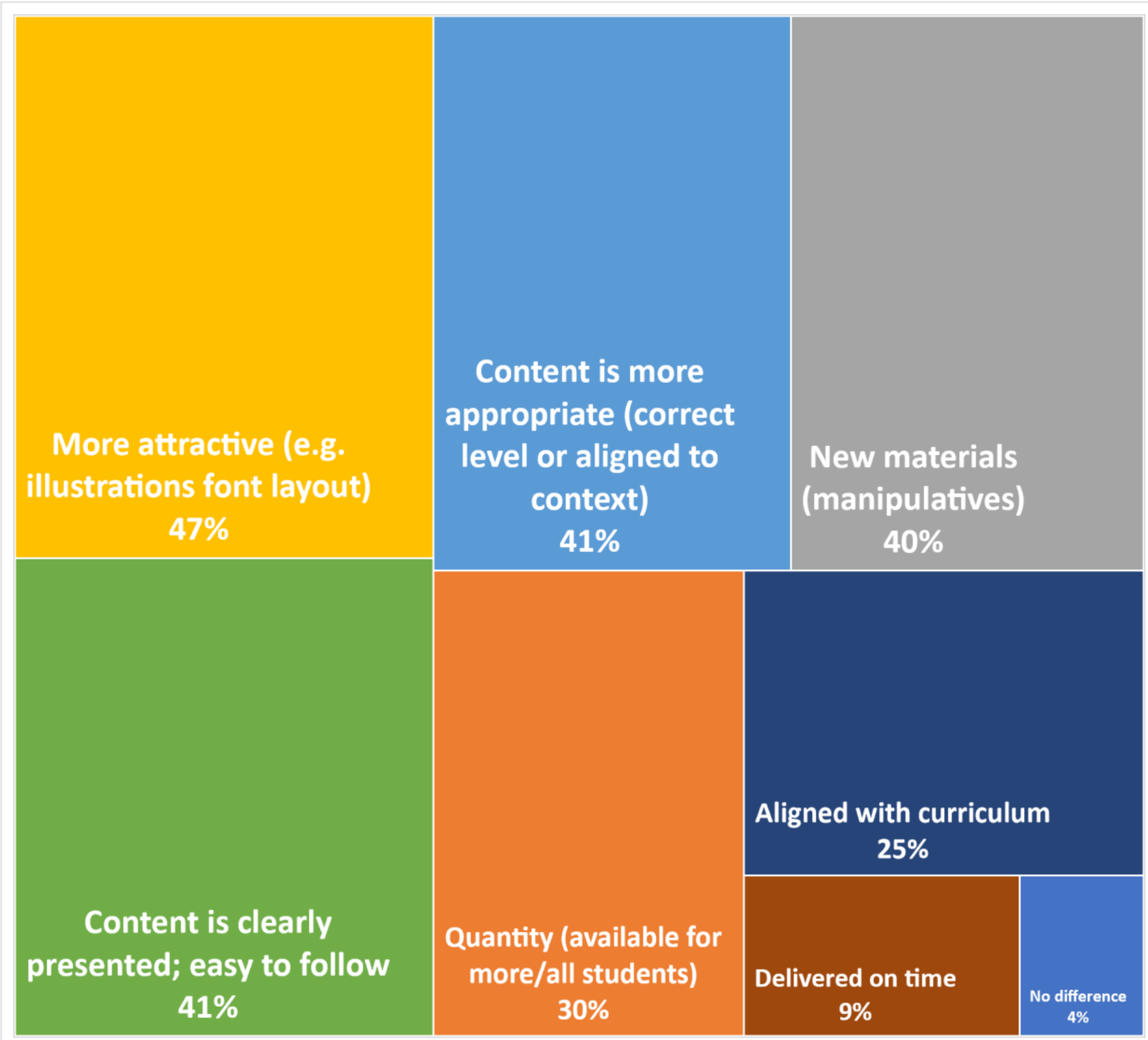


Figure 31 presents the overall average percent of teacher responses to a similar question about student materials. Teachers most often reported that the student materials they received under these programs, compared with previous materials, were newer and more attractive and that the content presented was easier to follow and more clearly aligned to the curriculum and context.

Figure 31. Comparison of programs' student materials with prior materials, as reported by teachers: Average across programs



Responses by program can be found in Annex A.

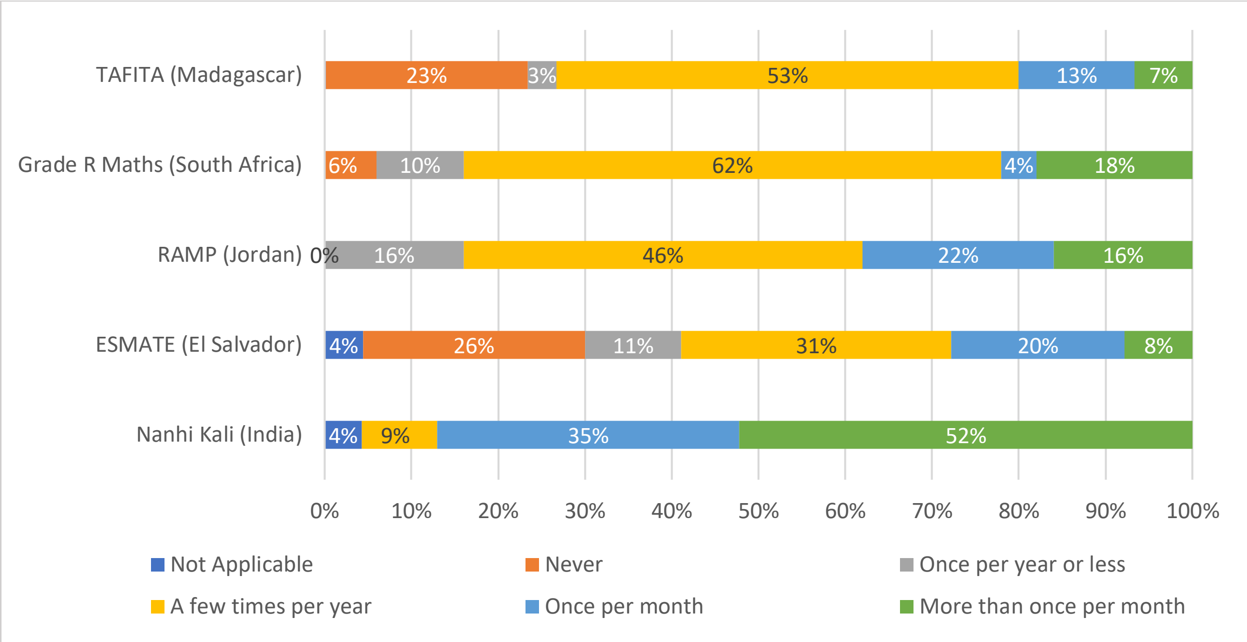
Theme 4: Ongoing support emphasizes feedback, problem solving, and learning new content over inspection and evaluation.

There was some variation in how teachers were supported across programs (including through teacher meetings, coaching, mentoring, and monitoring visits), but what they focused on in terms of the type of support provided to teachers, and what teachers found to be most helpful, was similar: Teachers had opportunities to get feedback, solve problems, and learn new content. Additionally, teachers reported that the individuals who provided professional development were more supportive and friendlier.

Coaching provided through support visits was a component of five of the programs, as more than one-half of the teachers from these programs reported receiving a coaching visit at least once per year or more. Coaching visits occurred most frequently in Nanhi Kali and

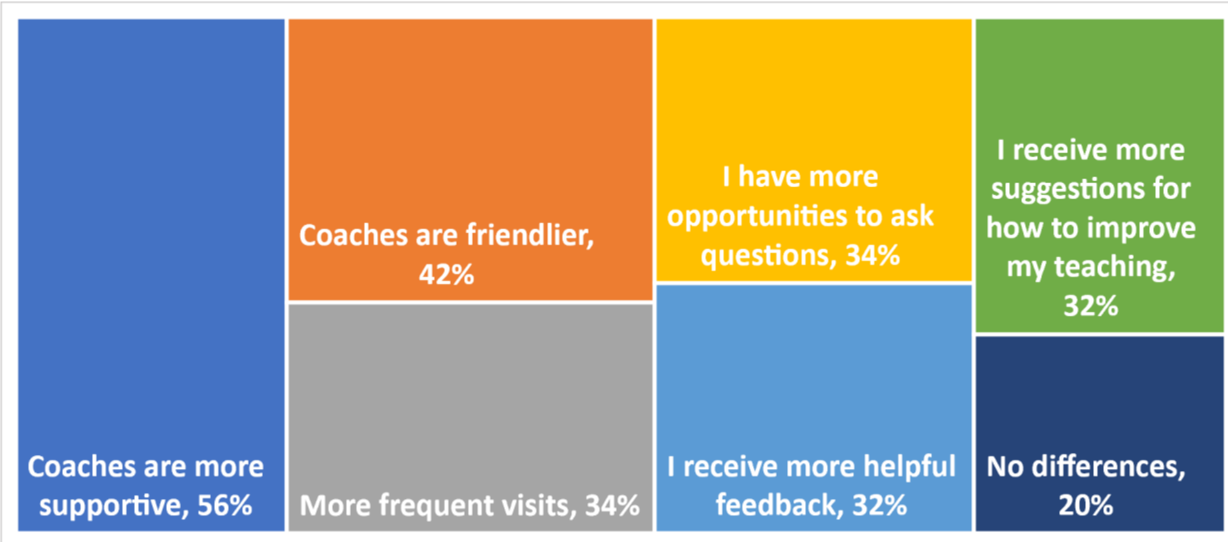
RAMP, with 38% of RAMP teachers and 87% of Nanhi Kali teachers saying that they were coached once per month or more (**Figure 32**).

Figure 32. Frequency of coaching visits, as reported by teachers (percentage), by program



Teachers for RAMP, R-Maths, and TAFITA were asked how the coaching they receive under these programs differed from what they've received in the past. **Figure 33** shows the average responses across programs (individual program data can be found in **Annex A**). Overall, teachers most commonly reported that under these programs, those who provided coaching were more supportive and friendlier. In addition to being visited more frequently, roughly one-third of teachers reported that they have more opportunities to ask questions, get more helpful feedback, and are provided with more suggestions for how they can improve instruction.

Figure 33. Differences between coaching before and during programs, as reported by teachers (average across programs)



Teachers from all five programs with coaching were then asked what they find to be most helpful from coaching visits (**Table 45**). In four of the five programs, teachers said that the most useful forms of support were the observations themselves, the positive feedback given, and the recommendations aimed at improving their instruction.

Table 45. Helpful activities in a typical coaching visit, as reported by teachers, by program

	Nanhi Kali (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Coach observes my instruction	83%	41%	64%	54%	36%
Coach provides positive feedback on my instruction	52%	39%	66%	64%	39%
Coach provides areas of improvement for my instruction	44%	31%	56%	56%	52%
I ask the coach questions about my instruction and how to improve	57%	19%	36%	32%	42%
Coach and I discuss progress from last observation	48%	16%	50%	26%	39%
Coach provides feedback on students' performance	39%	16%	40%	40%	27%

	Nanhi Kali (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Debrief with other teachers or head teacher	44%	14%	24%	28%	42%
Coach and I agree on skills and practices to focus on moving forward	35%	20%	28%	24%	21%
Coach assesses students	48%	11%	42%	8%	15%
Discussion of expectations at start of visit	30%	20%	24%	34%	9%

Gray cells denote the top three responses for each program.

In interviews with coaches for Nanhi Kali, RAMP, R-Maths, and TAFITA, we asked about post-observation conferences. When asked what they discussed with teachers after lesson observations, coaches across programs frequently said that they gave areas for improvement. Over two-thirds of coaches in Nanhi Kali, RAMP, and R-Maths also said that they discussed math content and the instructional approached used. The majority of RAMP and TAFITA coaches also reported addressing teachers' questions and giving at least one positive piece of feedback on the lesson.

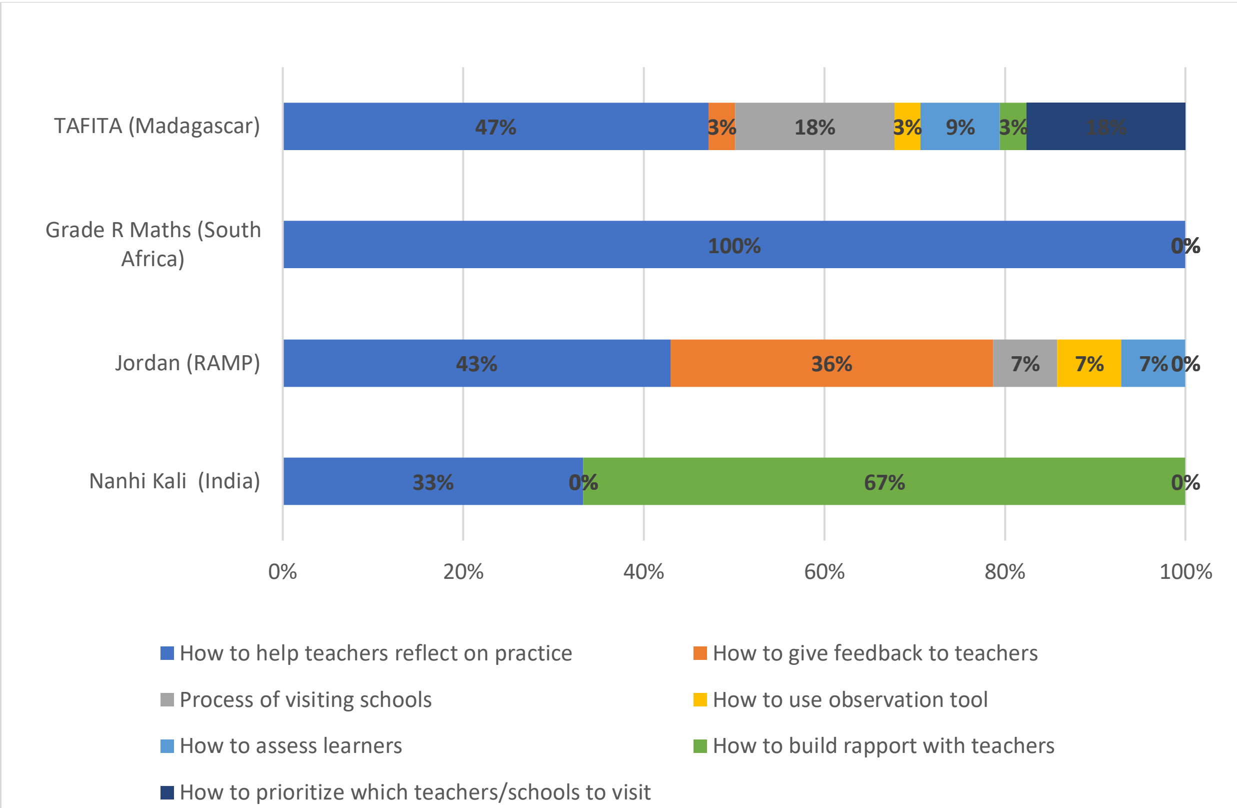
Table 46. Topics covered in post-observation discussions with teachers, according to coaches, by program

	Nanhi Kali (India)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Give areas for improvement	80%	93%	67%	93%
Support teacher to understand math content	60%	64%	100%	37%
Discuss lesson methodology or instructional approach used	60%	64%	67%	22%
Discuss classroom management	40%	64%		51%
Discuss teacher questions or challenges faced	40%	79%	0%	73%
Give at least one positive about the lesson	20%	79%	0%	81%
Ask teacher to reflect on their instruction	40%	71%	0%	59%
Discuss use of materials	20%	64%	0%	59%
Agree on way forward	0%	86%		20%
Discuss administrative matters	40%	36%		10%
Model activities	20%	57%	0%	34%
Discuss students' progress	20%	64%	0%	24%
Other	0%	0%	33%	0%

Gray cells denote the top three responses for each program.

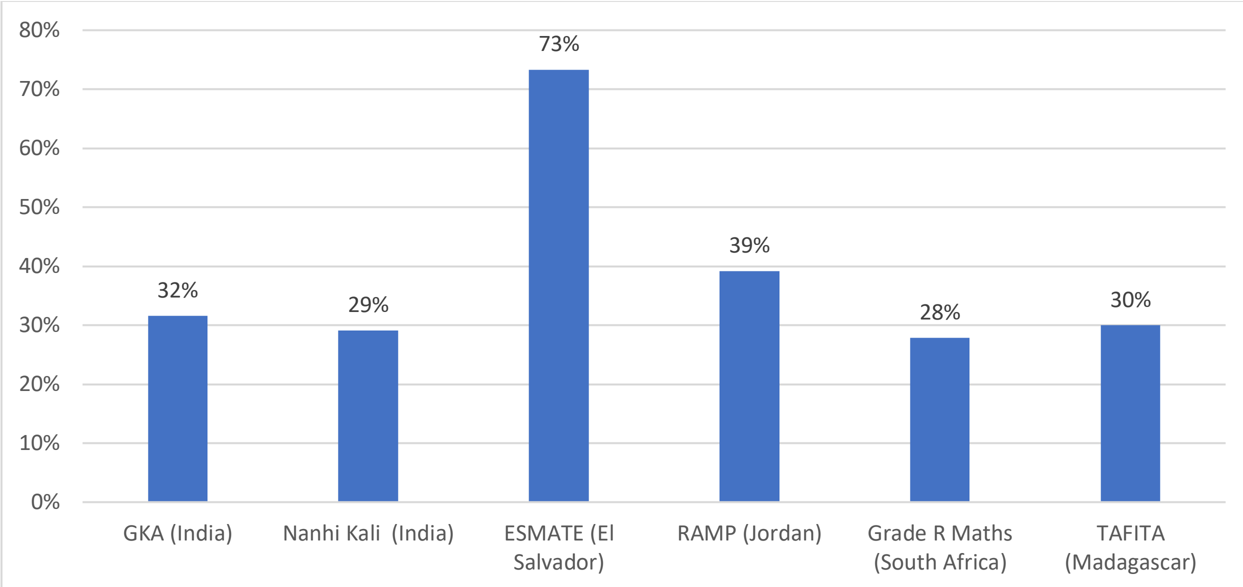
We then asked coaches about the training they receive as a coach (**Figure 34**). Overwhelmingly, coaches said that they found training on helping teachers reflect on their own practice to be the most helpful aspect of such training.

Figure 34**. Most helpful aspects of training on coaching, according to coaches, by program**



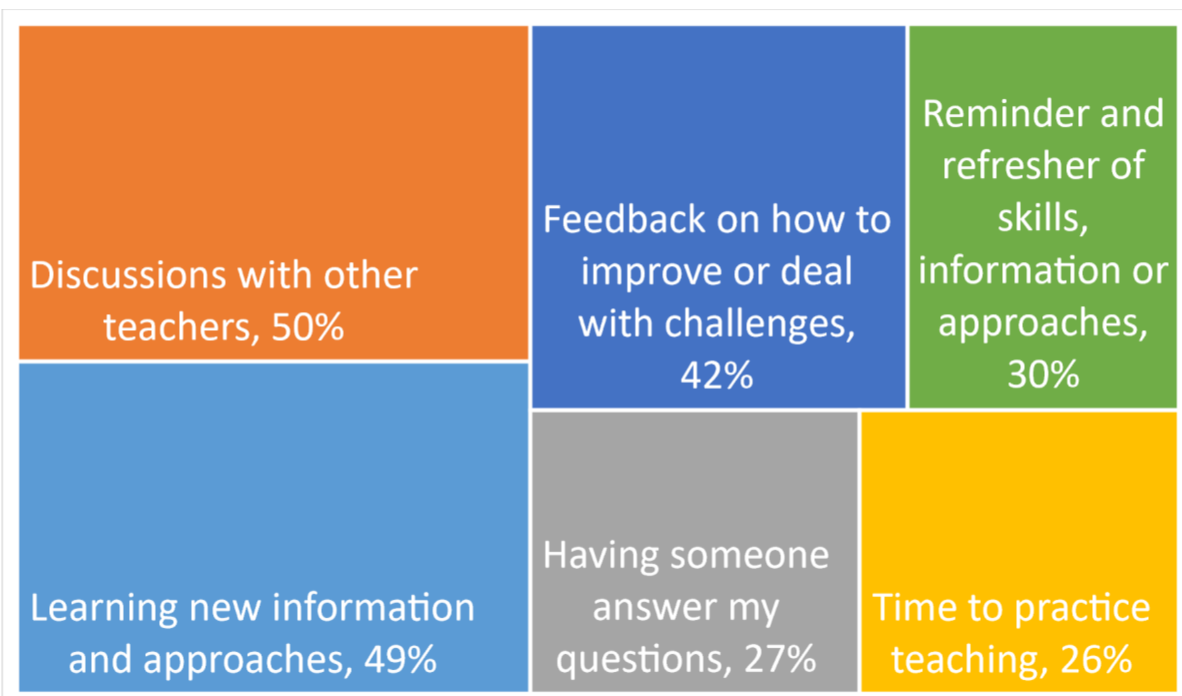
When asked what program supports they received, between 28% (R-Maths) and 73% (ESMATE) of teachers reported receiving teacher meetings. For some programs, teacher meetings postponed during COVID-19-related closures had not yet resumed at the time of data collection. It is important to note that in ESMATE, teachers may have thought of teacher meetings and trainings interchangeably, as both were held at the departmental level and focused on use of the materials and on instructional challenges faced by teachers.

Figure 35. Percent of teachers who report attending teacher meetings, by program



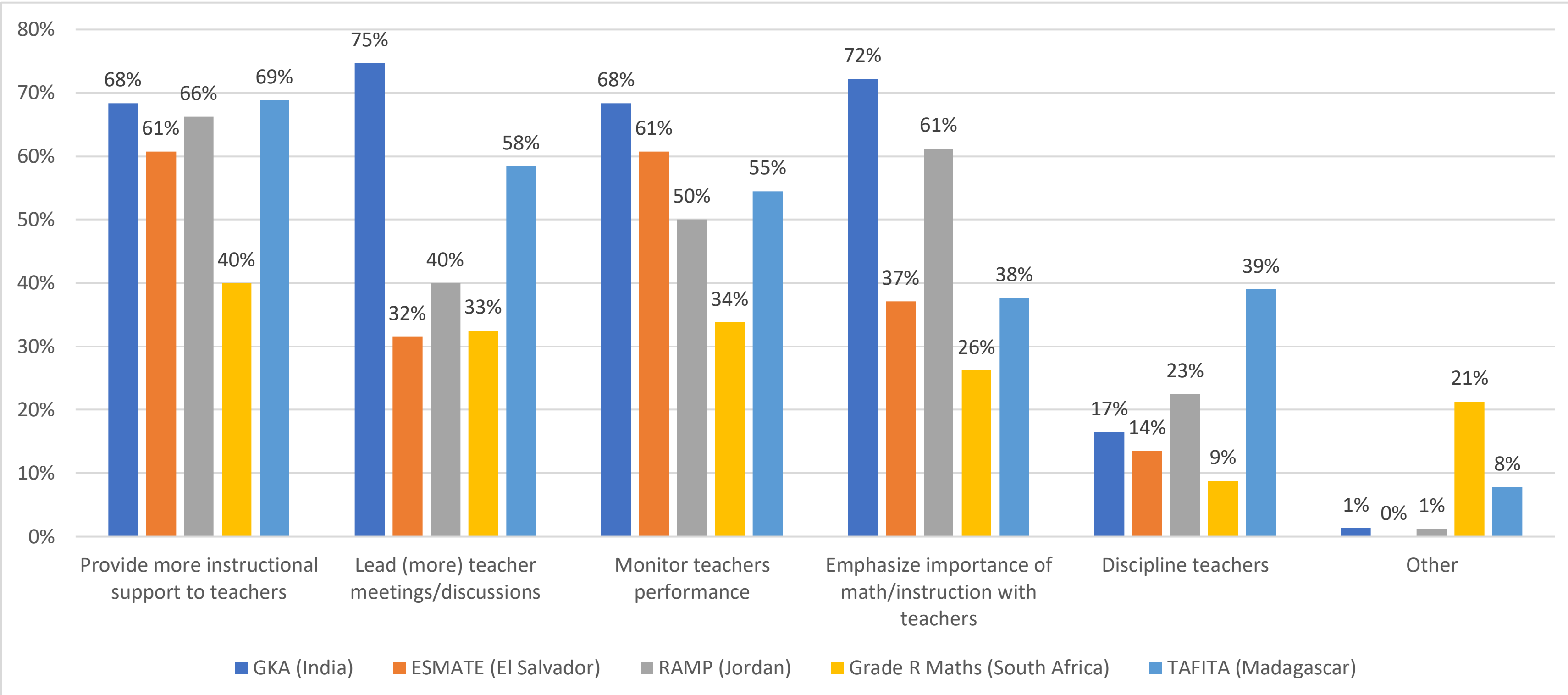
Of those teachers who have attended teacher meetings recently and could recall what went on, we asked what aspects of these meetings they found most useful. As **Figure 36** shows, teachers most commonly responded that discussions with other teachers, as well as learning new information and getting feedback on ways to improve or how to deal with challenges, were most helpful to them.

Figure 36. Most useful aspects of teacher meetings, according to teachers (average across programs)



Head teachers also play a critical role in supporting teachers in adopting new instructional practices. When head teachers were asked to identify the changes that they had made as a result of participating in the programs, they most frequently said that they now provide more instructional support to teachers. As **Figure 37** shows, over half of head teachers from GKA, ESMATE, TAFITA, and RAMP said that they also monitor teachers' performance more. The majority of head teachers from TAFITA and GKA also said that they lead more teacher meetings and discussions, while the majority of teachers from RAMP and GKA said that they now emphasize the importance of math and math instruction. It is important to note that varying responses are expected here, as the strategy for engaging head teachers is different under each program.

Figure 37. Changes that head teachers have made because of participating in their respective program, as reported by the head teachers



5.3.3 Teacher knowledge and attitudes

Table 47. Mathematical knowledge for teaching

	GKA (India)	ESMATE (El Salvador)	RAMP (Jordan)	TAFITA (Madagascar)
Math teacher knowledge number score, out of 23 questions (mean)	10.6	9.4	10.9	10.5
Math teacher knowledge percent score (mean)	45.9	40.7	47.4	45.7
Number sense percent score (mean)	35.7	40.4	45.4	39.2
Operations percent score (mean)	54.3	36.9	46	44.7
Geometry percent score (mean)	53.4	56.5	51.9	57.6
Measurement percent score (mean)	34.2	31.8	50.6	47.7
Developmental progressions percent score (mean)	48.6	47.4	51.8	47.1
Scaffolding percent score (mean)	50.6	35.6	52.6	43.9

Overall, teachers showed moderate knowledge of mathematical content and pedagogy that is important for teaching the early grades, scoring between 41% of items correct in ESMATE to 47% in RAMP, with teachers from other programs falling somewhere in between. Teachers scored particularly well in geometry across all programs. In other domains, the scores varied. The results suggest that while moderate MKT is important for program success, teachers need more support to develop MKT across the various domains.

Table 48. Teacher attitude scales

	High prevalence (76–100%)		Moderate prevalence (51–75%)		Low prevalence (26–50%)		Very low prevalence (0–25%)

Items	GKA (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Students in primary school can apply prior knowledge to solve new problems. *(vs. Students in primary school must always be shown how to solve a problem before they solve it.)*	54.4%	66.7%	84.8%	44.3%	51.2%
The process of solving a problem is the most important part of math class. *(vs. The correct answer is the most important part of math class.)*	84.8%	98%	84.8%	93.7%	86.3%
All children can be good at math if they try hard. *(vs. Some children are just naturally better than other children at math.)*	83.5%	66.7%	45.6%	62%	83.7%

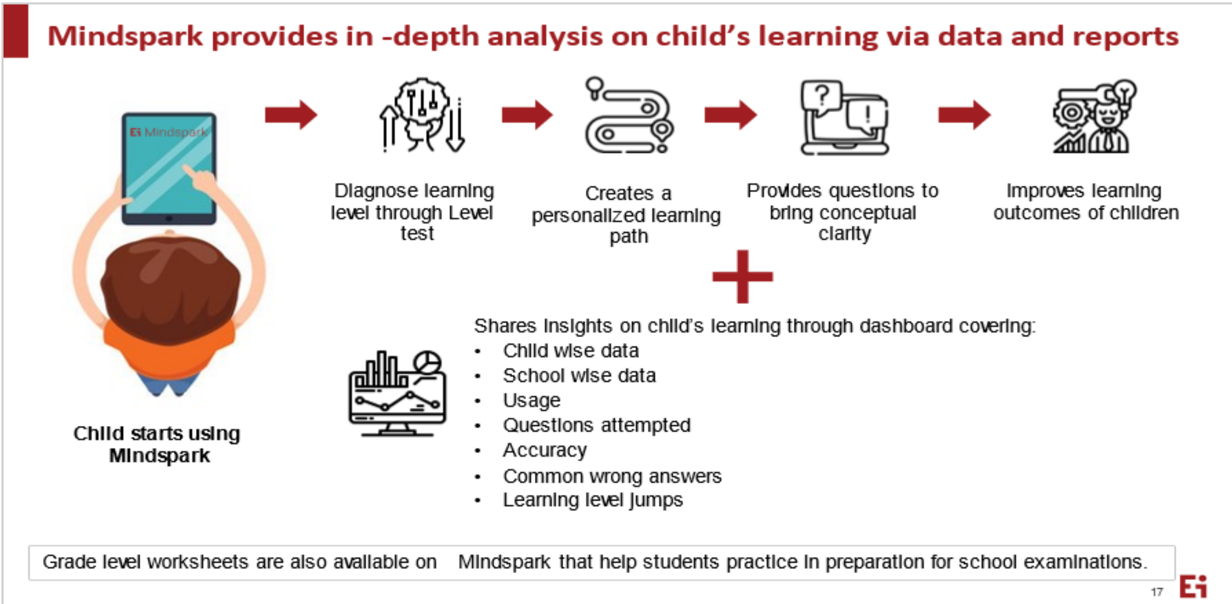
	High prevalence (76–100%)		Moderate prevalence (51–75%)		Low prevalence (26–50%)		Very low prevalence (0–25%)

Items	GKA (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
If a student answers a question incorrectly, the best way to help them is to work through the problem with them. *(vs. If a student answers a question incorrectly, the best way to help them is to ask another student to answer it correctly.)*	74.7%	95.3%	59.5%	91.1%	63.8%
Teaching multiple strategies to solve the same problem can help students. *(vs. Teaching multiple strategies to solve the same problem is confusing for students.)*	97.5%	88%	65.8%	91.1%	75%
Girls and boys both need to learn the same amount of math. *(vs. Girls don't need to learn as much math as boys.)*	100%	100%	83.5%	100%	97.5%

5.3.4 Findings on education technologies to support numeracy: Mindspark

As part of the Nanhi Kali after-school program, math is taught twice a week for approximately two hours each day. Each session is split into dedicated tablet-based instruction (i.e., tab time using Mindspark) and non-tablet-based instruction (i.e., non-tab time). While half the students are working on tablets, the other half are engaged in non-tab time, which doesn't follow a fixed curriculum but includes subject-related activities such as homework, worksheets, and practice problems generally aligned with what children are learning in school.

Figure 38. How Mindspark works



Credit: Education Initiatives

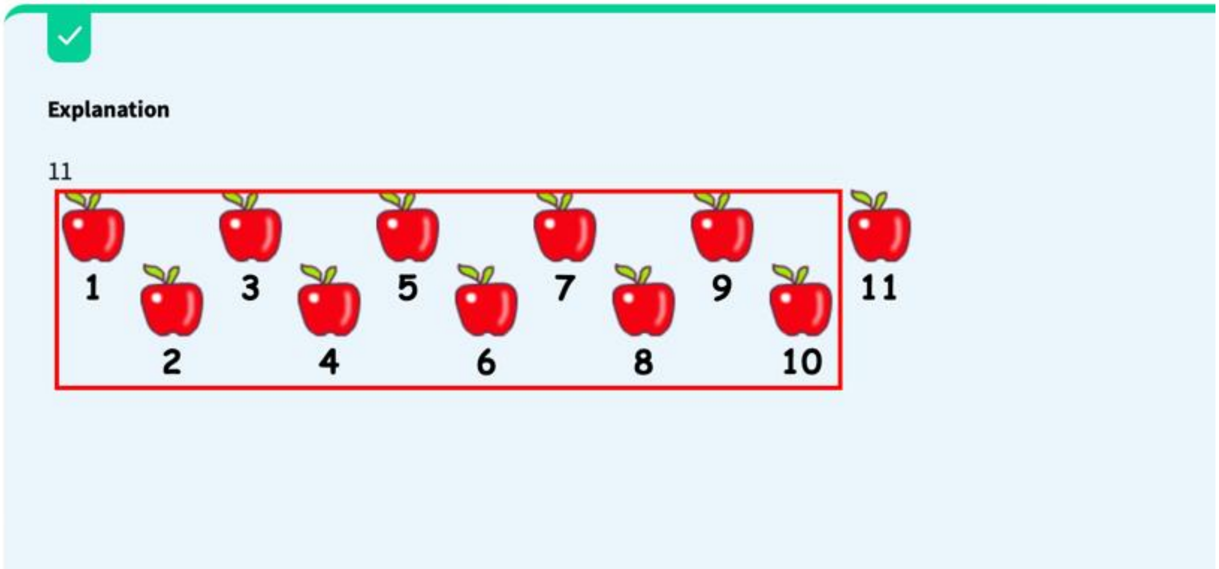
Mindspark's primary method of teaching and learning is through questioning rather than lecturing or modeling. Ultimately, the instructional approach focuses on practice and the application of concepts to help students move toward learning with understanding. The software doesn't teach content according to a child's grade level, so meeting curriculum standards or outcomes is not the driving force behind what children are learning. Rather, children learn at their own level and at their own pace. The software has a built-in adaptive flow that uses a child's response to decide if the child needs additional practice on a given topic or subtopic.

User data from more than 5,600 sessions across grades 1 to 3 in selected centers revealed that students averaged just under 30 minutes per tablet-based lesson. During that time, grade 1 students attempted an average of 41 problems, while grade 2 and 3 students attempted 33 and 31 problems, respectively. This means that students were averaging more than one problem per minute on the tablets. Perhaps even more important is the way that the Mindspark app handles correct and incorrect solutions to problems.

Central to the success of this app is its targeted and differentiated instructional model. Mindspark allows students to master key skills before they move on to more difficult ones. For example, in a unit on counting and numbers up to 20, the app starts with smaller numbers and then moves to larger numbers. When a student provides incorrect responses to larger numbers, the app responds by giving problems with lower numbers again and slowly works back up to larger numbers.

For some topics, the software also provides clear representations of the mathematical content. For example, in the same unit noted above, the app presents pictures of objects and asks students to count them. After the child answers, the app provides the correct answer, as seen in **Figure 39**. In this instance, the app shows the student how to count 11 apples, while also highlighting a strategy that students can use to put the objects in groups of 10.

Figure 39. Example of Mindspark teaching strategy

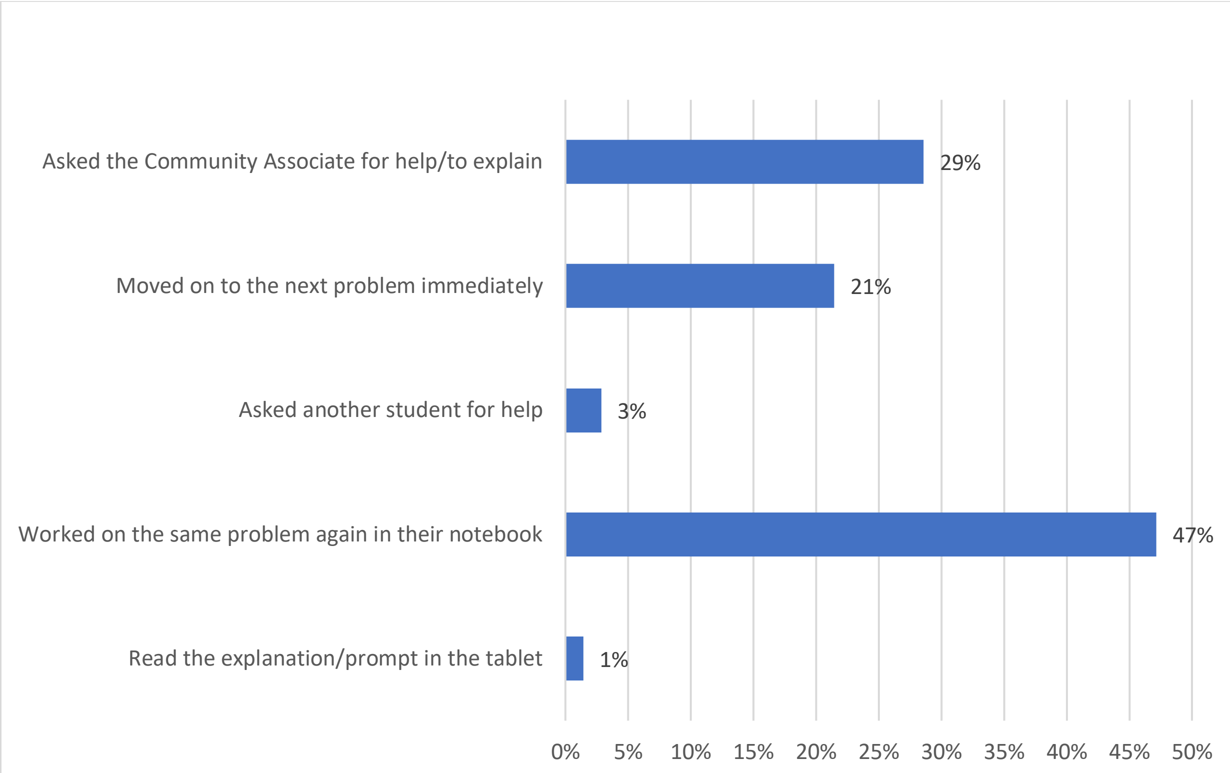


Revealingly, 81% of interviewed students reported that their math skills have improved as a result of the program, and 84% of students claimed that they like math more than before they started the program. Moreover, the majority of students (79%) reported that their improved skills and enjoyment of math are the result of their ability to practice problems using Mindspark.

By design, community associates are meant to support students with their work in ASCs. Despite not being trained teachers, community associates are relied on for instructional support in mathematics, as well as for classroom management and technological support (which was found to be necessary in nearly all observed classes).

During targeted observations of individual students, observers noted that 47% of students incorrectly answered at least one question on the tablet. This is similar to the average rate of incorrect responses from Mindspark user data (43%). When students were faced with an incorrect answer, their most common response was to work on the same problem again in their notebook (47%) before moving on to the next problem on the tablet (**Figure 40**). However, nearly one-third of the time (29%), students asked a community associate for assistance in solving the problem. Interestingly, students did not tend to read the explanation or prompt provided by the tablet (only 1% did so).

Figure 40. Student action after incorrectly answering a question



In 49% of the ASCs, observers indicated that students were paired or arranged in small groups according to ability. In multigrade classrooms, purposeful seating arrangements were helpful to students since they often relied on their peers for support during tab time. More specifically, 59% of students talked to another student while working on the tablet. These conversations almost universally revolved around the tablet or math content, thus providing an additional layer of support to students without negatively impacting their time on task.

5.4 Systems-Level Analysis

The following systems-level analysis seeks to understand the support from the education system that is required for effective teacher training and successful classroom practices. The analysis was informed by structured qualitative interviews with various stakeholders, including government officials, donor representatives, and program staff from five programs: ESMATE (El Salvador), R-Maths (South Africa), GKA (India), RAMP (Jordan), and TAFITA (Madagascar). Nanhi Kali (India), an after-school program, was excluded due to its limited engagement with the formal education system. The primary goal of the interviews was to understand program implementation, its intersection with the broader education system, and factors that contributed to success.

5.4.1 Coding and analysis process

RTI staff and consultants led the interviews and took written notes. The research team then transcribed the interviews and imported them into Dedoose, a qualitative data analysis software that allows several researchers to work collaboratively. Coding employed a mix of deductive and inductive methods, starting with codes drawn from predefined hypotheses (see **Section 3.1.1**), and then codes based on emerging themes. For consistency, team members reviewed code criteria and definitions, and the team member responsible for each of the programs coded a first set of interviews. Once the codebook was completed, the remaining transcripts were coded by one research consultant. After the coding was complete, the lead researcher reviewed the codes and excerpts. The analyses included reviewing the data for themes that were common across the different programs using analytic tools that come with Dedoose. Interpreting the codes, inter-code relationships, and frequency of occurrence helped the researchers find common themes.

5.4.2 Common characteristics across education systems

The main roles of the various government stakeholders are described in **Table 49**. Some programs are fully implemented by the government, with technical support from external partners, while others are led by external partners, with support and engagement by government. Yet certain characteristics were found to be consistent across all programs: (1) strong stakeholder collaboration; (2) an emphasis on monitoring and using data collected; (3) investment in the resources needed to bolster quality instruction; and (4) the institutionalization of practices and processes.

Table 49. Role of ministry and other stakeholders, by program

GKA	Government funds math kits and training of teachers; Akshara Foundation proves technical support and monitors the program
ESMATE	Implemented and funded by Ministry of Education, with technical support from JICA
RAMP	Implemented by RTI and funded by USAID, works though the government system
R-Maths	Implemented by WCED, with funding from the Zenex Foundation and the Maitri Trust and technical support from the University of Cape Town
TAFITA	Funded by JICA and implemented by the Malagasy Ministry of National Education. SOFIASIVE provides technical and administrative support, through a contract with Asuka, a Japanese consulting firm

Theme 1: Programs actively collaborate with key stakeholders.

Collaboration and coordination among key stakeholders emerged as a salient theme, which supports several of our initial hypotheses: key system actors are informed about the program; system communicates expectations for districts, schools, teachers, and students; and system actors play substantive roles in implementation. The collaboration started at the planning and design stages and continues today through the implementation. Even externally funded programs needed full collaboration from government officials, school staff, and, in some cases, parents and communities for effective large-scale implementation. For instance, a top Ministry of Education official from Jordan stated:

> *"We believe that RAMP is one of the best projects that's been conducted in the Ministry of Education. We believe that RAMP is part of the ministry; we cannot say that we are two different parties. We are part of planning, implementation, and sustaining of RAMP activities. We meet regularly and we*

> *review and discuss all plans together. We own RAMP, so there are no expectations about how to work 'with' it because it's just part of our work."*

The collaboration happened at all levels, both formally through steering committees and informally. In Madagascar, a high-level government official highlighted the importance of a steering committee in enhancing collaboration:

> *"Working through a steering committee led by the ministry enables the project and the ministry to plan together, review implementation progress, and adjust and manage implementation. The work of the project is tied to specific offices within the ministry: Basic Education and Early Childhood Directorate… Directorate of Pedagogy … Technical staff from these offices are the ones who report on project progress to the steering committee."*

There were regular meetings between program staff and government officials aimed at coordinating program activities and ensuring alignment with government policies and practices. As a ministry official from Jordan explained:

> *"[There were] regular meetings with RAMP from the start. NCCD sent all drafts to the RAMP team for feedback before they were printed. The goal of that was to ensure that the textbooks were aligned with the RAMP materials. We don't want the teachers to get confused by receiving conflicting information. We want teachers to have streamlined materials. RAMP textbooks [are]aligned with NCCD."*

Similarly, GKA in India exemplified collaboration between the Akshara Foundation and the government of Karnataka in implementing the program at scale. Akshara funded the pilot and developed technical specifications for the math kits, while the state government procured and distributed the kits statewide and trained teachers. A district official in Bangalore Rural described this collaboration:

> *"We give the training. Our main responsibility is to train the teachers. DIET [District Institute for Education and Training] coordinates the training. The child needs to have concept clarity in math. All concept clarity will come during the teacher training program. Training happens from block resource persons to cluster resource persons to teachers … DIET is responsible for getting kits to schools through the block resource center and cluster resource center. These centers are visiting schools and talking to teachers. They have a team that goes to the school."*

WCED in South Africa used existing policy statements to coordinate input from stakeholders. According to a high-level WCED official:

> *"When it came to R-Maths, we knew what we wanted—the concept guide and certain topics, and to retain the grade R system. But the funders and [Cape Town University] had their own ideas. We said, 'You can keep your ideas, but it must be consistent with the CAPS'—a 256-page document."*

In summary, stakeholder collaboration was identified as paramount in all programs, beginning from planning to implementation. Collaboration was both formal (e.g., through steering committees) and informal. Regular coordination meetings helped ensure that programs were consistent with existing government policies. In order to reach schools at scale, most programs worked through government officials and systems: the government provided the people, while external collaborators provided technical assistance and funding support.

Theme 2: Investments are made in resources to improve quality classroom instruction.

Regardless of the funding source—external donors or the government—every program invested in additional resources to support quality instruction. This theme aligns with the system hypothesis that necessary inputs and resources are reliably made available. The principal areas of investment were in the professional development of system actors (e.g., teachers and instructional coaches) and the provision of teaching and learning materials.

Professional development of system actors. All programs invested heavily in the professional development of teachers and other education professionals. The professional development process took several forms: in-person training workshops, online training, professional learning communities, and coaching support. The training topics aligned with the intended roles in supporting learning: instruction, monitoring, coaching, and the use of teaching and learning materials. A provincial education officer in South Africa described the training of teachers as follows:

> *"For new grade R teachers coming into the system, we run novice teacher orientation sessions. We also have a CAPS training that includes R-Maths. In fact, the province will be training 120 teachers during the upcoming June holidays … Every year, we run induction and refresher courses with grade R teachers either for R-Maths or e-Lit training."*

In addition to teachers, district-level personnel—"middle managers" who coach and supervise teachers—received significant amounts of training. For example, RAMP trained coaches in assessment and supervision tools such as eCoaching and the EGMA and EGRA. It also trained principals and supervisors to be instructional coaches for teachers, thereby improving relationships between teachers and their supervisors. High-level ministry officials also had the opportunity to receive support in data analysis, report writing, and dissemination of the findings. A district official in Jordan stated:

> *"Training and capacity building efforts have been huge from the beginning—how to train principals and teachers; how to conduct coaching; how to conduct LQAS [lot quality assurance sampling]."*

In Madagascar, school management committees (known as FEFFIs)—a vital actor in the program—also received capacity-building training on the establishment of FEFFIs, school action planning and community auditing. These contributed the functioning of FEFFIs and the implementation of remedial activities outside of school hours.

. Meanwhile, in India, GKA used a cascade training method to train teachers on how to use the provided math kits to teach mathematical concepts; the Akshara Foundation trained state education officials as "master trainers," who then cascaded the training down to district and schools. In addition, the foundation recorded all of the training content and made it available on YouTube so that teachers could access it on their own time.

Teaching and learning materials. Besides strengthening the capacity of system actors, all programs provided a variety of teaching and learning materials to support instruction. These included math kits (with a variety of manipulatives), student books, teacher's guides, tablets, and more. A subject advisor from South Africa described the support provided by R-Maths:

> *"R-Maths brings other resources—the teachers get books, which are part of a kit, including one activity guide per term, a poster book, and a concept guide, together with manipulatives that children can use, such as structure beads. These are a specific set of resources provided through the R-Maths program. Grade R teachers are also encouraged to use recyclable materials and dough. In addition to this, we provide teachers with Legos—this really supports the learning process. New teachers are provided with kits. We gave every school the kits for their grade R classes—principals received a circular to put the kits on their inventory lists. The kit must be kept in the classroom."*

GKA also provided math kits for teachers, which the Karnataka state government funded:

> *"[The state government] provides the money for procuring TLMs [teaching and learning materials] and training. The first chunk was from a dedicated fund for the [district most in need]. Funding went through the project approval board and came partly from the state and central government."*

In Madagascar, there were no manipulatives provided, but all teachers received a teacher's guide with tips on how to use low-cost and locally available materials to make learning aids. ESMATE distributed materials even though its budget appeared quite constrained. A government official explained how they had to make choices because of costs:

> *"The way to ensure financing for the printing of books has been to gradually reduce the number of notebooks that the school had. So, the child has seen fewer notebooks, but instead of a notebook, a textbook arrives, which more or less balances the costs."*

In short, we found that programs prioritized investments in the professional development of staff and in the provision of quality teaching and learning materials. Overall, these investments were celebrated as key contributors to program success. While professional development activities are common in education budgets, in these cases the professional development was closely aligned with the goals of improved instruction and learning outcomes, and the teaching and learning materials were evidence based, of high quality, and given to students and teachers in adequate quantities.

Theme 3: Programs emphasize continuous monitoring and use of data for system improvement.

Monitoring and the use of data to inform practice emerged as a common and important practice across all programs. The two most common focus areas of monitoring were students' learning outcomes and teachers' pedagogical practices. This theme aligns most closely with the research hypotheses that *expectations for system actors are specified* and that the *system monitors performance relative to stated expectations.*

Student learning outcomes. RAMP provides a good example of a program that systematically collected student assessment data using technology tools and then used data to make decisions about what support to provide to schools. As one high-level ministry official reported:

> *"RAMP is really successful because it is based on data that come from real, authentic assessment. Once the data are received, interventions are developed based on evidence. Results from the assessment are then shared throughout the system so other stakeholders can act on them."*

For example, the assessment data helped field directorate officials plan for remediation:

> *"We get a report released annually from the Ministry of Education. After that, as head of the field directorate, I will work with the head of supervision to work with schools to enact changes. Due to the eCoaching system, we know the results of all schools [in our field directorate and others]. This gives us evidence on the "effective instruction" indicator, based on which we can implement remediation approach."*

GKA uses a unique approach involving tens of thousands of community volunteers who hold math contests among a cluster of schools at the village level to strengthen the accountability relationships between the parents and the school. The data from the assessment are available to parents and teachers in real time, and they can have informed conversations about the results. The Akshara Foundation chairperson explained the process as follows:

> *"All the children come into one school. Questions are prepared by the Akshara Foundation … using almost 65,000 volunteers. Teachers are not allowed in the arena. In the first hour, there's a paper-and-pencil test. In hour two, papers are graded. In the third hour, prizes are given. A general announcement is made about the results. In the fourth hour, there are fights. Parents are saying I've been sending my child to school, and the test is telling us that they can't do anything. Teachers are saying we give them homework, and you are not giving them time to do it."*

We found that programs assessed students frequently and made use of this data to improve instruction and provide targeted support (e.g., remediation) for students.

Teacher practice. Beyond assessing student learning outcomes, decentralized government officials—such as subject advisors in South Africa, District Institute for Education and Training officials in India, and senior teachers in Jordan—also monitored teachers' practices.

Programs such as GKA and RAMP use an electronic system for classroom observation, with the data readily available via an online dashboard.

R-Maths is a good example of a program with a robust monitoring system focused on pedagogical practices. Subject advisors observe pedagogical practices to determine if teachers require any support or resources or to address any challenges related to the implementation of the program. A provincial education official conveyed this message:

> *"Subject advisors give a report after each visit. We expect them to submit every term. They do dedicated monitoring with a certain percentage of schools, and they look at everything in the classroom. Is it the way it should be? They complete a monitoring tool. Everything that is in the classroom, print-rich classes, it also includes the teacher practices."*

RAMP uses a combination of eCoaching, exams for teachers' promotion, and school visits to monitor teachers. Like R-Maths, RAMP also starts the monitoring process during the initial training program for teachers. As described by a high-level ministry official:

> *"Monitoring starts from being there in the trainings—attending trainings that are provided by RAMP to the core team. Then we monitor core team trainings to teachers. We monitor through coaching data and dashboards."*

In summary, all these programs were found to prioritize data collection and use. Whether the data are about student performance, instruction methods, or the availability and use of teaching and learning materials, these data play a key role in shaping and improving the program. For example, a high-level ministry official in Jordan explained that the program shifted its focus to basic math skills, including developing new materials, because of the data gleaned from student assessments:

> *"In order to ensure that we are on the right track, we kept applying formative and summative assessments. In 2017–18, the project developed a greater focus on basic skills, including new materials, based on the needs of students that were discovered from the assessments."*

Theme 4: Programs focus on systemically embedding and institutionalizing best practices, with an eye toward sustainability.

For donor-funded programs, institutionalization and sustainability are critical issues, as one of the measures of a program's success is its continuity after donor exit. Two of the programs, ESMATE and R-Maths, are examples of programs that transitioned to full-fledged government programs from smaller pilots. To improve their chances of sustainability, all of the programs studied adopted a range of strategies, including designing with institutionalization in mind, implementing through government systems, and conducting policy advocacy.

<u>Designing for sustainability and scale.</u> Most programs built the foundation of institutionalization by involving government officials from the outset. They actively involved government officials during the planning phase and ensured alignment with the country's educational goals and priorities. As explained by a provincial education officer from South Africa's R-Maths program:

"The R-Maths program fully aligned with the existing numeracy curriculum. Initially, it caused some confusion among teachers, but we had trained them to understand it as a complement to CAPS, which would enrich the curriculum."

Similarly, in the design of GKA, the Akshara Foundation made choices that would enable the government to be able to take the program to scale. The foundation's aim was always to develop an initiative that could be delivered at scale using government resources. As an Akshara Foundation official noted:

"The most popular word is sustainability, replicability. Whatever we do, we design for scale from the beginning. While we will do a pilot, we still do exactly what we will do if we have 50,000 schools. The pilot and ultimate thing we scale might have some differences, but the design philosophy is 'design for scale.' That is, you have to learn how to work with government, and that is not the easiest thing in the world ... Working with government has its own challenges. But the beauty is that if you crack that problem, it's very hard to get dislodged from the system."

Additionally, many of these programs were proactive in influencing government policies and practices. This is evident both in government-led programs and those implemented by external organizations, where program officials spent time advocating for some of the program practices to become general policies or practices. An El Salvadorian official shared their experience in ESMATE advocating for broad curriculum change:

"I think the impact it has right now, and I say this as curriculum director, is that from that experience, we have systematized the way to make curricular changes in the country. So, we proposed with the president that the educational system had to be reformed, and we have proposed an educational reform that is based on this curricular change, based on the experience we had with ESMATE."

Policy changes were not exclusive to ESMATE. TAFITA and RAMP, which are donor-funded programs, also led to policy changes. At TAFITA's urging, the government of Madagascar merged school accounts for streamlined financial management, and the government of Jordan integrated RAMP's best practices into continuous professional development policies. As noted by two high-level officials in Madagascar and Jordan, respectively:

"The policy concerning school management committees changed—creating a single school account instead of separate school accounts for capitation grants and for money raised locally; creation of the school management committee and establishment of its operating norms; official recognition of district and regional federations of school management committees."

"Yes, CPD [continuous professional development] policies have changed in recent years, and all of RAMP's best practices are included in these policies. All of RAMP's procedures are part of the official CPD programs now—tied to

teacher promotion. All of these programs are based on teaching standards developed by RAMP."

In India, GKA convinced the Karnataka state government to fund math kits for all public schools in the state. The Akshara Foundation funded the pilot and a third-party evaluation with its own money and then used data from the evaluation, testimonials from teachers, and a demonstration of the math kits to persuade a high-level state official of the merits of the initiative. Encouraged by the success in Karnataka, other states subsequently showed interest in the program.

Sustainability was easier to achieve with strong government leadership. The provincial government in Western Cape demonstrated ownership and commitment to the R-Maths program from the start. It was deeply engaged in the technical details and had a budgetary commitment to continue the program. All of these factors contributed to the sustainability of the program after external funding and technical support ended.

5.4.3 Challenges to scaling and sustaining interventions

While many programs expanded their reach, the scaling process varied and faced several obstacles. Typically, programs started small, demonstrated results, and then scaled up. The added resources for scaling up mostly came from the government, and, based on funding availability, decisions had to be made about what aspects of the program to scale up.

For example, a pedagogical coach from El Salvador said:

"[ESMATE] gradually expanded by levels and to the entire country. First it was a sample, a pilot test, and then it was extended to the entire country, increased by the new [financial] resources."

In India, the GKA program, which was already operating in all public schools in the state of Karnataka, expanded its reach by including more grade levels. In addition, the program has scaled to a few other states that expressed interest based on the successes in Karnataka. The program has most of its materials freely available online for other NGOs and governments to use. As one GKA program official explained:

"State officials say, 'Can you repeat [this program] for grades 6 through 8?' So, we are now creating a new kit for grades 6 through 8. The state gave 30 teachers to work with us. That kit is ready, and they are now buying it for 10,000 schools, so now the state owns all of it. We are just a huge catalyst. We've done it and put it up in other places, and other organizations can take it and run with it."

Like India, Jordan is also scaling up some of the activities and practices of RAMP into other grade levels. A district official from Jordan said:

"RAMP was a pioneer in the eCoaching system. The Ministry of Education is looking to expand it to the upper grades."

Another high-level ministry official explained the scale-up at different levels:

"We are trying to benefit from the RAMP methodology, worksheets, and materials to develop the support for teachers in grades 4 through 6. We are

working on a national remedial plan for education with the Ministry of Education—grades 4 through 11—and we recommended that the RAMP approach should be applied to grades 4 to 6, including a focus on foundational skills in training for teachers. I hear a lot from the teachers and principals that they love RAMP and that they want it to continue."

Scaling, however, presents challenges, including sourcing additional resources, responding to the diverse needs of various districts and schools, and establishing the operational infrastructure to support scale-up.

A representative from TAFITA illustrated some of the adjustments made for expansion:

"During the expansion from pilot to phase 1 districts, the program took into account how to reduce costs—both to stretch the available JICA funding and also to make it more likely to be sustained by the ministry. Adjustments included a reduction in the number of books used, from nine to three, and a reduction in the number of math posters provided to schools. Also, the organization of the teacher training sought to take advantage of existing ministry structures—teacher pedagogical days already programmed into the school calendar, so JICA and the ministry agreed to use one session [three days] for the TAFITA training."

Similarly, a RAMP representative explained how the program had to adjust its coaching model to make it more likely to be sustained by the ministry:

"In the initial years, every teacher had six visits per semester. But how could it be sustained? The first shift was to ensure that supervisors at least visit all teachers to understand their needs. High-performing teachers didn't need significant additional support; but [the idea was] to identify those teachers in need of additional support. The second shift was to establish coaching within schools [i.e., for senior teachers], which began in two districts—and the ministry is now trying to sustain this at the national level. The third shift was to use principals to build their capacity to be instructional leaders in schools. This also made principals accountable for the performance of their students. The first two shifts are institutionalized, while the latter two still need some work."

Scale-up also meant addressing the fact that districts and schools varied in terms of their capacity and needs, which can impact how well a program might work in another location. One donor official from the TAFITA program expressed concerns over the uneven capacities across schools:

"[We have] concerns about variations in capacity at the district and school levels—some schools and some zones and districts have less capacity to organize and support the remedial classes and the necessary follow-up support and data gathering."

To summarize, scaling was not an afterthought, and many of these programs were already thinking about and planning for scale during their design phase. Programs often began on a small scale, proved their effectiveness, and then sought to expand, financed either by the government or donors. However, scaling up is not without its challenges, as it requires additional resources, an ability to cater to the unique needs of different schools and students, and a supportive operational infrastructure. As these programs expanded, there were efforts to optimize costs and resources, which usually meant changes to the initial design, such as RAMP's shift in its coaching model.

5.4.4 Remediation and recovery efforts: COVID-19

The COVID-19 pandemic disrupted traditional instruction. While all programs mentioned facing pandemic-related challenges, RAMP, GKA, and ESMATE extensively discussed their responses to these challenges.

To counteract physical school closures, school systems pivoted to remote learning, and many programs also shifted to providing materials and tools to support students and teachers in this unfamiliar environment. RAMP, ESMATE, and GKA provided online videos, apps, and worksheets to help students learn remotely. A government official from El Salvador described this approach:

> *"The team was designing a lot of virtual material to work online, supporting applications so that the child had interactive resources for the different content from first grade to high school. A large number of videos were created, an incredible number of videos and animations where specific content was explained, from addition, subtraction, number formation, multiplication, division, fractions, decimals, geometry, statistics … Classes were also given on television and radio."*

In India, the GKA team converted all of the program's trainings into video format and posted them online for teachers to access. Some teachers went out to teach students in their villages and took the kits with them. Meanwhile, ESMATE set up a call center to provide support to students, teachers, and parents. As noted by an ESMATE representative:

> *"In pandemic times, we had a call center. At 11 at night, they wrote to us on Facebook and on other sites. So, we looked for mechanisms through which teachers, students, and even parents could request things, and we sent them to them. We sought to connect directly with them."*

Post-pandemic, programs resumed normal activities, but initiatives that had been started during the pandemic provided new opportunities to do things differently. For example, in Karnataka, India, the government rolled out a one-year "learning recovery program" in 2022 to address the learning loss during the pandemic. With this initiative's emphasis on functional literacy and numeracy, the demand for the math kits became stronger, and the Karnataka government began working with the Akshara Foundation to develop similar kits for grades 6–8. As a district official explained:

> *"For the learning recovery program, the kit is more effective for teaching. The teacher is using reading cards and the math kit. The focus is on competencies, not just the subject."*

A state-level officer in Karnataka reinforced this idea of the importance of the GKA math kits for the learning recovery program:

> *"During the pandemic, we had online support to the teacher. Now, we are implementing the learning recovery program for an entire year … different from other states, which are shorter. We are still using the kits to learn math, and it has helped very much with that. The materials have helped very much to address the gap. The learning recovery program selected a few important learning outcomes, and these are covered [by the kit]."*

In ESMATE as well, the ministry procured tablets to help with remote learning, which schools continue to use in the post-pandemic classrooms.

In summary, the COVID-19 pandemic disrupted traditional instruction and program activities due to school closures. In response, education systems transitioned to remote education. The programs adapted by providing a range of resources, such as online videos, apps, worksheets, and call centers. Nonetheless, there was significant learning loss during this time. Thus, as schools transitioned back to in-person instruction after the pandemic, some of the initiatives that were started during the crisis continued.

5.4.5 Other keys to program success

In addition to the common characteristics described above, respondents weighed in on the specific aspects of their programs that contributed to improved math outcomes at scale, where other similar programs did not have the same result. The responses were many and varied. For example, GKA respondents spoke about the high level of buy-in among teachers. As one high-level state official said:

> *"I have always believed we should not trust any program unless the teachers own it. We should say the fact that the teachers are involved is the reason GKA has been accepted."*

In Jordan, training teachers specifically on job-related competencies and specialized content was described as unique to RAMP, according to a high-level government official:

> *"I think that one of the things that we hear about RAMP a lot is that teachers were trained on specialized content knowledge. This is missed in the other grades. On mathematics, conceptual understanding, misconceptions, etc. This is key."*

Similarly, a high-level ministry official from South Africa described the R-Maths training as being better than that of other programs:

> *"The main thing was the module on training the trainer. With e-Lit, we had a lead teacher training. We had a session first with lead teachers. But this was a bit strenuous, and lead teachers wanted more money, which wasn't there. With R-Maths, we decided to train the subject advisors. I think that is why it is a stronger model. Their core job is to support and guide teachers."*

In Karnataka, India, GKA respondents focused on the math kits that allowed for play-based activities that helped translate abstract math content into tangible learning. These kits allowed students to "learn by doing," which many were convinced was a better way of learning. A district education manager explained:

> *"GKA is one of the best methods—concrete to abstract, known to unknown … similar to constructivism. Pupils have concrete experience in the classroom with the kit and teaching materials … The teacher is using [the kit] as a teaching aid. The teacher shows the students; students work in groups of five or six members; and every child has experience in handling the kits."*

Regular monitoring and evaluation activities were seen as a strength of programs in Jordan, South Africa, and Madagascar. Respondents in Jordan and South Africa mentioned using systemic assessments and tests in more consistent ways than other programs. For example, data from assessments and tests were used to decide which students needed remedial support, the content of textbooks and curricula, and the content of teacher trainings and support. A high-level ministry official from Jordan described this relationship of monitoring and evidence-driven intervention:

> *"RAMP is really successful because it is based on data that come from real, authentic assessment. Once the data are received, interventions are developed based on evidence. As they administer their intervention, they develop new tools and continually measure the impact and change. It's connected directly to the needs of teachers and students."*

Finally, TAFITA found the engagement of communities in improving learning to be a distinct factor for its success. Parents and communities were actively involved in school governance and management, with a focus on improving learning outcomes.

6 DISCUSSION AND RECOMMENDATIONS

The availability of program design data allowed us to understand how the Numeracy at Scale interventions worked across the five domains of program characteristics that we identified: *Materials*, *Pedagogy*, *Training*, *Teacher Support*, and *Systems*. We found that the program design structures shared some elements and that all included aspects of these five areas. While each program was unique in what it saw as critical to its success, we found 14 elements that were determined to be key for three or more programs and suggest that future interventions consider these elements as essential for program impact. These "essential elements" emphasize meeting students where they are and using different models and representations—including while students practice with one another—to support conceptual understanding in math. These elements also emphasize close alignment with the government, especially around curriculum, monitoring, and coaching. As part of a minimum package, these elements also point to the value of in-person training over multiple points in time. The 14 essential program elements are as follows:

- Learning aids for students (e.g., counters, number cards, place value materials, etc.)
- Program materials aligned to government curriculum
- Structured teacher's guides (scripted lessons)
- Continuous and formative assessment
- Instruction targeted to student level (differentiated instruction)
- Focus on developing conceptual understanding
- Pair work or group work
- Using concrete materials and resources (manipulatives)
- Coaches who are government staff
- Initial face-to-face training
- Refresher face-to-face training
- Teacher training (lowest level of cascade) done by government officers
- Government staff responsible for conducting monitoring
- Program invested in capacity building at the decentralized level

NOTE ON INTERPRETATION OF FINDINGS

While findings are presented for all six programs together, evaluative comparisons should *not* be made when interpreting these findings. Each program has a unique design and is operating in a unique context. Programs were selected for this study based on evidence of effectiveness, and they represent diverse education systems and populations across three continents. Program findings are shown side by side to allow us to identify commonalities and trends, not to determine which program is "better" in any given area.

Through rigorous analyses of the primary data collected for this study, we have identified a suite of instructional practices, instructional supports, and system supports that were essential for successful programs to improve mathematics learning outcomes at scale.

Instructional Practices

- **Teachers use multiple representations and models to support learning.** Across programs, a variety of concrete materials were used, as well as images and symbolic notation. Teachers modeled and explained how materials can be used to represent and understand a concept and explicitly linked different representations, including symbolic notation. In addition, concrete materials were in the hands of students and were not just used by the teacher for demonstration.

- **Instructional approaches include a specific focus on both conceptual understanding and procedural fluency.** These are essential components to effective mathematics instruction, and both are needed to ensure learning. Conceptual understanding and procedural fluency were developed in different ways, such as through questioning and discussion, connecting formal and informal mathematics, and using concrete materials appropriately.
- **Various approaches are used to ensure active student engagement throughout lessons.** All programs engaged students throughout the math lessons, and most students were engaged in all lessons. Teachers used a variety of methods to engage students during whole-class instruction, using combinations of questions directed to individual or small groups, questions with choral answers, and some repetition. During independent work, students spent ample time practicing the skills they had just been taught.
- **Teachers use assessment-informed instruction approaches to address differentiated needs**. Most of the programs incorporated assessment in some form, though there was variability in the degree to which assessment was used to inform instruction. Teachers across programs monitored students and provided help to students who struggled during lessons—a basic, core strategy for infusing assessment-informed instruction into the classroom. Several programs utilized assessment results to adjust what instruction (RAMP, TAFITA) or problems (Nanhi Kali) students received.

Instructional Supports

- **Teacher supports focused explicitly on math content and improving instruction.** This was particularly true for teacher trainings and coaching support visits. In lieu of a historically prevalent focus on classroom management and general classroom practices, trainings included a direct focus on introducing teachers to new math content while providing them with a variety of methods for improving students' conceptual understanding of mathematics.
- **Trainings emphasize modeling and practice over lecturing, providing teachers with opportunities to practice and discuss.** Teachers were provided with ample opportunities to practice and discuss new approaches during their trainings, which they consistently reported as a valuable change from prior trainings that they had attended. Teacher meetings (within and across schools) also provided teachers with significant opportunities to discuss their new instructional approaches, which they deemed in some programs to be the most valuable aspect of meeting with other teachers.
- **Teacher and student materials provide explicit guidance for instruction.** This was addressed through the provision of structured teacher's guides and lesson plans, as well as local-language materials and teaching and learning aids for teachers and students, respectively, in nearly all programs. Teachers reported that materials were better organized, easier to follow, and more engaging than traditional materials.

- **Ongoing support emphasizes feedback, problem solving, and learning new content over inspection and evaluation.** There was some variation in how teachers were supported (including teacher meetings, coaching, mentoring and monitoring visits), but what they focused on in terms of the type of support provided to teachers, and what teachers found to be most helpful, was similar: Teachers had opportunities to get feedback, solve problems, and learn new content. Additionally, teachers reported that the individuals who provided professional development were more supportive and friendlier.

System Supports

- **Programs actively collaborate with key stakeholders.** Collaboration among stakeholders (government officials, school staff, donors, external implementers, and school communities) was essential. This collaboration took place through formal channels (such as steering committees) and informal channels (such as WhatsApp groups). The roles and responsibilities of the various stakeholders were usually well understood.
- **Investments are made in resources to improve quality classroom instruction.** Programs invested heavily in the professional development of education actors such as teachers, instructional coaches, and school management committees. They also invested heavily in providing sufficient quantities of essential teaching and learning materials such as math kits, textbooks, tablets, and teacher’s guides.
- **Programs emphasize continuous monitoring and use of data for system improvement.** Programs emphasized the importance of collecting and using data for system improvement. They collected data on teacher practice and student math outcomes. The data were then used by various education actors to determine which students needed remedial or targeted support, content areas for teacher training, and where teaching and learning materials were needed, among other things.
- **Programs focus on systematically embedding and institutionalizing best practices, with an eye toward sustainability.** Sustainability was a priority for all programs. The programs were strategically designed with sustainability in mind, aligned their activities to the government’s goals and objectives, and advocated for new policies supportive to improving learning.

Finally, based on the findings presented in this report, we present a series of recommendations for future numeracy interventions to consider. The recommendations are divided into three sections: Pedagogy Considerations, Instructional Support Considerations, and Systems Considerations.

6.1 Pedagogy Considerations

There is no one prescribed way to improve math instruction. While there is no one recipe for improving math learning outcomes, there are several pathways that lead to success. These include asking students higher-order questions, making real-life connections, using concrete materials, and providing clear and accurate explanations of content.

Involving all students in modeling and explanation is important. Students are an essential part of knowledge creation around mathematics, not passive recipients of information from the teacher. Students’ active participation during all parts of a math lesson is key to supporting mathematical discussion. Programs should encourage student participation during the modeling and explanation of new content. Strategies to encourage participation include answering questions, using materials, and responding chorally to the teacher during whole-class instruction. The involvement of students provides opportunities to discuss incorrect responses and engage in the explanation and justification of mathematical ideas.

Students need time to practice. Every lesson should ensure that students have ample time to solve problems or engage in hands-on learning activities. It is not enough for students to watch a teacher explain how to solve a problem or observe a fellow student solve a problem at the board; each student must experience it for themselves and then have the opportunity to practice solving multiple similar problems.

It's not enough to just have manipulatives. While there tends to be recognition that concrete materials, or manipulatives, are important in mathematics classrooms, these materials will not support learning unless teachers are able to use them appropriately. Teachers need training and support on how to ensure that they understand and are able to show students how the manipulatives link to the underlying math concept and how they can be used to help solve problems. Manipulatives also need to be available for all students to use, whether individually or in groups. In this regard, teacher training should also cover the management of materials, as learning can be inhibited if materials management is too distracting or takes too much time away from instruction.

There must be a strong link between concrete materials, pictures or drawings, and abstract symbols. Linked to the appropriate use of manipulatives, the ultimate goal of using different representations is to ensure that students understand concepts, can apply them to novel problems, and can use abstract mathematical symbols (numbers, operation signs, etc.). Explicitly showing linkages between these representations can help students move from, for example, using counters to solve problems to being able to solve written problems and undertaking mental arithmetic. Ensuring that students move to the abstract stage is an important element in this progression, as continuing to rely on manipulatives for even basic operations will hamper students' ability to solve more complex problems.

Assessing students is not enough; teachers need to know how to use that information to inform their instruction. Ensuring improved learning for all students demands going beyond merely assessing students to helping teachers use that information to adjust instruction and to ensure that struggling students receive the help they need.

6.2 Instructional Support Considerations

Ensuring that classrooms have high-quality teaching and learning materials will make teachers' lives easier. Every new program or intervention asks teachers to do something new (which can often be seen as something "extra"). In addition to integrating program approaches into governments' teaching and learning materials, the burden on teachers can be eased by providing materials that allow for more effortless engagement with students. Such materials include teacher's guides, as well as high-quality, contextualized student materials (including manipulatives and other concrete materials that provide students with hands-on opportunities to practice new math concepts).

Just as students need time to practice new skills in the classroom, teachers need time to practice new skills during training and teacher meetings. Learning a new skill is difficult. Learning a new instructional approach or math content for teachers is no exception. When teachers are provided with significant opportunities to practice new skills or approaches during training (in place of more lecture-based trainings), they are more comfortable with the material when they return to the classroom. It is also helpful to give teachers opportunities to discuss new approaches during school-based or cluster-based teacher meetings.

Developing teachers' MKT through training and support should be a constant focus. Good math instruction relies on a firm understanding of mathematical concepts, as well as of how students learn those concepts. However, most teachers in the early grades are not math experts, nor do they typically receive extensive training in mathematics pedagogy. By focusing on continually improving MKT through teacher training and support, programs can further enhance math instruction and ensuing student outcomes.

Ensuring that teacher supports are complementary can help overcome financial, logistical, and capacity limitations. Due in large part to financial constraints, it is often difficult for coaches to visit schools as frequently as intended. While head teachers are based in schools, competing responsibilities can limit their ability to provide continuous, targeted instructional support to all teachers. School and cluster-based teacher meetings provide an important space for teachers to interact but can be complicated by logistical considerations. Ultimately, a single support is unlikely to be a viable option for providing teachers with the guidance they need—but redundancies put unnecessary strains on education systems. Therefore, the best approach is to understand the strengths and limitations of each actor in order to ensure that they all know their relative roles (and the roles of others) and to develop complementary supports across various actors. For example, head teachers can use classroom-based data to identify struggling teachers but then find another system actor to provide the necessary support (as opposed to providing it themselves). Or more rigorous data monitoring can allow for the remote identification of struggling teachers—to be supplemented with remote support.

Making coaches friendlier and more supportive is just the beginning. This is a particularly important consideration when existing system actors such as school inspectors are used as coaches (which involves a concomitant transition of focus from accountability to support). However, while friendlier coaching is a valuable start, the quality of coaching will ultimately depend on the expertise and support provided to coaches. Therefore, existing system actors may require revisions to their job descriptions—or recruitment procedures may need to be modified—to ensure long-term, high-quality coaching.

6.3 Systems Considerations

Data and evidence should be used to inform decisions. Data should be systematically collected from evaluations, learning assessments, and classroom observations and then used to inform program design, professional development topics, and the development of teaching and learning materials.

The education system should focus on professional development for education staff. Extensive professional development is important for teachers, school administrators, and middle managers. This professional development should include a mix of in-person trainings, classroom coaching, and virtual support for teachers. Additionally, the content and delivery should be intentionally designed to improve instruction and student learning.

Designing for scale and implementing through government systems are necessary steps for achieving improved learning outcomes at scale. For programs that are not directly initiated by the government, successful implementation at scale typically requires designing activities with scalability in mind. This means involving education system actors right from the outset, aligning with government policies and processes, and implementing through government systems and actors.

Governments and external funders should invest strategically in resources to bolster quality instruction. Governments and external funders alike should invest significant resources (financial, technical, and human) in areas that evidence shows contribute to quality instruction and improved learning outcomes. These include high-quality teaching and learning materials, professional development opportunities for system actors, and instructional support for teachers. For government-run programs, the government usually provides the resources. For programs involving external partners, the government can still contribute financially, such as through the procurement and distribution of materials. And whatever the type of program, the government can invest significant resources through the allocation of staff time to training, instructional support, and monitoring.

RISKS OF SCALING

While we were not able to compare pilot implementation to implementation at scale in all countries, our analyses did point to some risk of dilution as more schools receive resources and more teachers are trained. For example, in El Salvador, some of the strongest instructional practices were more apparent in original pilot schools (when compared to scaled schools).

Remediation programs can help build basic skills in students who are struggling, and TAFITA teachers demonstrated strong instructional practice during remediation sessions. However, evidence from this study did not show an impact of the remediation program on core instruction. The risk here is that every new cohort will continue to need remediation unless core instruction is also addressed.

REFERENCES

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching* (2nd ed.). Association for Supervision and Curriculum Development.

de Barros, A., Fajardo-Gonzalezez, J., Glewwe, P., Sankar, A. (2023). "The Limitations of Activity-Based Instruction to Improve the Productivity of Schooling." *The Economic Journal*, June 2023. https://de-barros.com/publication/de-barros-the-2023/de%20Barros%20et%20al.%20-%202023%20-%20The%20Limitations%20of%20Activity-Based%20Instruction.pdf

JET Education Services & Kelello Consulting. (2018). *Evaluation of the Grade R Mathematics Project implemented in the Western Cape: Endline report July 2018*. Zenex.

Global Reading Network. https://shared.rti.org/content/instructional-strategies-mathematics-early-grades

JET Education Services & Kelello Consulting. (2018). *Evaluation of the Grade R Mathematics Project implemented in the Western Cape: Endline report July 2018*. Zenex.

Loewenberg Ball, D., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*(5), 389–407 https://doi.org/10.1177/0022487108324554

Maruyama, T., & Kurosaki, T. Forthcoming. Developing Textbooks to Improve Math Learning in Primary Education: Empirical Evidence from El Salvador. *Economic Development and Cultural Change*. https://doi.org/10.1086/721768

Maruyama, T., & Igei, K. Forthcoming. Community-wide support for primary students to improve Foundational Literacy and Numeracy: Empirical evidence from Madagascar. *Economic Development and Cultural Change*. https://doi.org/10.1086/726178

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276–282. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/

Muralidharan, K., Singh, A., & Ganimian, A. J. (2019). "Disrupting education? Experimental evidence on technology-aided instruction in India." *American Economic Review*, *109*(4), 1426–1460. DOI: 10.1257/aer.20171112

Ralaingita, W., Mamytova, A., & Sitabkhan, Y. (2023). Capturing teachers' mathematical knowledge for teaching. https://shared.rti.org/content/mathematical-knowledge-teaching-survey-cies-2023-presentation

Roberts, N., & Mawoyo, M. (2020). *R-Maths consolidation phase evaluation report*. Kelello.

RTI International. (2019). *Early Grade Reading and Mathematics Initiative: Endline survey report*. USAID. https://pdf.usaid.gov/pdf_docs/PA00WPQD.pdf

University of Michigan Teaching & Learning Exploratory. (n.d.) *High leverage practices overview*. https://tle.soe.umich.edu/HLP

Vaijayanti, K., Suma, M.N., Mondal, A. (2016). *The Impact of Akshara Ganitha: A Longitudinal Study 2012-13 to 2014-15.* Akshara Foundation.

ANNEX A. ADDITIONAL TABLES FROM INTERVIEWS

A.1 Classroom Observation (Additional Items)

Table A-1. Classroom Observations

	GKA (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
	n= 78	n= 154	n= 80	n= 80	n= 70
Teacher Gender					
Male	47.40%	22.10%	11.20%		25.70%
Female	52.60%	77.90%	88.80%		74.30%
Grades					
Grade R	0%	0%	0%	100%	0%
Grade 1	0%	0%	0%	0%	18.60%
Grade 2	0%	55.20%	98.70%	0%	77.10%
Grade 3	0%	31.80%	1.30%	0%	2.90%
Grade 4	23.10%	0%	0%	0%	0%
Grade 5	43.60%	0%	0%	0%	1.40%
Mixed grades (multiple)	32.10%	13%	0%	0%	0%
No grade - afterschool group	1.30%	0%	0%	0%	0%
How many times did the teacher ask learners a question that does not have one correct answer? (mean)	2	2.6	3.2	1.5	3.7
Did any learners answer a teacher question with an incorrect or no response?					
Yes more than 3 times	21.80%	47.40%	55%	15%	48.60%
Yes 1-2 times	64.10%	34.40%	36.20%	55%	28.60%
No	14.10%	18.20%	8.80%	30%	22.90%

	GKA (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Did the teacher model/explain how to solve a problem, a concept, an activity?					
Yes	100%	90.90%	100%	93.70%	90%
No	0%	9.10%	0%	6.30%	10%
Did the teacher ever make a maths error?					
Yes	15.40%	2.90%	6.30%	8%	6.30%
No	84.60%	97.10%	93.70%	92%	93.70%
Did the teacher ever connect the math concept to real-life examples or the lives of learners?					
Yes	93.60%	41.40%	71.30%	60%	44.40%
No	6.40%	58.60%	28.70%	40%	55.60%
Did the teacher intentionally ever introduce an incorrect solution?					
Yes	17.90%	5.80%	8.80%	13.80%	2.90%
No	82.10%	94.20%	91.20%	86.20%	97.10%
Did the teacher introduce or use a learners' incorrect response to discuss why it is incorrect?					
Yes	34.60%	11.70%	55%	23.80%	7.10%
No	65.40%	88.30%	45%	76.30%	92.90%
Look at 3-5 books (or observe students doing an activity) from students seated around you. Are the students solving the problems/doing the activity correctly?					
Yes most of the students are solving problems/doing the activity correctly	96.20%	88.90%	93.90%	98.40%	90.50%
No most of the students are NOT solving problems /doing the activity correctly	3.80%	11.10%	6.10%	1.60%	9.50%

	GKA (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
What math-related materials were visible in the classroom?					
Posters	59%	70.10%	80%	96.20%	75.70%
Concrete materials	14.10%	27.90%	57.50%	71.30%	24.30%
None	30.80%	20.80%	13.80%	2.50%	12.90%
Other	7.70%	3.20%	1.30%	12.50%	12.90%
What was the sitting arrangement?					
learners at individual desks	17.90%	85.10%	5%	10%	2.90%
learners on the floor	41%	0.60%	0%	90%	4.30%
learners at desks with multiple learners per desk/table/chair facing front of room	66.70%	5.20%	88.70%	6.30%	87.10%
small groups (learners facing each other)	2.60%	11%	6.30%	63.70%	2.90%
other	0%	0.60%	0%	1.30%	5.70%
Did the teacher refer to a teacher's guide during the lesson?					
Yes	37.20%	72.70%	6.30%	16.30%	51.40%
No	62.80%	27.30%	93.70%	83.70%	48.60%
Did the learners use a textbook during the lesson?					
Yes	64.10%	82.50%	52.50%	0%	4.30%
No	35.90%	17.50%	47.50%	100%	95.70%
Did learners use a workbook during the lesson?					
Yes	59%	27.90%	21.30%	7.50%	20%
No	41%	72.10%	78.80%	92.50%	80%
If yes, did learners write in the learner's workbook?					
Yes	95.70%	95.30%	88.20%	100%	92.90%
No	4.30%	4.70%	11.80%	0%	7.10%

	GKA (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Did learners use concept cards/flash card?					
Yes	44.90%				38.60%
No	55.10%				61.40%
Number of boys present (mean)	9	8.5	12.3	11.6	10.5
Number of girls present (mean)	12.1	7.7	13	12.6	24.2
Number of textbooks (mean)	18.7	13.3	23.8	0.1	5.6
Number of workbooks (mean)	17	6.5	20.1	1.2	13.1

A.2 Coaching/Mentoring Interview (Additional Items)

Table A-2. Coaching/Mentoring Interview

Description	GKA (India)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
	n= 5	n= 15	n= 4	n= 50
What is your main job?				
Teacher coach – program	100%	0%		
Teacher coach – government	0%	33.30%		
Supervisor / inspector	0%	66.70%		
Other government official	0%	0%		
Other program staff	0%	0%		
Teacher	0%	0%		
Head Teacher	0%	0%		
Other	0%	0%		
In as few words as possible, how would you describe your main purpose as a coach?				
School inspection	20%	6.70%	0%	9.80%

Description	GKA (India)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
To improve teaching in schools	80%	86.70%	33.30%	75.60%
To evaluate teachers	20%	60%	0%	39%
To ensure that teachers are teaching the right lesson	40%	66.70%	0%	61%
To inform teachers about the mistakes they make in the classroom	40%	53.30%	0%	41.50%
Other	0%	6.70%	66.70%	19.50%
Have you been a coach or provided teacher support in another role prior to the program?				
Yes	60%	86.70%	100%	90.20%
No	40%	13.30%	0%	9.80%
Do you provide regular support to the same number of classrooms as before the program (more/fewer)?				
No change		7.70%	33.30%	27%
I support fewer classrooms now		7.70%	0%	10.80%
I support more classrooms now		84.60%	66.70%	62.20%
Have you received training as a coach? (If so, how often do you receive training on coaching?)				
I have never received training on coaching	40%	6.70%	33.30%	17.10%
Less than once per year	0%	0%	0%	22%
Once per year	0%	6.70%	0%	17.10%
More than once per year	20%	86.70%	66.70%	43.90%
What was the most useful aspect of the training you received as a coach?				
How to give feedback to teachers	0%	35.70%	0%	2.90%
How to help teachers reflect on practice	33.30%	42.90%	100%	47.10%
How to build rapport with teachers	66.70%	0%	0%	2.90%

Description	GKA (India)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Process of visiting schools	0%	7.10%	0%	17.60%
How to use observation tool	0%	7.10%	0%	2.90%
How to assess learners	0%	7.10%	0%	8.80%
How to prioritize which teachers/schools to visit	0%	0%	0%	17.60%
Do you feel that you have too many, too few, or just the right number of teachers to support?				
It is too many teachers	40%	53.30%	100%	46.30%
It is too few teachers	0%	0%	0%	9.80%
It is the right number of teachers	60%	46.70%	0%	43.90%
What materials, if any, does the program provide to help you work with the teachers?				
Tablet	100%	80%	0%	24.40%
Observation tools	0%	60%	0%	51.20%
Teachers' guides	20%	73.30%	66.70%	4.90%
Student books	40%	46.70%	0%	
Manuals	40%	66.70%	66.70%	
Videos	0%	33.30%	0%	
Chat groups (whatsapp telegram etc)	0%	60%		
Student usage data/reports	40%			
Other	0%	0%	66.70%	41.50%
Which of the following do you regularly do when you visit a school?				
Review lesson plan	20%	93.30%	0%	58.50%
Check teacher materials	80%	86.70%	66.70%	65.90%
Check student materials	40%	73.30%	0%	34.10%
Observe lesson	60%	100%	66.70%	95.10%

Description	GKA (India)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Assess children	80%	93.30%	0%	17.10%
Debrief session with individual teacher	20%	93.30%	66.70%	61%
Debrief session with all teachers together	20%	80%	0%	46.30%
Debrief session with head teacher	20%	60%	33.30%	58.50%
What do you normally do during lesson observation?				
I do not observe lessons	0%	0%	0%	0%
Sit and observe lesson	80%	73.30%	33.30%	78%
Fill out observation form	0%	80%	0%	53.70%
Take notes	20%	93.30%	33.30%	75.60%
Correct teacher mistakes (during lesson)	20%	73.30%	0%	36.60%
Check-on and help learners	0%	66.70%	0%	29.30%
Other	0%	13.30%	100%	2.40%
After you observe a lesson do you talk with the teacher about what you saw?				
Yes	100%	93.30%	100%	100%
No	0%	6.70%	0%	0%
What are the steps or topics you cover in the post-lesson conversation?				
Ask teacher to reflect on their teaching	40%	71.40%	0%	58.50%
Give at least one positive about the lesson	20%	78.60%	0%	80.50%
Give areas for improvement	80%	92.90%	66.70%	92.70%
Model activities	20%	57.10%	0%	34.10%
Discuss teacher questions/challenges faced	40%	78.60%	0%	73.20%
Discuss student progress	20%	64.30%	0%	24.40%
Discuss lesson methodology/the instructional approach used	60%	64.30%	66.70%	22%

Description	GKA (India)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Support teacher to understand math content	60%	64.30%	100%	36.60%
Discuss use of materials	20%	64.30%	0%	58.50%
Discuss classroom management	40%	64.30%		51.20%
Discuss administrative matters	40%	35.70%		9.80%
Agree on way forward	0%	85.70%		19.50%
Other	0%	0%	33.30%	0%
Do you regularly follow up on teacher's progress from previous visits?				
Yes	80%	100%	66.70%	68.30%
No	20%	0%	33.30%	31.70%
How do you know what topics to follow up on?				
It’s recorded on the form	25%	66.70%	100%	28.60%
I take notes each time and refer back	0%	93.30%	50%	71.40%
I'm prompted by the tablet to do so	25%	13.30%	0%	0%
I remember things we previously discussed	50%	33.30%	0%	39.30%
Teacher records notes in lesson plan/book	25%	53.30%	0%	25%
other	0%	0%	50%	10.70%
Why don’t you follow up on teacher’s progress?				
The form doesn't require it	0%			15.40%
Too much time between visits	100%		0%	7.70%
I haven't received training on how to do this	0%		0%	7.70%
Other	0%		100%	84.60%
What do you do with the information you collect during a coaching visit?				
Nothing	0%	0%	0%	0%

Description	GKA (India)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Send written report to program/ministry	0%	80%	100%	39%
Provide written report/comments to school	0%	66.70%	100%	29.30%
Use in debrief/discussion	40%	60%	33.30%	24.40%
Use for follow-up visits	0%	80%	33.30%	48.80%
Feed into school improvement plans	40%	80%	0%	22%
Upload results	20%	60%	0%	2.40%
Use during cluster meetings or training	40%	80%	33.30%	24.40%
Other	0%	6.70%	0%	17.10%
What do you do if a teacher is struggling with implementing their instruction?				
Nothing	0%	0%	0%	2.40%
Visit that teacher more often than others	20%	26.70%	33.30%	29.30%
Provide targeted support/training for the teacher (including modeling best practices)	80%	93.30%	33.30%	61%
Inform others (e.g. school director, program staff)	20%	46.70%	33.30%	26.80%
Provide teacher with additional resources	40%	66.70%	33.30%	19.50%
Suggest that teacher reaches out to fellow teachers or have other teacher check on them	60%	53.30%	33.30%	26.80%
Other	0%	0%	33.30%	4.90%
What motivates you to visit classrooms?				
It is my job	40%	60%	0%	61%
To help teachers	60%	46.70%	33.30%	65.90%
To improve teaching	100%	86.70%	66.70%	65.90%
To improve learning	80%	80%	0%	29.30%
Reimbursement/allowances	20%	0%	33.30%	4.90%
I am passionate about my work	20%	53.30%		29.30%

Description	GKA (India)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Other	0%	0%	66.70%	9.80%
Do you feel the coaching is effective in improving the teachers' instruction compared to before the program started?				
Yes	80%	100%	100%	100%
No	20%	0%	0%	0%
What about the coaching is helpful for improving teaching?				
Post-observation discussion	0%	66.70%	33.30%	51.20%
Providing teacher feedback (praise/improvement)	50%	93.30%	33.30%	46.30%
Being able to follow-up/visit regularly	50%	66.70%	0%	39%
Teachers improve when they are observed	50%	73.30%	0%	56.10%
Other	0%	0%	33.30%	0%
How do you think the coaching process could be improved?				
Increased frequency of visits	60%	66.70%	100%	56.10%
Additional training for coaches	20%	53.30%	33.30%	56.10%
Increased allowances (e.g. transport etc.)	0%	6.70%	0%	63.40%
Increased support from program staff	40%	40%	33.30%	29.30%
Reduced workload (number of teachers; other responsibilities)	40%	33.30%	0%	19.50%
Other	0%	13.30%	33.30%	12.20%
Does anyone from the government ever come to support you during a school/site monitoring visit?				
Yes	20%	13.30%		12.20%
No	80%	86.70%		87.80%

A.3 Trainer Interview

Table A-3. Trainer Interview

	GKA (India)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
	n= 2	n= 15	n= 4	n= 50
What is your main job?				
Teacher coach – program	0%		0%	7.50%
Teacher coach – government	50%		50%	42.50%
Trainer	0%		50%	10%
Other government official	0%		0%	5%
Other program staff	0%		0%	30%
Head teacher	0%		0%	0%
Teacher	0%		0%	0%
Other	50%		0%	5%
Have you been a trainer in any other program or position?				
Yes	50%	66.70%	100%	90%
No	50%	33.30%	0%	10%
How do program trainings compare to other teacher trainings you've been involved in?				
No differences	100%	0%	0%	5.60%
Better overall	0%	80%	50%	58.30%
Worse overall	0%	0%	0%	2.80%
Frequency - more regular	0%	50%	25%	2.80%
Frequency – less regular	0%	10%	0%	8.30%
Organization - better	0%	70%	25%	44.40%

	GKA (India)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Organization – worse	0%	70%	50%	0%
Preparation (manual training etc.) - better	0%	10%	0%	27.80%
New training approaches used	0%	0%		44.40%
Content is more useful	0%	60%		47.20%
Allowances - better (transport per diem etc.)	0%	80%		22.20%
Other	0%	0%	25%	5.60%
Who trained or prepared you to train the teachers?				
Program staff	100%	73.30%		65%
Government staff	0%	40%	100%	30%
University lecturer	0%	0%	75%	30%
International specialist	0%	20%	0%	2.50%
Other	0%	0%	0%	0%
Did you receive a training manual?				
Yes	0%	100%	100%	95%
No	100%	0%	0%	5%
What aspects of the training manual make it easy for you as a trainer?				
None/not useful		0%	0%	0%
Timetable		60%	25%	28.90%
Step-by-step instructions		66.70%	75%	71.10%
Useful activities and/or practice lessons		80%	25%	55.30%
Clear explanations of concepts		73.30%	75%	44.70%
Roles are clearly defined		60%	75%	28.90%
The entire manual is helpful		80%	0%	50%
Other		0%	50%	13.20%

	GKA (India)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
What aspects of the training manual would you improve?				
None		33.30%	50%	28.90%
Too much information		20%	0%	10.50%
Too long		26.70%	0%	7.90%
Not enough instructions		20%	0%	21.10%
No clear timetable		6.70%	0%	0%
No digital copies		6.70%	0%	15.80%
Too complicated		0%	0%	2.60%
Other		26.70%	50%	42.10%
Does anyone from the program staff come to support you during training?				
Yes	100%	66.70%		42.50%
No	0%	33.30%		57.50%
What do they (Program Staff) do at the training?				
Nothing	0%	20%		0%
Observe	50%	40%		58.80%
Give feedback/advice	50%	50%		64.70%
Help train or model teaching	100%	60%		58.80%
Evaluate your training	0%	60%		29.40%
Have discussions with teachers	0%	40%		41.20%
Other	0%	0%		17.60%
Does anyone from the government come to support you during training?				
Yes	50%	40%	100%	27.50%
No	50%	60%	0%	72.50%

	GKA (India)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
What do they (Government Staff) do at the training?				
Nothing	0%	0%	0%	0%
Observe	0%	66.70%	25%	36.40%
Give feedback/advice	100%	83.30%	0%	45.50%
Help train or model teaching	0%	66.70%	50%	63.60%
Evaluate your training	0%	33.30%	25%	0%
Have discussions with teachers	0%	50%	0%	18.20%
Give a speech	0%	0%	0%	
Other	0%	0%	100%	0%
Do teachers practice new skills during trainings?				
Yes	100%	93.30%	100%	95%
No	0%	6.70%	0%	5%
Do teachers practice in front of a large group during trainings?				
Yes	50%	80%	100%	92.50%
No	50%	20%	0%	7.50%
Do teachers practice in small groups or pairs during trainings?				
Yes	50%	80%	75%	85%
No	50%	20%	25%	15%
Do teachers practice solving math problems to learn new math content?				
Yes	100%	86.70%	100%	95%
No	0%	13.30%	0%	5%
What do you do if teachers don't know how to do the math in the activities?				
Pair them with another teacher	100%	20%	25%	35%

	GKA (India)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Re-teach the content one on one	100%	66.70%	25%	27.50%
Review the content with all teachers	50%	53.30%	25%	62.50%
Direct them to the content in a teachers guide or textbook	0%	73.30%	0%	25%
Refer them to another trainer	0%	6.70%	0%	5%
Other	0%	0%	50%	7.50%
Nothing	0%	0%	0%	0%
How many teachers practiced implementing instructional activities during your last training?				
All (or nearly all)	0%	46.70%	100%	60%
More than half	100%	33.30%	0%	20%
Approximately half	0%	13.30%	0%	15%
Less than half	0%	6.70%	0%	5%
None	0%	0%	0%	0%
How many teachers demonstrated instructional activities to at least one other teacher during your last training?				
All (or nearly all)	0%	20%	50%	32.50%
More than half	0%	26.70%	50%	40%
Approximately half	50%	13.30%	0%	10%
Less than half	50%	40%	0%	10%
None	0%	0%	0%	7.50%
Which training method do you use most in program trainings? (Read options only if needed for clarification)				
Lecture	0%	0%	0%	0%
Discussion	0%	6.70%	25%	2.50%
Modeling/demonstration	100%	46.70%	50%	52.50%
Small-group practice	0%	13.30%	0%	20%

	GKA (India)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Large-group practice	0%	26.70%	25%	22.50%
In-school practice	0%	6.70%	0%	2.50%
Which method do you think is the most useful to teach a new instructional approach? (Read options only if needed for clarification)				
Lecture	0%	6.70%	0%	5%
Discussion	100%	6.70%	0%	5%
Modeling	0%	60%	100%	40%
Small-group practice	0%	26.70%	0%	22.50%
Large-group practice	0%	0%	0%	22.50%
In-school practice	0%	0%	0%	5%
Can you tell me 2-3 things that make program trainings more effective than others?				
Nothing	0%	6.70%	0%	2.50%
More focus on skills	50%	80%	0%	20%
More practice	100%	73.30%	50%	57.50%
More frequent training	50%	60%	0%	22.50%
Better allowances (transport per diem etc.)	0%	26.70%	25%	27.50%
More accountability	0%	46.70%	25%	10%
More teachers attending	0%	40%	0%	20%
More teacher engagement	50%	53.30%	25%	37.50%
Knowledgeable/well-trained trainers	100%	20%	50%	25%
Other	0%	0%	75%	20%
Name 2-3 things that you would change to improve the training?				
Nothing	0%	0%	75%	0%
Include more activities	100%	46.70%	0%	5%

	GKA (India)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Include more practice	100%	60%	0%	27.50%
Less lecture	0%	33.30%	0%	7.50%
Better allowances (transport per diem etc.)	0%	46.70%	0%	60%
Reduce number of participants	0%	20%	0%	5%
Increase use of technology	100%	33.30%	0%	27.50%
Include more discussion	0%	40%	0%	10%
More frequent trainings	0%	60%	0%	35%
Other	0%	13.30%	25%	45%
Did training activities change due to COVID-19?				
Yes	100%	93.30%	50%	55%
No	0%	6.70%	50%	45%
If yes, how did trainings change?				
Additional health precautions taken	0%	57.10%	0%	54.50%
Smaller (in person) trainings	0%	57.10%	0%	31.80%
Moved to virtual meetings (via Zoom Skype Google Meeting etc.)	100%	71.40%	50%	0%
Moved to new technology-based trainings that teachers did on their own	0%	42.90%	50%	4.50%
Sent training guides and other materials to teachers for self-study	0%	35.70%	0%	18.20%
Set up chat groups (via Whatsapp Telegram FB)	100%	57.10%	50%	4.50%
Posted training videos (via YouTube Facebook etc)	100%	42.90%	0%	0%
Phone calls with teachers	0%	57.10%	0%	18.20%
Other	0%	0%	50%	22.70%
Are any of the changes introduced during COVID being sustained by the program? If so, which ones?				
/None	0%	14.30%	100%	

	GKA (India)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Additional health precautions taken	50%	35.70%	0%	45.50%
Smaller (in person) trainings	0%	14.30%	0%	22.70%
Moved to virtual meetings (via Zoom Skype Google Meeting etc)	100%	35.70%	0%	4.50%
Moved to new technology-based trainings that teachers did on their own	0%	35.70%	0%	4.50%
Sent training guides and other materials to teachers for self study	0%	28.60%	0%	9.10%
Set up chat groups (via Whatsapp Telegram FB)	100%	50%	0%	0%
Posted training videos (via YouTube Facebook etc)	50%	35.70%	0%	0%
Phone calls with teachers	0%	21.40%	0%	9.10%
Other	0%	0%	0%	50%

A.4 Teacher Interview (Additional Items)

Table A-4. Teacher Interview

	GKA (India)	Nanhi Kali (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
	n= 79	n= 79	n= 150	n= 79	n= 79	n= 80
Teacher Gender						
Male	48.1%	15.2%[11]	23.3%	12.7%	0%	23.8%
Female	51.9%	84.8%	76.7%	87.3%	100%	76.3%
What grades(s) do you teach?						

[11] in the total Nanhi Kali program there are 0.3% male CAs.

	GKA (India)	Nanhi Kali (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Pre-primary	0%	12.7%	2.7%	0%	3.8%	2.5%
Grade 1	3.8%	77.2%	6%	8.9%	6.3%	16.3%
Grade 2	3.8%	86.1%	60.7%	96.2%	1.3%	88.7%
Grade 3	8.9%	86.1%	38%	7.6%	0%	6.3%
Grade 4	70.9%	83.5%	4%	0%	0%	5%
Grade 5 or above	77.2%	60.8%	6%	0%	0%	3.8%
Grade R					96.2%	
What is your highest level of education?						
University degree (or higher)	77.2%	55.7%	90.7%	96.2%	17.7%	0%
Post secondary (not university)	20.3%	30.4%	9.3%	3.8%	64.6%	0%
Secondary	2.5%	13.9%	0%	0%	16.5%	7.5%
Below Secondary	0%	0%	0%	0%	1.3%	61.2%
How many years have you been a teacher? (mean)	19.5	3.9	23.1	13	13.4	9.6
Which of the following supports do you receive from this Program?						
Training (in-person)	46.8%	68.4%	84.7%	91.1%	68.4%	81.2%
Program Training (virtual)		16.5%	36%	17.7%	19%	0%
Coaching/Mentoring (in-person; internal to school)			36%	39.2%	24.1%	28.7%
Coaching/Mentoring (in-person; external coach such as Subject Advisor or Curriculum Advisor)		29.1%	36.7%	46.8%	49.4%	27.5%
SMS or other remote coaching/advice/ guidance		13.9%	20%	25.3%	12.7%	5%
Teacher materials (teachers' guide; lesson	59.5%	40.5%	55.3%	50.6%	78.5%	60%

	GKA (India)	Nanhi Kali (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
plans; etc.)						
Student materials (textbooks storybooks manipulatives etc.)	88.6%	44.3%	18.7%	38%	73.4%	16.3%
Teacher meetings (e.g. cluster or block meetings held for teachers; NOT staff meetings)	31.6%	29.1%	73.3%	39.2%	27.8%	30%
Teaching and Learning Materials Kit	72.2%	72.2%	20.7%	21.5%	38%	
Podcast	7.6%	12.7%	32%			
Videos	8.9%					
None of the above	2.5%	0%	1.3%	1.3%	8.9%	2.5%
What teacher materials have you been given by the program?						
None	0%	0%	7.30%	7.60%	6.30%	3.80%
Teachers' guide	53.20%	59.50%	52.70%	72.20%	81%	87.50%
Tablet	0%	81%	60.70%	1.30%	1.30%	0%
Manipulatives and other teacher aids (e.g. flash cards counters base-ten blocks etc.)	92.40%	12.70%	13.30%	63.30%	81%	32.50%
Supplies to develop your own materials (e.g. posters markers scissors)	26.60%	27.80%	45.30%	43%	36.70%	25%
Instructions or training on making materials using low-cost local materials.	20.30%		30.70%	26.60%	49.40%	6.40%
Teacher's version of textbook	24.10%	60.80%	18.70%	20.30%	59.50%	12.50%
Lesson plans	24.10%	29.10%		10.10%	27.80%	8.80%
Lesson notes	10.10%	15.20%				3.70%
Other	0%	0%	3.30%	2.50%	7.60%	6.30%

	GKA (India)	Nanhi Kali (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Which one of these teacher materials do you feel is the MOST useful?						
Teachers' guide	10.1%	10.1%	38.5%	37%	16.7%	54.5%
Tablet	0%	54.4%	22.2%	0%	0%	9.1%
Manipulatives and other teacher aids (e.g. flash cards counters base-ten blocks etc.)	73.4%	6.3%	9.6%	54.8%	58.3%	14.3%
Supplies to develop your own materials (e.g. posters markers scissors)	2.5%	6.3%	8.9%	4.1%	8.3%	5.2%
Instructions or training on making materials using low-cost local materials.	3.8%	0%	16.3%	2.7%	0%	9.1%
Teacher's version of textbook	5.1%	20.3%	4.4%	0%	8.3%	3.9%
Lesson plans	3.8%	1.3%	0%	0%	8.3%	2.6%
Lesson notes	0%	1.3%	0%	0%	0%	0%
Other	0%	0%	0%	0%	0%	1.3%
None	1.3%	0%	0%	1.4%	0%	0%
What student materials have you been given by the program?						
Textbooks	31.6%	20.3%	90%	26.6%	48.1%	2.5%
Workbooks	57%	45.6%	42%	51.9%	36.7%	1.2%
Manipulatives	91.1%	5.1%	9.3%	50.6%	83.5%	10%
Student exercise books	38%	65.8%	32%	49.4%	24.1%	2.5%
Stationary (notebooks etc.)	7.6%		37.3%	35.4%	19%	1.2%
Other	0%	1.3%	2%	3.8%	6.3%	8.7%
None	2.5%	0%	2%	11.4%	8.9%	80%
Not applicable/Unable to respond					0%	

	GKA (India)	Nanhi Kali (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Which one of these student materials do you feel is the MOST useful?						
Textbooks	3.9%	3.8%	74.10%	8.6%	13.9%	5.6%
Workbooks	18.2%	15.2%	8.20%	20%	8.3%	5.6%
Manipulatives	71.4%	3.8%	6.10%	50%	72.2%	88.9%
Student exercise books	6.5%	15.2%	8.20%	15.7%	1.4%	0%
Stationary (notebooks etc.)	0%	0%	2%	2.9%	1.4%	0%
Tablet and Mindspark/eLearning application	0%	60.8%	0%	0%	0%	0%
Other	0%	1.3%	1.40%	2.9%	2.8%	0%
None	0%	0%	0%	0%	0%	0%
Now, how is the program helping students to catch up and improve their learning since coming back to school after school closures?						
Not doing anything different	1.3%	1.3%	4.7%	12.8%	16.7%	16.3%
Remedial/catch up materials	44.3%	22.8%	68.9%	79.5%	52.8%	33.7%
Learning recovery model	72.2%					21.2%
Reduced competencies to focus on	15.2%	32.9%	11.5%	16.7%	11.1%	2.1%
Additional time for math	70.9%	77.2%	58.1%		37.5%	35%
More homework	43%	32.9%	21.6%	16.7%	31.9%	47.5%
Other	0%	0%	6.8%	5.1%	18.1%	7.5%
Has your regular class instruction changed since you started working with this Program?						
No difference	1.30%		3.40%	5.10%	6.90%	8.80%
More focus on having students explore and solve problems	78.50%		57.40%	39.70%	54.20%	32.50%
More focus on using multiple strategies	75.90%		52%	64.10%	61.10%	46.20%

	GKA (India)	Nanhi Kali (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
More active learning/less lecture	29.10%		46.60%	33.30%	43.10%	48.80%
More pair/group work	29.10%		46.60%	33.30%	43.10%	48.80%
New methodology/instructional approach	48.10%		25.70%	60.30%	36.10%	28.80%
Involves more materials /activities (given to me)	54.40%		60.10%	21.80%	47.20%	40%
Involves more materials /activities (that I make)	43%		29.90%	51.30%	36.10%	35%
Not applicable	0%		0%	0%	2.80%	2.50%
Other	0%		1.40%	2.60%	8.30%	0%
What change have you seen in your students since the program started (If so, how?)						
No impact	0%	0%	4.1%	6.4%	1.4%	5%
Learners understand basic math concepts better	77.2%	51.9%	55.4%	42.3%	69.4%	50%
Learners have improved performance on basic operations	59.5%	38%	52.7%	56.4%	51.4%	52.5%
Learners are better at solving problems	81%	34.2%	45.9%	48.7%	65.3%	40%
Learners are better at explaining how they solved a problem	51.9%	24.1%	34.5%	32.1%	30.6%	28.7%
Student attendance has improved	45.6%	62%	16.2%	23.1%	22.2%	50%
Student behavior is better	29.1%	53.2%	27.7%	28.2%	29.2%	18.8%
Students are more engaged	59.5%	38%	35.8%	52.6%	59.7%	62.5%
Students like the lessons/activities	55.7%	25.3%	46.6%	34.6%	51.4%	43.8%
Students enjoy math more	45.6%	40.5%	52.7%	55.1%	54.2%	63.8%
Students are more confident	34.2%	27.8%	44.6%	16.7%	43.1%	26.2%

	GKA (India)	Nanhi Kali (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Not applicable	0%					
Other	0%	0%	.7%	1.3%	2.8%	0%
What part of your instruction has had the biggest impact on student learning?						
More focus on having students explore and solve problems	35.4%	22.8%	43.1%	21.6%	26.9%	6.6%
More focus on using multiple strategies	31.6%	17.7%	16.7%	39.2%	19.2%	5.3%
More learner-centered, less lecture	7.6%	6.3%	8.3%	13.5%	21.8%	23.7%
More pair/group work	11.4%	3.8%	9%	8.1%	12.8%	5.3%
New methodology/instructional approach	5.1%	46.8%	13.2%	9.5%	7.7%	25%
Involves more materials or activities	8.9%	2.5%	9%	8.1%	6.4%	31.6%
Other	0%	0%	.7%	0%	5.1%	2.6%
Was the last training program you received in-person or virtual?						
In-person attended training	45.9%	86.7%	84.6%	72.6%	94.8%	100%
In-person given by a teacher at my school	13.5%	0%	.7%	2.7%	0%	0%
Virtual	37.8%	11.7%	14.7%	24.7%	5.2%	0%
Don’t know	2.7%	1.7%	0%	0%	0%	0%
Has your teaching changed as a result of the coaching you’ve received as part of the program?						
Yes		95.7%	94%	90%	89.4%	96.2%
No		4.3%	6%	10%	10.6%	3.8%
How often did you receive remote coaching/guidance (e.g. SMS, email, etc.) in this program now?						

	GKA (India)	Nanhi Kali (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Daily		39.1%	4.4%	4.2%	0%	0%
Weekly		17.4%	6.7%	8.3%	4%	3.1%
Monthly		21.7%	22.2%	25%	12%	18.8%
A few times per year		17.4%	22.2%	20.8%	32%	31.2%
Yearly (or less)		0%	2.2%	6.3%	8%	3.1%
I don't remember		0%	4.4%	0%	10%	0%
Never		4.3%	37.8%	35.4%	34%	43.7%
Tablets and software/eLearning application	0%	48.1%	0%	0%	0%	0%
Performance dashboards/data/reports	0%	1.3%	0%	0%	0%	0%
Other	0%	0%	2.7%	2.6%	2.8%	2.5%

A 5. Head Teacher Interview

Table A-5. Head Teacher Interview

	GKA (India)	Nanhi Kali (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
	n=79	n=79	n=89	n=80	n=80	n=77
Headteacher gender						
Male	51.9%	67.1%	46.1%	11.3%	53.8%	33.8%
Female	48.1%	32.9%	53.9%	88.7%	46.2%	66.2%
Has this program been successful in improving children's learning in your school?						
Yes	100.0%	98.7%	98.9%	94.9%	87.5%	98.7%

	GKA (India)	Nanhi Kali (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
No	0.0%	1.3%	1.1%	5.1%	12.5%	1.3%
What role have you played in helping the program improve learning in your school?						
Nothing	0.0%	2.6%	0.0%	3.9%	5.7%	0.0%
Provide information to teachers on program expectations.	63.3%	59.0%	53.4%	50.0%	60.0%	53.9%
Monitor teachers in their classrooms	81.0%		52.3%	59.2%	51.4%	61.8%
Provide information about the school's progress to program staff.	59.5%	44.9%	29.5%	32.9%	37.1%	30.3%
Provide information about the school's progress to ministry staff.	27.8%	29.5%	23.9%	15.8%	14.3%	36.8%
Coach/mentor teachers	44.3%		26.1%	69.7%	31.4%	64.5%
Facilitate professional development activities (training, learning circles, etc.)	36.7%		59.1%	50.0%	44.3%	10.5%
Other	0.0%	6.4%	2.3%	2.6%	18.6%	13.2%
What type of support, if any, was given to you to perform these roles?						
Training from program	79.7%	75.3%	75.3%	49.4%	28.9%	74.0%
Training from district office	21.5%	15.6%	29.2%	16.9%	27.6%	63.6%
Coaching support	59.5%	48.1%	22.5%	24.7%	28.9%	35.1%
Admin or logistics support	16.5%	9.1%	27.0%	28.6%	18.4%	22.1%
Financial support	8.9%	9.1%	6.7%	7.8%	22.4%	11.7%
None	6.3%	7.8%	10.1%	27.3%	43.4%	3.9%
What type of information has been communicated to you or your school about the program?						

	GKA (India)	Nanhi Kali (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
None	1.3%	2.5%	10.1%	3.7%	31.3%	0.0%
Goals and objectives of the program	84.8%	67.1%	62.9%	76.3%	41.3%	88.3%
Roles and responsibilities of teachers	73.4%	43.0%	49.4%	67.5%	47.5%	67.5%
Roles and responsibilities of school director	44.3%	20.3%	34.8%	57.5%	7.5%	61.0%
Official curriculum and materials	55.7%		42.7%	48.7%	43.7%	31.2%
Training expectations	30.4%		29.2%	35.0%	28.8%	27.3%
Expectations about student outcomes	54.4%	32.9%	37.1%	46.2%	21.3%	55.8%
Policy changes around instruction.	5.1%	35.4%	11.2%	22.5%	10.0%	22.1%
Do you meet with parents/community/school governing body? (If so, how often?)						
No	0.0%	0.0%	1.1%	0.0%		2.6%
More than once per month	34.2%	51.9%	10.1%	25.0%		13.0%
Once per month	60.8%	44.3%	37.1%	21.3%		7.8%
A few times per year	5.1%	2.5%	50.6%	52.5%		72.7%
Yearly or less	0.0%	1.3%	1.1%	1.3%		3.9%
Have you received training as a part of this program? If so, who provided the training?						
No training received.	36.7%		39.3%	33.8%	78.8%	3.9%
Project Staff	53.2%					
Ministry of Education Staff	26.6%		29.2%	22.5%	8.7%	36.4%
Other	3.8%		2.2%	1.2%	0.0%	3.9%
Which of the following did you receive training on from the program?						

	GKA (India)	Nanhi Kali (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Early Grade Reading or Mathematics (same as teachers)	38.0%		55.6%	50.9%	58.8%	95.9%
Student Assessment	46.0%		51.9%	39.6%	70.6%	82.4%
Providing instructional support to teachers	60.0%		53.7%	60.4%	64.7%	35.1%
Management and Leadership Skills	66.0%		29.6%	50.9%	52.9%	28.4%
Engaging the community	40.0%		16.7%	43.4%	23.5%	40.5%
Other	0.0%		7.4%	3.8%	17.6%	0.0%
How would you rate the effectiveness of program trainings compared to other in-service trainings you've attended?						
More effective	94.0%		56.2%	87.5%	76.5%	60.5%
Less effective	2.0%		3.4%	3.8%	0.0%	14.0%
About the same	4.0%		23.6%	5.0%	11.8%	20.9%
Never received other in-service training	0.0%		16.9%	3.8%	11.8%	2.3%
Doesn't know	0.0%		0.0%	0.0%	0.0%	2.3%
What do you do differently as a head teacher as a result of this Program?						
Provide more instructional support to teachers.	68.4%		60.7%	66.2%	40.0%	68.8%
Lead (more) teacher meetings/discussions	74.7%		31.5%	40.0%	32.5%	58.4%
Monitor teachers' performance.	68.4%		60.7%	50.0%	33.8%	54.5%
Emphasize importance of reading/instruction with teachers	72.2%		37.1%	61.2%	26.2%	37.7%
Discipline teachers	16.5%		13.5%	22.5%	8.8%	39.0%

	GKA (India)	Nanhi Kali (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Other	1.3%		0.0%	1.2%	21.3%	7.8%
Nothing	0.0%		2.2%	10.0%	27.5%	6.5%
How do you know if a teacher in your school is not performing as expected?						
Regular interaction/discussion with teachers	63.3%	34.2%	33.7%	30.0%	70.0%	54.5%
Observing the teacher	84.8%	40.5%	74.2%	71.2%	70.0%	67.5%
Student results	81.0%	68.4%	56.2%	87.5%	65.0%	80.5%
Coaching data	10.1%	10.1%	13.5%	25.0%	11.3%	15.6%
Coaching feedback	12.7%	38.0%	10.1%	52.5%	28.8%	15.6%
Other	0.0%	0.0%	11.2%	5.0%	27.5%	1.3%
What action do you take if a teacher is not performing as expected?						
Provide additional support to the teacher yourself	70.9%	39.2%	82.0%	71.2%	71.2%	89.6%
Request additional coaching support for the teacher.	67.1%	35.4%	24.7%	71.2%	62.5%	26.0%
Observe the teacher (more frequently)	58.2%	48.1%	44.9%	61.2%	53.8%	41.6%
Provide a report to ministry official	15.2%	16.5%	5.6%	16.3%	15.0%	6.5%
Other	0.0%	2.5%	6.7%	3.7%	33.8%	15.6%
Is there a community of practice meeting for teachers held at your school or at the cluster/block level? [If yes, how often are these meetings held?]						
No	11.4%	2.5%	32.6%	17.5%	7.5%	3.9%
Once per week	12.7%	5.1%	1.1%	5.0%	12.5%	19.5%

	GKA (India)	Nanhi Kali (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
More than once per month	29.1%	35.4%	9.0%	26.2%	22.5%	6.5%
Once per month	45.6%	48.1%	15.7%	20.0%	8.8%	58.4%
Once per term	1.3%	7.6%	36.0%	28.8%	45.0%	10.4%
Less than once per term	0.0%	1.3%	5.6%	2.5%	3.7%	1.3%
Can you show me the minutesmeeting logagenda of the last meeting that occurred?						
Head teacher was able to show record	68.6%	85.7%	50.0%	97.0%	73.0%	59.5%
Head teacher was NOT able to show record	31.4%	14.3%	50.0%	3.0%	27.0%	40.5%
Did instruction continue during the Covid-19 schools closure period?						
Yes	92.4%	25.3%	98.9%	100.0%	87.5%	15.6%
No	7.6%	74.7%	1.1%	0.0%	12.5%	84.4%
If Yes, what platforms/channels were used to ensure continuity in instruction?						
Television	30.1%	20.0%	25.0%	23.8%	0.0%	0.0%
Radios	19.2%	5.0%	4.5%	0.0%	0.0%	11.1%
Teleconferencing platforms	95.9%	25.0%	44.3%	93.7%	34.3%	0.0%
Other	13.7%	0.0%	26.1%	7.5%	21.4%	83.3%
How did the government support the resumption of learning after the Covid-19 period?						
Additional Assessments	46.8%	20.3%	33.7%	26.2%	25.0%	5.2%
Modified Curriculum/Timetable	73.4%	53.2%	31.5%	43.8%	85.0%	16.9%
Additional time for remediation	69.6%	36.7%	56.2%	70.0%	21.3%	83.1%

	GKA (India)	Nanhi Kali (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Additional support staff	22.8%	25.3%	19.1%	22.5%	27.5%	0.0%
Additional materials	20.3%	20.3%	57.3%	28.8%	17.5%	7.8%
Other	1.3%					15.6%
Does this school have a school management committee, parent teacher association, or School Board?						
Yes	100.0%	97.5%	97.8%	96.2%	97.5%	100.0%
No	0.0%	2.5%	2.2%	3.7%	2.5%	0.0%
How often does the SSC, SMC, or School Board meet? (Please show minutes/attendance)						
Once a month or more	74.7%	76.6%	65.5%	26.0%	47.4%	7.8%
Once every two months	15.2%	19.5%	27.6%	15.6%	10.3%	16.9%
Once a quarter	7.6%	3.9%	1.1%	22.1%	39.7%	67.5%
Once a semester	2.5%	0.0%	4.6%	36.4%	2.6%	7.8%
Never	0.0%	0.0%	1.1%	0.0%	0.0%	0.0%
What aspects of the school's functioning does the SMC/PTA/School Board contribute to? Please describe.						
Infrastructure	72.2%	44.2%	60.9%	49.4%	61.5%	75.3%
School Administration/Budget	39.2%	31.2%	72.4%	37.7%	82.1%	68.8%
Teacher Management	36.7%	35.1%	33.3%	44.2%	57.7%	32.5%
Teacher Instruction	27.8%	31.2%	17.2%	58.4%	37.2%	5.2%
Teaching and Learning Materials	68.4%	37.7%	23.0%	50.6%	46.2%	54.5%
Hygiene and Sanitation	60.8%	63.6%	50.6%	63.6%	39.7%	37.7%
School Feeding Program	63.3%	13.0%	60.9%	33.8%	48.7%	40.3%

	GKA (India)	Nanhi Kali (India)	ESMATE (El Salvador)	RAMP (Jordan)	R-Maths (South Africa)	TAFITA (Madagascar)
Student Performance/Exams	55.7%	35.1%	28.7%	50.6%	35.9%	20.8%
Student Enrollment and attendance	57.0%	37.7%	13.8%	49.4%	33.3%	29.9%
Supporting students who struggle	45.6%	22.1%	27.6%	68.8%	46.2%	40.3%
Outreach to parents/community	32.9%	37.7%	34.5%	59.7%	46.2%	54.5%
Other	1.3%	0.0%	8.0%	2.6%	20.5%	7.8%
None	2.5%	0.0%	0.0%	1.3%	1.3%	0.0%
Does your school have any of the following plans?						
School Strategic Plan/School Improvement Plan	44.3%	74.7%	75.3%	93.7%	100.0%	
School Annual Operational Plan	89.9%	74.7%	82.0%	62.5%	67.5%	
Other	1.3%	1.3%	31.5%	7.5%	11.3%	
None	0.0%	0.0%	3.4%	0.0%	0.0%	

ANNEX B. LEARNING AT SCALE PROGRAM INTERVIEW QUESTIONS

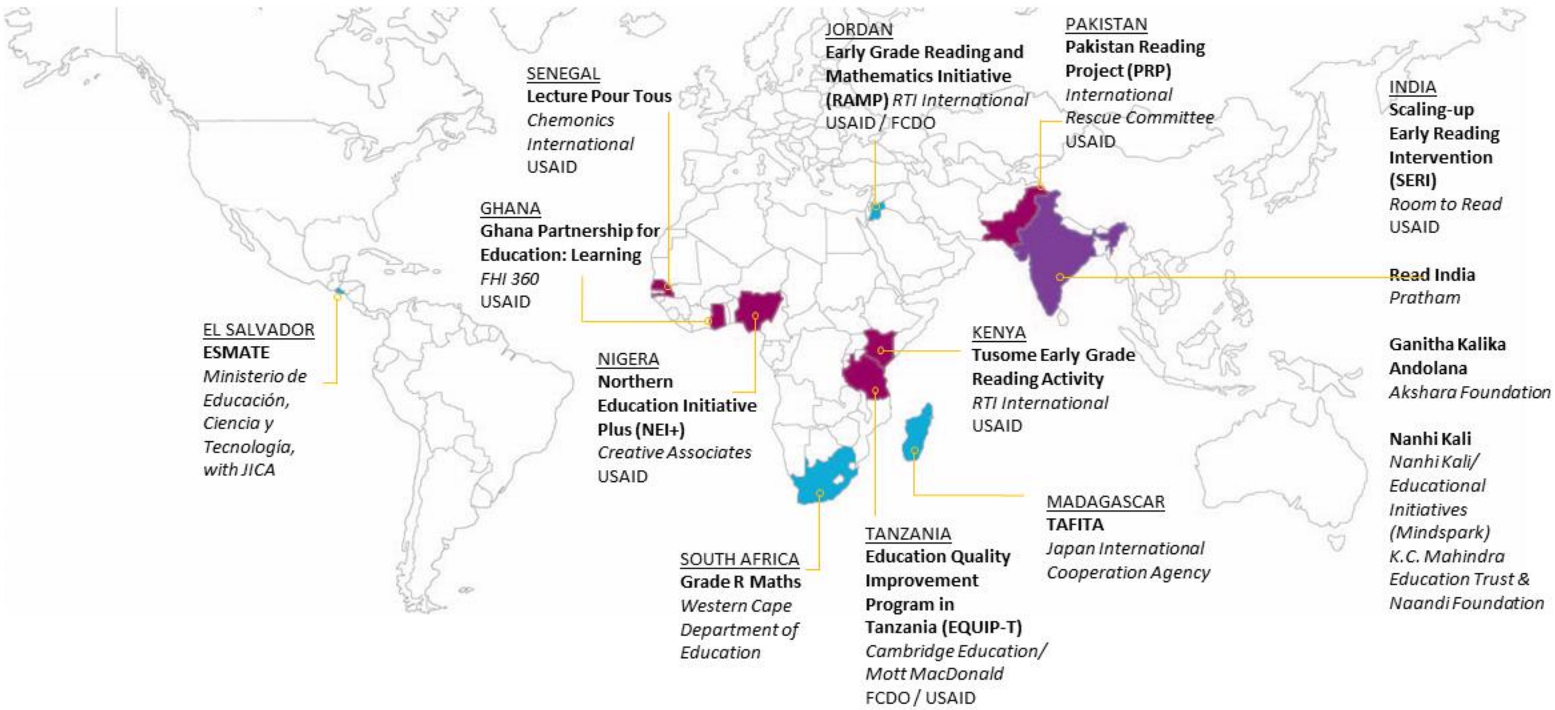
1. **Please describe your intervention**
    a. What level of schooling (e.g., primary; secondary)?
    b. Does the program involve a numeracy or literacy component? Briefly describe.
    c. What is the timeline for the program? When will it end?
    d. What is the instructional strategy in the classroom that makes this program successful?
    e. What makes this program unique?
2. **What learning outcome evaluation data exists?**
    a. What was the scope of the evaluation (e.g., how many schools were in the evaluation? how many schools is the outcome data externally valid to?)?
    b. What is the impact of the program (as defined by the program)?
    c. What is the impact of the program (as defined by effect sizes in standard deviations)?
    d. When was the impact evaluation done?
    e. What is the design of the impact study?
3. **What is the scope of the program?**
    a. How many schools are you implementing in currently? How many schools will you be implementing by the end of the program?
    b. How many entire districts or clusters is the program implementing in?
4. **What is the role of the government in program implementation?**
    a. If not fully government-led, is the program integrated into the system?
    b. What evidence of integration can you provide?
5. **If selected, can we have access to the raw impact data?**
6. **If selected, can we have access to cost data?**
7. **What else do we need to know to understand your program and its eligibility for learning at scale?**

ANNEX C. FINAL PROGRAM SELECTION

Program	Country *(Duration)*	Implementer	Funder	Effectiveness	Current Scale	Evidence of Impact at Scale?	Key Takeaways
ESMATE	El Salvador *Ongoing:* *2015 pilot* *2019 scale*	Ministerio de Educación de El Salvador, El Salvador	Pilot: JICA Scale: Ministerio de Educación de El Salvador	0.48 SD impact on math scores for 2nd grade students from impact evaluation (small-scale); 0.18 SD increase in math scores for lower secondary	As of 2019, all 4,666 primary schools, 2,726 Junior High Schools and 705 Senior High Schools in El Salvador	No, but because the program continues to implement in pilot regions while scaling, there is the opportunity to compare (done under N@S)	ESMATE 2 is a national take-up of the program that is being implemented in all schools. There were minor shifts due to COVID but the overall package remains the same.
Ganitha Kalika Andolana	India *Ongoing* *2011 - present*	Akshara Foundation	Akshara FoundationGov't of Karnataka	Internal longitudinal evaluation (from prior phase) found 0.27 to 0.43 SD impact on numeracy; External randomized evaluation found 0.14 effect size for girls but no average impact.	48,000 in Karnataka, 54,000 in Odisha; 2,000 in Andhra Pradesh	Yes – for girls only. While the external evaluation detected only significant impacts for girls, implementation was disrupted by flooding, and children received only 66 hours of the intervention	Pilot work began in 2011 (with evaluation in 2015). State governments adopted the program and Akshara now provides some technical support. Mixed evidence from recent external evaluation. Implementation greatly impacted by COVID.
Projet d'Appui à la Gestion Participative et Décentralisée de l'écoles (TAFITA)	Madagascar *Phase 1: 2016-2020* *Phase 2:* *2020-2024*	Japan International Cooperation Agency (JICA)	JICA	0.47 standard deviation increase in grade 3 math scores based on RCT (endline in 2019)	Full program in 2 regions (2,725 schools; 288,896 grade 2-5 students) The reading and SMC work has expanded to 9 additional regions (10,680 schools)	No, but because the program continues to implement in pilot regions while scaling, there is the opportunity to compare.	Program combines PMAQ and TaRL, using mobilization of local resources for educational development. Hybrid of government curriculum and remedial lessons.
Grade R Math	South Africa *Ongoing* *2017 - present*	Western Cape Education Department (WCED)	Pilot: WCED and University of Cape Town Schools' Development Unit Scale: WCED	0.21 SD effect size for rural district (no impact overall on urban). 0.17 to 0.24 SD effect size combined, accounting for covariates (quasi-experimental design)	All schools in Western Cape (8 districts; ~1000 schools)	Yes	Initially this pre-primary maths program ended in 2018 but has been integrated into government system. A similar model is also being taken to scale in Gauteng.
Nanhi Kali (using Mindspark)	India *Nanhi Kali began in 1996, introduced Mindspark in 2020*	Naandi Foundation, Education Initiatives	K.C. Mahindra Education Trust (funded by Mahindra & Mahindra, with other corporations and individuals)	Evaluations of Mindspark in both after-school centers and in-school models found large impact on numeracy scores from a small pilot (1.7 SD effect size).	Nanhi Kali (using Mindspark) is now operating in 6,100 centers across 8 states; reaching 160,000 girls	No. Pilot evidence for Mindspark only (200 schools)	Nanhi Kali uses the Mindspark app as its core instructional approach in an after-school program for girls. Use of Mindspark is facilitated by trained community members.

Program	Country *(Duration)*	Implementer	Funder	Effectiveness	Current Scale	Evidence of Impact at Scale?	Key Takeaways
RAMP	Jordan *2015 - 2022*	RTI International	U.S. Agency for International Development (USAID)/ UK Foreign, Commonwealth and Development Office (FCDO)	10 percentage point (50%) increase in students who can do grade-level mathematics with understanding. Internal endline, no comparison.	All public schools in Jordan (2,230 schools)	Yes	National scale and strong endline improvements in numeracy.

Learning at Scale: Studying the Most Successful Literacy and Numeracy Programs at Scale



ANNEX D. FULL DATA ON PROGRAM ELEMENTS AND KEY ELEMENTS FOR PROGRAM IMPLEMENTATION

		ESMATE		R Maths		TAFITA		GKA		Nanhi Kali		RAMP	
Category	**Items**	**Component of Program Implementation**	**Key to Program Implementation**	**Component of Program Implementation**	**Key to Program Implementation**	**Component of Program Implementation**	**Key to Program Implementation**	**Component of Program Implementation**	**Key to Program Implementation**	**Component of Program Implementation**	**Key to Program Implementation**	**Component of Program Implementation**	**Key to Program Implementation**
Materials	Program materials aligned to government curriculum												
Materials	Structured teacher's guides (scripted lessons)												
Materials	Teaching aids for Students (e.g., counters, number cards, place-value materials, etc.)												
Materials	Consumable student books (workbooks)												
Materials	Local language materials												
Materials	Lesson Plans (little to no scripting or structure)												
Materials	Materials developed with Gov't												
Materials	Students books (textbooks) for all students (1:1)												
Materials	Textbook taken home												
Materials	Online or soft-copy materials for teachers												
Materials	Student books (textbooks)												

Materials	Teaching aids for Teachers (e.g., number charts, posters, etc.)												
Pedagogy	Continuous/formative assessment												
Pedagogy	Instruction targeted to student level / differentiated instruction												
Pedagogy	Focus on developing conceptual understanding												
Pedagogy	Pairwork or Groupwork												
Pedagogy	Using concrete materials/resources (manipulatives)												
Pedagogy	Teacher model/explanation followed by student practice												
Pedagogy	Focus on developing procedural knowledge												
Pedagogy	Implemented together with literacy program												
Pedagogy	Increased instructional time in lessons												
Pedagogy	Linking informal (out-of-school) and formal mathematics												
Pedagogy	Supporting student discussion/explanation of math concepts												
Pedagogy	Using multiple models/representations												
Support	Coaches are government staff												
Support	Caregiver/community involvement in school management (e.g. PTA/SMC meetings)												

Support	Coaches have structured tools												
Support	Coaches meet in groups / with supervisors												
Support	Communities of practice meetings (across schools)												
Support	Program oversight/support for coaches during school visits												
Support	Caregiver involvement in student learning at home (e.g., materials, support, homework)												
Support	Coaches are provided with program/teacher materials												
Support	Coaches use tablets or other devices (smart phones etc)												
Support	Internal to school coaching/mentoring												
Support	Virtual communities of practice (whatsapp, SMS, not face to face)												
Support	Coaches are program hired												
Support	Coaches reimbursed by program (e.g., transport support)												
Support	External to school coaching												
Support	School-based communities of practice meetings (within school)												
Support	Virtual coaching (whatsapp, SMS, not face to face)												
System Support	Government staff responsible for conducting monitoring												

System Support	Program invested in capacity building - decentralized level
System Support	Government uses monitoring data to make decisions about implementation
System Support	Program developed and used dashboard for result/data sharing
System Support	Gov't responsible for monitoring frequency of coach visits
System Support	Program designed to align with existing education plans
System Support	Program has regional staff
System Support	Program uses monitoring data to make decisions about implementation
System Support	Gov't responsible for distribution of materials
System Support	Program gave community/local grants
System Support	Program invested in capacity building - central level
System Support	Program mapping/scoping exercise prior to intervention
System Support	Program mobilized additional local resources to support schools
System Support	Program responsible for distribution of materials
System Support	Program responsible for monitoring frequency of coach visits

System Support	Program sent funds to Gov't
System Support	Program shares achievement Data with gov't decision makers
System Support	Program staff embedded in Gov't offices
Training	Initial face to face training
Training	Refresher face to face training
Training	Teacher training (lowest level of cascade) done by government officers
Training	Structured training manuals
Training	Teacher training emphasizes modeling/practice
Training	Training for Head Teachers
Training	Residential teacher training
Training	School-based training
Training	ToT done by program staff
Training	Non-residential teacher training
Training	Pre-service training component
Training	ToT done by government staff
Training	Virtual teacher training (whatsapp, SMS, not face to face)

ANNEX E. CLASSROOM OBSERVATION INSTRUMENTS DEVELOPED FOR NUMERACY AT SCALE

E.1 Quantitative Observation Tool

Classroom Observation – Early Grade Mathematics (REVISED 14 Dec 2022)

Date: _______________ Assessor Code: __________________ Official Language: _____________________

School: _______________________________________Grade(s): ____________________

Number of girls present _______
Number of boys present _______

Number of learners with textbook _________

Number of learners with a workbook ___________

Start time: _________________________ End Time: __________________________

Total Number of Minutes: __________________________

The observation form should be completed in class during a mathematics lesson. If the teacher indicates that there is not a separate mathematics lesson, ask to observe a lesson that focuses on mathematics.

When arriving to class, find a seat at the back of the class. Try not to interrupt or disturb the class.

1. Lesson Structure		
1a	What parts were present in the lesson? (Mark all that apply)	☐ 1. whole class- teacher talking/choral recitation/singing/some learners solving problems in front of class/ teacher is leading as learners working on an activity or textbook/workbook problems ☐ 2. independent/group work: learners working on their own ☐ 3. non-instructional time (passing out materials, classroom management)

		☐ 4. Other ___________________
1b	Put the above parts you observed in order from beginning to end of the lesson. Some parts may occur more than once. Do not include 3. Non-instructional time	(enter order of approaches

2. Whole Class (Teacher model, explanation, discussion, a few learners working at board, teacher leading while learners working while at desk)		
2a	By the end of the lesson, how many minutes was the whole class section?	______________Enter number of minutes
2b	How often did the teacher ask learners to repeat after her? [Mark one]	☐ Never ☐ 1 – 3 times ☐ 4 – 6 times ☐ 7 – 10 times ☐ 11 – 15 times ☐ 16 – 20 times ☐ More than 20 times
2c	**Teacher Questioning:** How many times did the teacher ask learners a question that does not have one correct answer?	How many times did the teacher ask a question that did not have one correct answer? ________
2d	**Patterns in Learner Engagement**: **Teacher questions** 1. Approximately how many times did an individual student respond to a question? [If never, move to 2d-3]	☐ Never ☐ 1 – 3 times ☐ 4 – 6 times ☐ 7 – 10 times ☐ 11 – 15 times ☐ 16 – 20 times ☐ More than 20 times

	2. Who did the teacher call on to answer questions? [Mark one]	☐ a few of the **same** learners ☐ a variety of learners ☐ only one learner
	3. How often did the teacher ask students to answer a question in partners (think pair share) or small groups?	☐ Never ☐ 1 – 3 times ☐ 4 – 6 times ☐ 7 – 10 times ☐ 11 – 15 times ☐ 16 – 20 times ☐ More than 20 times
	4. How often was there a choral response to a question? [Mark one]	☐ Never ☐ 1 – 3 times ☐ 4 – 6 times ☐ 7 – 10 times ☐ 11 – 15 times ☐ 16 – 20 times ☐ More than 20 times
	5. How often did the teacher answer her own question? [Mark one]	☐ Never ☐ 1 – 3 times ☐ 4 – 6 times ☐ 7 – 10 times ☐ 11 – 15 times ☐ 16 – 20 times ☐ More than 20 times
2e	**Patterns in Learner Engagement: Learner Incorrect Response** 1. Did any learners answer a teacher question with an incorrect or no response? [Mark one; If no, skip to 2e]	☐ yes, more than 3 times ☐ yes, 1-2 times ☐ no
	2. How did the teacher respond to the incorrect answer or no response? ((note how teacher responded) [Mark all that apply]	☐ Teacher asked another learner to answer ☐ Teacher gave correct response ☐ Teacher discussed why answer was incorrect ☐ Teacher asked student/students to solve the same problem again without any additional help ☐ Teacher helps student/students solve the problem through questions, model, clarifying questions, etc.

		☐ Teacher ignores incorrect response and moves on to a new question ☐ Other ______________________
2f	**Patterns in Learner Engagement: Learner Content Questions** 1. How many times did learners ask questions about the content of the lesson? [If 0, skip to 2g]	Note number of times______
	2. If yes, who answered? [Mark all that apply]	☐ The teacher ☐ Other learners ☐ No one- question is not answered
2g.	**Teacher Model** 1. Did the teacher model/explain how to solve a problem, a concept, an activity? [If no, skip to 2h].	☐ yes ☐ no
	2. Did the teacher ever make a math error?	☐ yes ☐ no
	3. Did the teacher ever connect the math concept to real-life examples or the lives of learners (e.g. linked to everyday examples like shopping in the market, or playing sports)?	☐ yes ☐ no
	4. Did the teacher intentionally ever introduce an incorrect solution?	☐ yes ☐ no
	5. Did the teacher introduce or use a learners' incorrect response to discuss why it is incorrect (e.g., any discussion of WHY a solution is not correct)?	☐ yes ☐ no
2h.	**Manipulatives Use (concrete materials)**	☐ counters ☐ Shapes

	1. What materials, if any, were used during the teacher model or discussion? [Mark all that apply] [If none, move to 3 (independent work)]	☐ number cards ☐ place value materials ☐ Number chart 1-100	☐ Dice ☐ Varied objects for sorting and making patterns ☐ Real or concrete objects for fractions ☐ Other: ________________ ☐ none
	2. Who used the materials? Mark all that apply.	☐ teacher ☐ only a subset of learners ☐ all learners	

3. Independent work (learners or groups of learners working independently without the teacher leading)		
3a.	**Time** 1. By the end of the lesson, about how many minutes consisted of independent work [If 0, skip to 4 (other)]	_____________Enter number of minutes
3b.	**Monitoring** 1. Did the teacher monitor learner work while they were working? [If no, skip to 3c]	☐ yes ☐ no
	2. Did the teacher spend more time with any one learner/group? [If no, skip to 3c]	☐ yes ☐ no
	3. If yes, why was the teacher spending more time with the learner? [Mark all that apply]	☐ to help the learner with management issues (finding the page, getting pencil or materials, personal issues, etc.) ☐ to help the learner solve the math problem

		☐ can't tell ☐ other ________________	
3c	**Learning Tasks** 1. What types of tasks did the learners work on during independent work time? [Mark all that apply]	☐ Solving problems written on the board/ from the textbook/workbook/worksheet ☐ Active learning activity (e.g., games and puzzles, measuring objects around class, using manipulatives, drawing, cutting out shapes) ☐ Only copying from board or textbook ☐ Other ________________	
	2. Did the teachers give all learners the same activity or problems to work on?	☐ yes ☐ no	
	3. Look at 3-5 books (or observe students doing an activity) from students seated around you. Are the students solving the problems/doing the activity correctly?	☐ yes, most of the students are solving problems/doing the activity correctly ☐ no, most of the students are NOT solving problems /doing the activity correctly	
3d	**Manipulatives Use (concrete materials)** 1. What materials, if any, were used during the independent work time? [If none, skip to 3e]	☐ counters ☐ number cards ☐ place value materials ☐ Number chart 1-100	☐ Shapes ☐ Dice ☐ Varied objects for sorting and making patterns ☐ Real or concrete objects for fractions ☐ Other: ________________ ☐ none
	2. Who used the materials during independent work time?	☐ teacher ☐ subset of learners ☐ all learners	

3e	1. How many learners are engaged the whole time?	☐ yes, all learners are engaged the whole time ☐ yes, more than half of learners were engaged the whole time ☐ yes, less than half of learners were engaged the whole time ☐ No learners were engaged
	2. If some learners finished early, what did the teacher do?	☐ Nothing ☐ He gave them more of the same problems to solve ☐ He gave them a different math activity or problem ☐ He gives them other non-math related work ☐ Other ________________ ☐ No learners finished early

Other		
4a Physical space	1. What math-related materials were visible in the classroom? [Mark all that apply].	☐ posters ☐ concrete materials ☐ other ____________________________ ☐ none
	2. What was the sitting arrangement? [Mark all that apply].	☐ learners at individual desks ☐ learners on the floor ☐ learners at desks with multiple learners per desk/table/chair facing front of room ☐ small groups (learners facing each other) ☐ other ____________________________

4b Communicating with learners	1. What is the topic(s) of the lesson?	☐ Numbers and counting Note specific sub-topic: ______________ ☐ Operations Note specific sub-topic: ______________ ☐ Geometry Note specific sub-topic: ______________ ☐ Measurement Note specific sub-topic: ______________ ☐ Statistics/data analysis Note specific sub-topic: ______________ ☐ other Note specific sub-topic: ______________
	2. Did the teacher refer to a teacher's guide during the lesson?	☐ yes ☐ no
	3. Did the learners use a textbook during the lesson?	☐ yes ☐ no
	4. Did learners use a workbook during the lesson? [If no, skip to 4c].	☐ yes ☐ no
	5. If yes, did learners write in the learner's workbook?	☐ yes ☐ no
4c Using assessment for instruction	Did the teacher do any of the following assessment activities? [Mark all that apply]	☐ Checking learner work ☐ Collecting learner work for the purposes of reviewing learner work ☐ Giving an assignment for the purposes of grading ☐ Other __________________

E.2 Qualitative Observation Tool

Qualitative Classroom Observation – Early Grade Mathematics

For each pattern, note **Evidence** and corresponding **Impressions**

Pattern: Clarity and effectiveness of math/model/explanation/use of representation	• Did the teacher model clearly? • What types of representations did teacher use? • Were manipulatives used? How and by whom? • Did the teacher explain why an incorrect response is incorrect? • What, if any, math errors were there? • Was the teacher confident in delivery of the model? • Were there opportunities for students to develop their own knowledge before teachers showed them how to solve a problem?

Pattern: Appropriateness/difficulty/ sequencing of content	• Were content progressions appropriate? • Was the content of the lesson developmentally appropriate? • Did the students seem to understand the content? • What gaps were there in the progressions?

Pattern: Task types during whole class and independent work	• Were the tasks too easy, too hard, or right at students' level? • How many tasks were assigned during independent or group work? Was it enough for the amount of time given? • How successful were students at the tasks? • Were the tasks aligned to the lesson teachers were teaching?

Pattern: Student Engagement	• What kind of questions did the teacher ask? • Who answered questions? Was it students who raised hand, volunteers, variety of children or only a few? • What did teacher do with incorrect responses? • What types of choral responses were there? When did they happen? • How did teachers engage students during whole class? • Did students do group or individual work? • How did they work together? • What did teachers do to help students work together?

Pattern: Other	• How did the teacher deal with students who did not understand the content? • Did the teacher struggle with management? At what times? • Anything else?