October 2023

Learning at Scale: Final Report

Photo credit: USAID Lecture Pour Tous, Senegal

RTI INTERNATIONAL

RTI Project Number
0216840.000.000

Learning at Scale: Final Report

October 2023

Prepared for

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

Prepared by

Jonathan Stern, Matthew Jukes, Joseph DeStefano, Jessica Mejia, Peggy Dubeck, Bidemi Carrol, Rachel Jordan, Cosmus Gatuyu, Tabitha Nduku, Christine Van Keuren, Maitri Punjabi, Fatima Tufail
RTI International
3040 E. Cornwallis Road
Research Triangle Park, NC 27709
USA

CONTENTS

LIST OF FIGURES

LIST OF TABLES

viii

LIST OF ABBREVIATIONS

3Rs	reading, writing, arithmetic (Tanzania program)
5 T's	tongue, text, teaching, time, test
AAM	assessor accuracy measure
AJK	Azad Jammu and Kashmir Province (Pakistan)
ASER	Annual Status of Education Report
BRC	block resource coordinator (India)
CAMaL	Combined Activities for Maximized Learning
CBC	competency-based curriculum (Kenya)
CGD	Center for Global Development
CL	Curriculum Leader (Ghana)
CLASS	Classroom Assessment Scoring System
CLASSIC	Classroom Language Arts Systematic Sampling and Instructional Coding
COL	Community of Learning (Tanzania)
COVID-19	coronavirus disease 2019
CPD	continuous professional development
CRC	cluster resource coordinator (India)
CS	Circuit Supervisor
CSO	Curriculum Support Officer
cwpm	correct words per minute
DFID	United Kingdom's Department for International Development (merged into FCDO in 2020)
DOE	Department of Education
DTST	District Teacher Support Team (Ghana)
EDC	Education Development Center, Inc.
EGR	early grade reading
EGRA	Early Grade Reading Assessment
ELLCO	Early Language and Literacy Classroom Observation
EQUIP-T	Education Quality Improvement Program in Tanzania
FATA	Federally Administered Tribal Area (now called Newly Merged Districts; Pakistan)
FCDO	United Kingdom's Foreign, Commonwealth and Development Office
GB	Gilgit-Baltistan (region, Pakistan)
GoT	Government of Tanzania
GPE	Global Partnership for Education
GPS	global positioning system
HQPD	[Observation Checklist for] High-Quality Professional Development
HT	head teacher
ICT	Islamabad Capital Territory (Pakistan)

INSET	in-service education and training
IRC	International Rescue Committee
IRR	interrater reliability
ITTSI	In-service Teacher Training Survey Instrument
K–3	kindergarten–grade 3
KG2	second level of kindergarten
KICD	Kenya Institute of Curriculum Development
KIE	Kenya Institute of Education
KNEC	Kenya National Examinations Council
KP	Khyber Pakhtunkhwa Province (Pakistan)
L1, L2	first language, second language
LANES	Literacy and Numeracy Education Support (GPE program, Tanzania)
LAYS	learning-adjusted years of schooling
LEMA	Local Education Monitoring Approach
LF	Literacy Facilitator (SERI, India)
LGA	Local Government Authority (Tanzania; equivalent of district)
LGEA	local government education authority (Nigeria)
LMICs	low- and middle-income countries
LOI	language of instruction
LPT	Lecture Pour Tous
LS	Literacy Specialist (Ghana)
MGML	multigrade, multilevel
MOE	Ministry of Education
MOU	memorandum of understanding
MSI	Management Systems International
MT	mother tongue
n/a	not applicable
NEI+	Northern Education Initiative Plus (Nigeria)
NGO	nongovernmental organization
NICHD	U.S. National Institute of Child Health and Human Development
NMD	Newly Merged Districts (formerly known as Federally Administered Tribal Areas, or FATA, Pakistan)
NSC	National Steering Committee (Kenya)
OL	oral language
OPERA	*Observation des Pratiques Enseignantes dans leur Rapport avec les Apprentissages des élèves* (Observation of Teaching Practices in Relation to Student Learning)
OPM	Oxford Policy Management (external evaluator of EQUIP-T)
ORF	oral reading fluency
P1, P2	primary grades 1, 2 (Ghana)
PA	phonological awareness

PKR	Pakistani rupee
PLATO	Protocol for Language Arts Teaching Observations
PO-RALG	President's Office for Regional Administration and Local Government (Tanzania)
PRIEDE	Primary Education Development (numeracy project funded by the World Bank and the Global Partnership for Education)
PRIMR	Primary Mathematics and Reading Initiative (Kenya)
PRP	Pakistan Reading Project
QASO	Quality Assurance and Standards Officer (Kenya)
RCT	randomized controlled trial
RTI	RTI International (registered trademark and trade name of Research Triangle Institute)
SAGA	Semi-Autonomous Government Agency (Kenya)
SBI	school-based in-service education and training (aspect of Ghana Learning)
SCOPE-L	Standards-based Classroom Observation Protocol for Educators in Literacy
SD	standard deviation
SERI	Scaling-up Early Reading Intervention
SIM	subscriber identification module
SMS	short message service
SSA	School Support Associate (Pakistan)
SSO	School Support Officer (Nigeria)
TaRL	Teaching at the Right Level
TIE	Tanzania Institute of Education
TIG	teacher inquiry group (Pakistan)
TIPPS	Teacher Instructional Practices and Processes System
TLMs	teaching and learning materials
ToT	training of trainers
TPD	teacher professional development
TPDS	Targeted Professional Development Scale
TSC	Teachers' Service Commission (Autonomous Government Agency affiliated with Ministry of Education, Kenya)
TSh	Tanzanian shilling
TTI	teacher training institute (Pakistan)
TWG	Technical Working Group
UNICEF	United Nations Children's Fund
USAID	United States Agency for International Development
WEC	Ward Education Coordinator (Tanzania)
WEO	Ward Education Officer (Tanzania)

CONFLICT OF INTEREST DISCLOSURE

The Learning at Scale research study is led by RTI International and is part of the Center for Global Development (CGD) education research consortium, funded by the Bill and Melinda Gates Foundation. RTI International is also the lead implementer of one of the programs (Kenya Tusome) selected for inclusion in Learning at Scale.

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

RTI adopted a transparent process for developing research questions, methodologies and selection partner programs. All proposed questions and methodologies were reviewed by an independent advisory board. RTI presented two different methods of using selection criteria and presented the results to the advisory board with programs anonymized. Accordingly, final selection of the Learning at Scale programs was determined by this independent advisory committee.

RTI's first principal investigator for Learning at Scale was the former Chief of Party for the Tusome program in Kenya (which was selected for inclusion in this study). He was precluded from participating in the data collection in Kenya, as well as in initial data analysis and writing of results. His input into the original interim report was to make clarifications in the final write-up, similar to that of partners from all included Learning at Scale programs. He did not contribute in any way to the report update or final report.

RTI also took critical 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 is not accessible to RTI staff outside 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 leader staff and the Learning at Scale team, which clearly lay out what personnel, documents and data cannot be used for business development and proposal 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

Learning outcomes are low and instruction has significant room for improvement in many low- and middle-income countries (LMICs). Although the number of successful programs is growing, relatively few have demonstrated impact at large scale. The Learning at Scale study was designed to identify existing programs with demonstrated impact on basic skills at scale and to conduct in-depth investigations of these programs to determine what makes them successful. This research is being led by RTI International and is part of the Center for Global Development (CGD) education research consortium, funded by the Bill and Melinda Gates Foundation.

The Learning at Scale study focuses on two areas: (1) identifying instructional strategies that are essential for improving learning outcomes at scale in LMICs and (2) learning about the characteristics of the systems within which successful, scaled-up programs operate. Our research is designed to achieve these two goals by studying eight of the most effective large-scale education programs in LMICs. After an extensive search, the eight programs listed in **Table ES-1** were selected for inclusion in the study.

Table ES-1.   Selected programs for inclusion in Learning at Scale

Program	Country	Lead implementer
Scaling-up Early Reading Intervention (SERI)	India	Room to Read
Education Quality Improvement Program in Tanzania (EQUIP-T)	Tanzania	Cambridge Education/ Mott MacDonald
Partnership for Education: Learning (Ghana Learning)	Ghana	FHI 360
Tusome Early Grade Reading Activity (Tusome)	Kenya	RTI International
Pakistan Reading Project (PRP)	Pakistan	International Rescue Committee (IRC)
Read India	India	Pratham
Lecture Pour Tous	Senegal	Chemonics International
Northern Education Initiative Plus (NEI+)	Nigeria	Creative Associates

Given that complete data could not be collected from all eight programs as scheduled because of COVID-19, this report presents findings and implications based on initial discussions with all eight programs, systems interviews with all eight programs, and from school level data collected for five programs: (1) EQUIP-T, (2) Tusome, (3) SERI, (4) Lecture Pour Tous, and (5) NEI Plus.

1.2 Research Questions

The aim of the Learning at Scale Project 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 consisted of causal relationships between 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 hypotheses using quantitative and qualitative data that describe the nature of the causal relationship. In addition, we used a quasi-experimental design to identify impacts of the program on teacher behavior and classroom activities. Our theory of change can be found in the Research Design section (Section 3) of the report.

1.3 Findings

1.3.1 Program selection

Although the selection of programs included in this study resulted from an exhaustive search, two characteristics were unintentionally common across nearly all selected programs. First, all the programs were primarily managed by implementing partners (i.e., none of the programs could be characterized as government-only programs). This finding likely stemmed from the fact that government-only programs are seldom rigorously evaluated, which was an important part of our selection criteria. We remain interested in examining programs that are run, managed, and implemented by governments and will reexamine this set of options during later Learning at Scale activities.

The second primary characteristic of the selected programs was their funding source. Six of the programs were funded at least in part by the United States Agency for International Development (USAID); one was funded by the United Kingdom's Department for International Development (DFID),[1] and one had multiple donors. This result was found despite the careful examination of programs sponsored by the World Bank, the Global Partnership for Education (GPE), United Nations Children's Fund (UNICEF), and several bilateral donors. Programs funded by these donors were typically not included in the Learning at Scale study because of the smaller scale of their interventions, because of the

---

[1] Note that DFID ceased to exist when its functions were merged into a new Foreign, Commonwealth and Development Office (FCDO) in mid-2020.

lack of rigorous impact evaluation data, or because the programs did not show substantial impacts on learning.

1.3.2 High-level analyses

We were interested in understanding what program design elements were included in each of the eight programs and which of these elements the programs deemed as key to their success. These data were collected from program documents, program visits, and interviews with program teams. We identified 10 elements that were determined to be key for four or more programs and suggest that future interventions consider these 10 elements as essential for program impact, as follows:

1. Program's teacher training focused on modeling and practicing new skills (seven programs).
2. Program included structured teachers' guides (six programs).
3. Coaches were provided structured tools to support teachers (six programs).
4. Program used face-to-face training methods for their initial trainings (six programs).
5. Program used direct-instruction pedagogical methods (five programs).
6. Student books were available at a 1:1 ratio for all students (four programs).
7. Program utilized a phonics-based instructional methodology (four programs).
8. Program increased the amount of instructional time in reading lessons (four programs).
9. Program provided capacity building at a decentralized level (four programs).
10. Program was designed to align with existing government education plans (four programs).

Although the characteristics of the programs differed, we found similarities in a few areas. Six of the programs could be characterized as structured pedagogy programs. The programs in India (SERI), Pakistan (PRP), Kenya (Tusome), Ghana (Ghana Learning), Nigeria (NEI+), and Senegal (Lecture Pour Tous) were all substantially similar. They typically had structured learning materials for students matched with teachers' guides for teachers, focused training on new instructional methods for teachers, and teacher support systems that included coaching. The Tanzania (EQUIP-T) and India (Read India) interventions were quite different and might suggest other avenues for successful program design that do not fit into the structured pedagogy category. Read India, for example, used a teaching at the right level (TaRL) program design (Global Education Evidence Advisory Panel, 2020).

1.3.3 Quantitative findings from instruction data

Our analysis of classroom observations and teacher, head teacher, coach, and facilitator interviews is based on data from five of the eight programs. However, some key patterns did emerge, related to how teachers use instructional time and what students spend time

doing during reading lessons. In addition, there are clear indications of what teachers see as valuable in their training, coaching, and support.

In four of the five programs, teachers spent more class time focused explicitly on reading instruction than on any other instructional focus, with a quarter of class time spent on phonics-related topics (e.g., letters, letter sounds, word parts). Students were observed to be actively reading during 20% to 59% of lesson time. Teachers were observed to still spend most (between 66% and 84%) of a lesson teaching to the whole class, but students were noted as being on task at a very high rate (with most students in the classroom on task more than 80% of the time). Teachers also appeared to be more responsive to students, and they self-reported that their instruction is more "student-centered."

While programs varied in the amount of class time dedicated to different instructional aims, materials used, and student activities, there was significant overlap in the instructional approaches that stakeholders saw as key for the success of these programs. For example, in all five programs, over two-thirds of interviewed teachers claimed that a greater focus on letters, sounds, and blending, as well as a new methodology and/or instructional approach were the instruction-related factors that had the greatest impact on student performance (**Table ES-2**). It is important to note that given the design of these programs, "new methodologies and instructional approaches" may also be interpreted to include a focus on letters, sounds, and blending as well as a gradual release model or structured/semi-structured lesson plans.

Table ES-2. Instruction-related factors for improved student performance, as reported by teachers, by program

What part of your instruction has had the biggest impact on student learning? (Select only one)	Tusome	EQUIP-T	SERI	Lecture Pour Tous	NEI+
More focus on letters, sounds, and/or blending	50.8%	25.4%	57.6%	63.3%	55.9%
More student centered and/or less lecture	13.6%	6.8%	3.4%	1.7%	10.2%
More pair and/or group work	6.8%	10.2%	10.2%	5.0%	6.8%
New methodology and/or instructional approach	27.1%	39.0%	18.6%	20.0%	18.6%
Involves more materials and/or activities	1.7%	11.9%	10.2%	3.3%	8.5%
Other	0	5.1%	0	3.3%	0
Program did not impact student learning	0	1.7%	0	3.3%	0

The area with the strongest convergence across programs was teacher views on training. Teachers across all five programs reported that the training sessions used more discussion (ranging from 56.9% to 88.9%) and small-group practice (ranging from 53.4% to 96.2%) than previous teacher training sessions they had attended (**Figure ES-1**). Approximately

three-quarters of Tusome and SERI teachers and close to two-thirds of Lecture Pour Tous teachers also reported that the program training sessions included more modeling than previous training sessions. As the findings from the Quantitative Analyses (Section 5.5.2) further illustrate, these successful programs covered more focused content, through multiple modes of presentation, in a more organized way than teachers had previously experienced.

Figure ES-1. Percentages of teachers reporting program training methods used more frequently than in previous training sessions
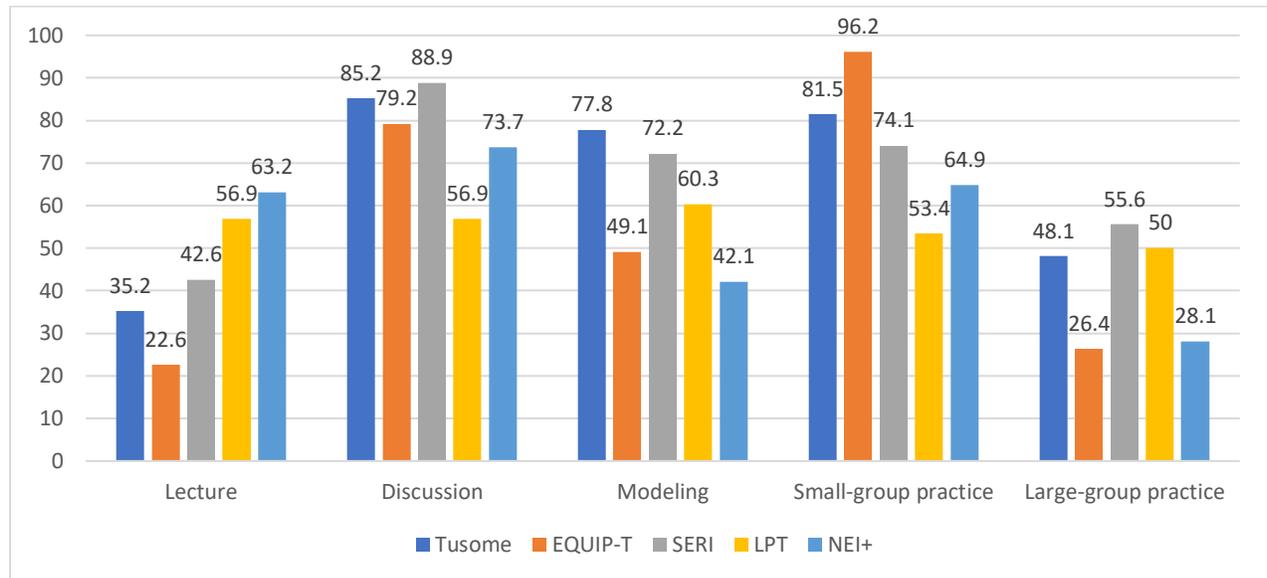
External coaching (i.e. coaching conducted by non-school-based personnel such as school inspectors or quality assurance officers) was a core component of Tusome, NEI+, and SERI, and over 85% of teachers in these three programs said coaching helped change their instructional practice. Coaching did not necessarily always increase in frequency, but teachers in these three programs pointed to the importance of receiving guidance from coaches on how to teach (76.1% to 91.4% in **Table ES-3**) and also noted that coaches under their respective programs were more supportive than coaches or inspectors prior to the program (44.4% to 71.7% in **Table ES-4**). Coaching under Lecture Pour Tous relied on a combination of head teachers and school inspectors (many of whom had no prior coaching experience), while teacher support for EQUIP-T was primarily focused on Communities of Learning, in lieu of coaching. However, under both programs, teachers noted that that coaches were more supportive and that receiving guidance from coaches had the most positive impact on their teaching **(Table ES-3 and ES-4)**.

Table ES-3.   Coaching factors that impacted teaching, as reported by teachers, by program

What about the coaching made you change your teaching? (Mark all that apply)	Tusome	EQUIP-T	SERI	Lecture Pour Tous	NEI+
Regular feedback from coach	50.0%	25.0%	25.7%	14.5%	25.0%
Receiving guidance from coach on how to teach	76.1%	83.3%	91.4%	30.9%	94.4%
Asking coach questions about my teaching	37.0%	66.7%	65.7%	20.0%	41.7%
Coach modeling instruction	41.3%	41.7%	65.7%	20.0%	33.3%
Coach helps with lesson planning	13.0%	50.0%	77.1%	10.9%	33.3%
Coach helps with classroom management	26.1%	50.0%	71.4%	3.6%	41.7%
Other	8.7%	8.3%	0	0	0

Table ES-4.   Differences between coaching before and during programs, as reported by teachers, by program

How are your interactions with the coach different than with coaches (or inspectors) before [program]? (Mark all that apply)	Tusome	EQUIP-T	SERI	LPT	NEI+
More frequent visits	34.8%	8.3%	28.9%	7%	41.7%
Coaches are more supportive	71.7%	66.7%	68.4%	41%	44.4%
I have more opportunities to ask questions	19.6%	50.0%	55.3%	33%	38.9%
I receive more helpful feedback	41.3%	58.3%	63.2%	26%	47.2%
I receive more suggestions for how to improve my teaching	30.4%	66.7%	76.3%	19%	55.6%
Coaches are friendlier	67.4%	58.3%	44.7%	4%	5.6%
This is my first experience with a coach or inspector	6.5%	16.7%	15.8%	11%	27.8%
No differences	2.2%	0	23.7%	30%	0

Lastly, by administering Early Grade Reading Assessments to students from observed classrooms, we were able to examine the relationship between teacher instructional practices and student performance. Although our analyses produced relatively few statistically significant associations, some relationships between classroom observation results and learning outcomes were identified in at least two of the three programs.

Using the instructional focus options from the Learning at Scale classroom observation tool, we found that "additional instructional time spent on *Grammar and Assessment*" was

negatively associated with performance (relative to a focus on *Reading*), that using no instructional *Materials* was negatively associated with learning (relative to the use of *Books*), that spending additional time focused on *Small Group* instruction was positively associated with learning but *Large Groups* were negatively associated with learning (relative to the *Whole Class*), and that teachers doing more *Demonstration* activities was negatively associated with learning. These findings were all statistically significant but their magnitudes were relatively small. Additionally, these results do not mean that these activities are ineffective on their own; instead, they mean that increased use of these activities relative to alternative activities (noted above) within these interventions was negatively associated with reading outcomes.

1.3.4 Qualitative findings from instruction interviews

In addition to the quantitative instructional interviews described above, qualitative interviews were conducted (during the same data collection window) with teachers, head teachers, coaches, meeting facilitators, ministry officials, and district officials for three programs: EQUIP-T, SERI, and Tusome. These qualitative interviews were planned but had to be dropped in the other five programs (Ghana Learning, Pakistan Reading Program, NEI+, Lecture Pour Tous and Read India) due to COVID-19 related limitations on the timing and staffing of these data collections.

Findings from these qualitative interviews largely confirm the quantitative interview findings. All three programs aimed to support teachers in changing their instructional methods by prioritizing practice at the trainings. The initial training, follow-ups, and ongoing peer support were structured so that teachers had time to practice new methods. Tusome teachers referred to practice as "microteaching" and discussed how the use of this practice differed from previous trainings in which facilitators merely gave instructions.

According to the trainers and others working with the teachers, the trainings improved teachers' self-efficacy. In some cases, trainers said the content of the training was new but that the training gave teachers the confidence to implement new instructional techniques. They also noted that teachers' confidence grew over time, which contributed to high levels of program implementation.

Regardless of the training model used (e.g., centralized or school-based), interviewees noted that the interactions between trainers and teachers were positive and collaborative in pursuit of a shared goal. These respectful relationships contributed to an environment where teacher concerns were addressed, and teachers were motived and engaged in the trainings.

When teachers began to implement the instructional approaches they had learned in training, their knowledge and skills were reinforced and improved by ongoing support and interactions through coaching visits and school-based teacher meetings. These interactions

provided support and motivation and solved problems. Interviewees' main concern about coaching was the decrease in interactions that occurred over time in some programs. In addition to teachers, coaches also reported growing professionally because of the programs. In the coaches' case, this growth was achieved explicitly through training, which helped them better serve their teachers.

In all three programs, instruction included teaching students the relationship between sounds and letters in a way that would support their ability to read words and increasingly complex text. The programs also encouraged students to be more active in their learning. Teachers learned and used instructional methods that shifted the focus from the teacher to the students, so students had a chance to demonstrate their proficiency with the new skills and practice them. As one teacher in Tanzania put it, "*It's not just me talking. Students are talking and doing things too.*" An additional benefit of students demonstrating their facility with a new skill was that it enabled the teachers to monitor progress and adjust their instruction accordingly. Teachers were also motivated by their students' demonstrations because they saw how their instruction contributed to positive student learning outcomes.

Fundamental to these programs were guides, books, and learning materials that made it easier for teachers to engage students. Depending on the program, the materials were either provided or teacher-made, but in all programs, they were central to the instructional methods for developing reading skills and were regularly used. Among materials for students, teachers spoke most highly of engaging stories and materials that allowed students to practice skills and connect words to real-world objects. The materials provided to teachers were also critical. Teachers' guides were recognized for including step-by-step instructions and for providing content in a student-friendly order with understandable lesson plans. Their benefits were best summed up by a coach in the SERI program who claimed that "*With the teacher guide, all the problems went away.*"

1.3.5 Qualitative findings from systems interviews

Qualitative interviews were conducted with program staff and government officials at all levels of the system from the ministry to subdistricts for all eight programs. The main aim of the interviews was to understand the relationship between what system actors did and program success.

Programs were able to achieve success when they were seen as important priorities by education system officials. Having evidence to demonstrate the success of the instructional model being introduced was one way to convince ministry counterparts of the value of the program. This was the case for Tusome in Kenya, which built on a prior program designed to research key aspects of instructional improvement that were ministry priorities. In Pakistan, government officials overcame initial reluctance when PRP began showing evidence of effectiveness. Demonstrating how a program is linked to existing ministry priorities or initiatives was another way to gain the support of ministry leadership, as was

the case of SERI in India—made easier because district officials had come to trust Room to Read based on their experience with an earlier program and the trusted data that it generated. In contrast, the government in Tanzania initially saw EQUIP-T as a separate program, with objectives that were not aligned to the ministry's priorities. This issue was overcome when the program adapted by working more directly with and through government structures and when the funding mechanism switched to one that permitted ministry line of sight into and some control over how resources were used.

Several programs systematically developed government capacity and transferred responsibility to government structures in stages. A good example of this is how SERI's "I do, we do, you do" approach progressively handed over responsibility for school support to local officials. NEI+ signed an MOU with state governments that resulted in states taking more responsibility for funding aspects of the program, such as student materials. Even in Tusome, which was designed to be implemented through government structures from the start, a progressive approach to developing, road testing, and reinforcing key capacities at the county and subcounty levels was needed to enable the system to provide productive support to schools and teachers. PRP was successful in revising the national curriculum for reading, while Ghana Learning had a lasting impact on approaches to school monitoring and coaching. Lecture Pour Tous successfully informed changes to the design of a new bilingual education model for the country.

High-level commitment and engagement from the government were found to translate into success at the school level when programs helped ensure adequate communication of that commitment down the system. Formal, official communication from ministry authorities is required, but following up those official messages with reminders and reinforcements through program-supported informal channels served to engage actors at the subordinate levels of the system. For example, district and subdistrict officials in Tanzania were important communicators, convincers, and influencers of teachers' attitudes regarding the new approach (as was the case for Curriculum Support Officers [CSOs] in Kenya), and importantly, they engaged in regular follow-up and communication directly with teachers. In addition, Tusome program communications via WhatsApp served as reminders to CSOs and county officials. The leader of the government Lecture Pour Tous program oversaw coordination, through an internal committee, of all arms within the Ministry of Education concerned by the program. Communications came directly from Ministry leadership to the regional and district levels, where whatsapp messenger groups helped accelerate delivery.

In addition to engaging key system actors at various levels, programs achieved success by building or reinforcing the capacity of the lower levels of the system to interact effectively with schools. This meant that the roles of individuals at these levels changed. The Ghana Learning program engaged system actors at all levels for developing and validating the learning materials, increasing stakeholders' capacity to develop and evaluate materials, as well as deepening their understanding and ability to train on these materials. In addition to

building the capacity of teachers, head teachers, and SSOs, NEI+ trained district officials were trained as master trainers so that they could train and supervise teachers. Additionally, district officials were developed into master trainers through SERI's work in India, while Tusome helped redefine the role of CSOs in Kenya to focus more explicitly on teacher support. In short, training and equipping these individuals and providing the resources needed for them to fulfill their roles were essential.

In addition to providing support, a key aspect of capacity at these lower levels of the system was monitoring progress—in terms of both implementation progress (e.g., teachers attending trainings, school visits being carried out) and student outcomes. Systems for collecting and using data to make sure activities were happening as planned (whether managed by the programs or by the government) appear to have helped reinforce the message that the system attached importance to program activities. Indeed, in each program, monitoring by district officials proved to be important for success.

Capacity development in key technical areas is one aspect that each program contributed to. Specifically, they helped strengthen capacity related to curriculum and materials development, teacher training (by building up cadres of master trainers), or school support services. We found no evidence that this kind of system capacity development contributed to program effectiveness, but it may play a role in sustaining aspects of these programs (for example, revised curriculum or updated approaches to teacher training).

Ultimately, the key ingredient, from a systems perspective, for producing measurable improvements in student learning outcomes was the change in organizational culture supported by the factors mentioned above:

Prioritizing program implementation

Setting clear expectations about what implementing the program would mean for actors throughout the system

Providing training to teachers and teacher support staff

Supplying the necessary inputs to allow counterparts to fulfill their roles, especially teachers

Engaging in ongoing follow-up and communication to reinforce new behaviors.

Combined, these aspects of the programs seem to have made it possible to alter the roles of key system actors in ways that placed greater emphasis on teaching and learning (as opposed to administrative compliance) and, thus, contributed to greater levels of support to schools and teachers.

1.4 Discussion and Recommendations

The implications and recommendations of this report are based on high-level findings from all programs, as well as Learning at Scale primary data collection efforts for seven programs. We will revisit these implications and recommendations in the final Learning at

Scale report. Here, we present these recommendations in three main sections: Effective Program Design, Program and Instruction Considerations, and Systems Issues.

Effective Program Design

- **Invest in Learning at Scale-type programs in response to COVID-19 learning loss.** COVID-19 has caused substantial learning loss in systems that were already producing poor learning outcomes. The Learning at Scale programs are examples of the sort of large-scale, highly effective programs that should be the primary focus of policy makers, donors, and implementers. We recommend that designing, implementing, and monitoring effective, large-scale programs to improve learning be the primary—if not exclusive—focus of the sector in 2022 and beyond.

Implement large-scale programs with rigorous evaluations. One of the main difficulties in undertaking the Learning at Scale study was the lack of evidence about effectiveness for some existing large-scale programs and the lack of scale of many rigorously evaluated programs. To improve outcomes at the scale needed, more programs must be implemented at scale and use designs that will allow us to measure impact and rethink implementation when and if programs struggle.

Fund what works. Our survey of the education sector revealed that for some multilateral and biliteral donors, we were unable to identify any large-scale, highly effective interventions, despite substantial funding investments. Donors should consider using the findings from Learning at Scale to design, implement, and evaluate future programs. Without evidence, we cannot be sure programs work. Without scale, we are not supporting the vast majority of students in LMICs who need better teaching to overcome upsettingly low learning outcomes.

Consider equity and regional differentiation. We were unable to identify effective, large-scale interventions in lusophone (Portuguese-speaking) contexts or Arabic-speaking countries, and we found only one such intervention in a francophone country. While several donors find working in anglophone contexts easier, the sector should not ignore countries or portions of countries that are not English-speaking. In fact, it may be that needs are larger in non-English-speaking LMICs, and we cannot be certain that what works in anglophone contexts will work elsewhere.

Consider structured pedagogy and teaching-at-the-right-level programs. Seven of the eight Learning at Scale programs can be characterized as either structured pedagogy programs (six) or teaching at the right level (one) in design. These programs, while relatively new in their large-scale implementation in LMICs, are showing substantial evidence of impact.

Program and Instruction Considerations

- **Utilize key program elements.** Our findings from across the eight programs show that ten program elements were included in several Learning at Scale interventions and identified as being key to program success. These elements are listed in detail on page 3, but we name the top five here:

    1. Teacher training focused on modeling and practicing new skills
    2. Structured teachers' guides
    3. Coaches' use of structured tools to support teachers
    4. Face-to-face methods for initial trainings
    5. Explicit and systematic teaching of skills using direct instructional pedagogical methods.

To improve reading, use phonics and spend time actually reading. Our classroom instructional findings are based on five programs. We found that that these programs focused the majority of instructional time on reading activities, used reading materials in classrooms as the primary resource, and primarily taught using an explicit and systematic phonics-based approach.

Incorporate practice, modeling, and discussion for effective teacher training. Train teachers differently, using more practice, modeling, and discussion. We observed that teachers in the Learning at Scale programs reported having more time for small-group practice and discussion. These teachers saw modeling, practice, and discussions of instructional methods as the most useful training methods. This type of focused skills-based teacher training contrasts starkly with the generalized teacher training pervasive in many LMICs.

Coach differently to impact outcomes. The consensus in the sector is that coaching works, if implemented appropriately. Our findings revealed that the type of coaching observed to work was more supportive, more friendly, and focused on pedagogical improvement rather than inspection. Engaging coaches should be coupled with coaches receiving training and structured tools to support teachers.

Avoid some classroom instruction "don'ts." We identified some instructional practices that were negatively related to learning outcomes within the broader interventions. Reducing time spent on reading in order to spend more time on Grammar or Assessment in the classroom was negatively associated with learning outcomes, as was using no instructional Materials compared with using Books. Additionally, while demonstration in the classroom is good, too much demonstration may limit the amount of time available for student practice and therefore negatively impact learning outcomes.

Use learning materials that work. Give students books, supply teachers with teachers' guides, and provide supplementary readers. We collected qualitative evidence that teachers and other educators saw the books, teachers' guides, and, in some cases, supplementary readers as contributing to program impact. With respect to student materials, teachers identified engaging stories and materials that allowed students to practice and connect to real-world objects as essential. Regarding teacher materials, they were better organized, easier to follow and teachers' guides with lesson plans were deemed important.

Monitor student progress throughout lessons. We found that teachers across all programs showed very high rates of 'responsiveness' to student needs in the classroom. This was demonstrated by teachers consistently checking for student understanding, circulating to monitor student progress, and rephrasing explanations when students didn't understand the content being taught.

Systems Issues

- **Align with government priorities.** Work closely and strategically with governments so that the program is a government priority. Our qualitative results spoke to government leaders as being essential to program success, particularly when the program and its requirements were seen as priorities for the officials themselves. This strategy works most effectively when the program is linked directly to existing ministry priorities or initiatives. The successful programs analyzed here found ways to do that, although those ways varied.

Communicate through the system to increase awareness and expectations of the program. We recommend working with government not only to gain buy-in to the program but also to communicate about program activities regularly through normal ministry

channels and informal ones. Such communication can be essential in convincing teachers of the value of the new methods and materials.

Enlist ministry counterparts in delivering and managing inputs needed to effect classroom change. Successful programs build clarity about how the program relates to the daily activities of midlevel civil servants in the government system. Work with the system to identify the roles of key actors in the system and set clear expectations for implementation for actors throughout the system. Monitoring can communicate the importance of expectations and provides a basis for accountability.

Implement specific capacity building related to school support and monitoring. Many studies recommend capacity building, but our findings go beyond that. Our evidence suggests that equipping low-level education officers with tools and knowledge about exactly how to support teachers implementing the program was essential and that this was most effective when government officers knew how to effectively monitor the program at the school, district, and subnational levels. Having ready access to supporting data and clear lines of sight between individual and system performance regarding implementation quality are also important.

Change the organizational structure to support pedagogical improvement. We recommend that efforts to focus on teaching, coaching, support, and monitoring go beyond piecemeal trainings. Instead, programs should work closely with government to reorient the entire system so that all actors understand their roles, implement ongoing support, and continually communicate to reinforce the new instructional behaviors needed to maintain program impact.

2 INTRODUCTION

2.1 Background

Learning outcomes are low and instruction has significant room for improvement in many low- and middle-income countries (LMICs). This situation is particularly concerning given the substantial learning loss due to the coronavirus disease 2019 (COVID-19) from which many systems are suffering. Although the number of successful programs is growing, relatively few have demonstrated impact at large scale. The Learning at Scale study was designed to identify existing programs with demonstrated impact on basic skills at scale and to conduct in-depth investigations of these programs to determine what makes them successful. This research is being led by RTI International and is part of the Center for Global Development (CGD) education research consortium, funded by the Bill and Melinda Gates Foundation.

The Learning at Scale study focuses on two areas: (1) identifying instructional strategies that are essential for improving learning outcomes at scale in LMICs and (2) learning about the characteristics of the systems within which successful scaled-up programs operate. Our research is designed to meet these two goals by studying eight of the most effective large-scale education programs in LMICs. The study entails collecting new data on classroom practices and education systems and reanalyzing existing evaluation data.

Given that complete data could not be collected from all eight programs as scheduled because of COVID-19, this report presents findings and implications based on initial discussions with all eight programs, systems interviews with all eight programs, and from school level data collected for five programs: (1) EQUIP-T, (2) Tusome, (3) SERI, (4) Lecture Pour Tous, and (5) NEI Plus.

2.2 Report Outline

After this introductory section, the report is organized as follows: **Section 3** covers the research design, with attention to the methods used and the process for instrument development. **Section 4** describes how the research team prepared for and carried out data collection (including training, data monitoring, and completed data collection efforts). **Section 5** discusses the key preliminary findings, with subsections that elaborate on program selection implications; the features of each of the eight selected programs; a matrix that cross-walks various characteristics found in each program, along with program-described key components; and quantitative and qualitative analyses of cross-program instructional and system approaches, stemming from our primary data collection efforts. **Section 6** tells the story that ties together the findings and offers recommendations based on them. Among the supplementary materials at the end of the report are a list of references cited and various technical annexes.

2.3 Program Selection

The first step in selecting programs for the Learning at Scale study was to develop inclusion criteria. After extensive discussions and reviews (both internally at RTI and with the CGD and Bill and Melinda Gates Foundation), we developed 10 selection criteria (**Table 1**). Rather than guaranteeing inclusion in the study, these criteria were designed to serve as screening measures for initial consideration.

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

Effectiveness:	▪ Evidence of causal impact at scale or causal impact at pilot with evidence of effective scale-up ▪ Local demand for the program
Scale:	Operating in most/all schools in at least two administrative subdivisions
Level of schooling:	Lower primary, upper primary, and secondary
Subject:	Includes a literacy component (may include other subjects as well)
Geography:	LMICs
Type of program:	Program aims to improve classroom teachers' effectiveness
Data available for analysis:	▪ Impact evaluation data ▪ Raw data on cost or cost-effectiveness
Time frame:	Active through 2019
Sector:	Public sector, civil society, or private sector
Access requirements:	▪ Key personnel and key stakeholders available for interviews ▪ Schools available for site visits (in high- and low-performing areas)

Initially, we were agnostic as to which academic subject(s) the candidate programs focused on, but we wanted to include programs that were similar enough to allow us to make useful comparisons. That goal ultimately led us to target programs that included at least a literacy component and that aimed to improve classroom teachers' effectiveness. Although we began with a more expansive search, we also narrowed the scope to programs working at the lower primary level. This was in no small part because there were so few large-scale, effective programs at the upper primary, secondary, and tertiary levels. Additionally, for programs to be considered, they had to be based in LMICs, have learning impact data available for analysis, and be able to give researchers access to schools and key personnel for interviews and site visits.

The most difficult criteria to meet were "effectiveness" and "scale." In the case of effectiveness, to make recommendations based on the examples we were studying, we had to make sure the programs led to meaningful improvements in learning. This meant that

they had to have conducted an impact evaluation on learning outcomes, preferably using either an experimental or quasi-experimental design. That said, there was some inherent conflict between the effectiveness and scale criteria because programs that had been scaled nationally could not be evaluated with a control group. For example, the Tusome Early Grade Reading Activity in Kenya was implemented in every public primary school in the country. In such cases, we relied on previous pilot evaluations and data on improvements in achievement over time as evidence of effectiveness. Another challenge was finding programs that had been evaluated and were still in operation. For example, Pratham's Read India government partnership program was evaluated in a randomized controlled trial (RCT) in Haryana State in 2012–2013 but was no longer working in collaboration with the government in that state as of 2019. However, large improvements in reading scores in other states—improvements that greatly surpassed national rates—served as strong evidence that similar programs functioning in those states were successful.

Additionally, we sought a measure of effectiveness that could be in some way comparable across programs. For this reason, we sought to identify effectiveness based on effect sizes—and due to its prevalence in program reporting, we opted to focus on oral reading fluency (ORF) scores (in terms of words read correctly per minute). Although we made this decision in order to have consistent measures of effectiveness across programs, it is important to note that (1) oral reading fluency scores should not be directly compared across languages and contexts; (2) effect sizes can be considerably impacted by differences in baseline values (and particularly in terms of oral reading zero scores and non-normal distributions); and (3) the focus on oral reading fluency scores as a selection criterion does not mean that it is the only measure of effectiveness for these programs or that it was even the primary focus for these programs.

We defined scale as "operating in at least two administrative subdivisions" (ideally with full coverage and no fewer than 500 schools), but we were also looking for initiatives that had been integrated with the government education system.

After contacting implementing partners, foundations, bilateral aid agencies, and university and think-tank researchers directly, we also issued a broad call for programs via a blog post. We ultimately spoke with more than 60 organizations about programs they supported, funded, or were aware of (**Annex A**).

For those programs with potential for meeting all criteria, we set up a one-hour interview between a program contact and a Learning at Scale team member. The interview focused on gathering information related to the 10 criteria but also included additional questions about the program's integration into the system in which it functioned, its sustainability, and its overall approach. For a list of key questions included in all interviews, see **Annex B.**

2.3.1 Initial selection

Initial and follow-up discussions led to the development of a preliminary list of 52 programs. Because the activity was designed to include only eight programs in total, it was necessary to determine a method for distinguishing the merits of each program to guide the selection of the final eight programs. We developed two approaches for review and consideration by CGD and the Learning at Scale Technical Working Group (TWG): (1) the "Raw Data Approach" and (2) the "Defined Category Approach." Both approaches were designed to simplify decision making by ranking all programs to be considered for inclusion. Ultimately, the Defined Category Approach was selected, though the results of both approaches were very similar.

2.3.2 Defined category approach

As a starting point for the Defined Category Approach, scale and effectiveness were each defined in terms of four distinct categories (Ideal, Moderate, Weak, and No), as shown in **Table 2.** The scale categories were intended to take into account both the number and coverage of schools to ensure that the size and scale of all potential programs were appropriate.

Aside from the Ideal category, which was based solely on effect size, the other effectiveness categories were defined in terms of both effect sizes and meaningful impact on reading ability (e.g., large increase in correct words per minute [cwpm] or the proportion of students[2] reading at benchmark levels).

Table 2.   Scale and effectiveness categories

Indicator:	Ideal	Moderate	Weak	No
Scale:	At least 500 schools AND universal (or near-universal) coverage in multiple subdivisions	At least 500 schools; coverage not universal but evidence of working through the government system	<500 schools OR working outside the government system	<100 schools or not systemic
Effectiveness: (Requires at least pre/post design; ORF or similar measure)	Effect sizes ≥0.4 standard deviation (SD) at scale	Effect sizes <0.4 SD but ≥0.15 SD AND meaningful impact on reading ability	Effect size <0.15 SD OR no clear meaningful impact on reading ability	No clear at-scale impact data (or no pilot impact with evidence of scale-up remaining the same)

These scale and effectiveness categories served as the basis for program scoring, along with the evaluation design criteria (**Table 3**). These criteria were developed and used only for

---

[2] Students: For ease of reference, we have standardized on the term "students"; in practice, the programs and countries of interest may prefer "learner," "pupil," or some other designation.

the Learning at Scale program selection process and should not be seen as evaluative descriptions of the programs themselves. Once the programs were ranked, if an organization was implementing more than one program in the list for consideration, we noted whether the program was that implementer's top-scoring program. This information did not factor into the score or the ranking itself but was recorded in an effort to potentially promote the diversity of programs, should a tiebreaker be needed. The reasoning behind using this factor as a tiebreaker was that implementers likely use similar models (for instructional and systems change) across contexts, and prioritizing diversity could generate more information about viable, effective approaches.

Table 3.   Program scoring: Overview

Component	Score range	Explanation
Total score	Minimum = 0; Maximum = 7	Score calculated by summing points for scale, effectiveness, and evaluation design
Scale	Minimum = 0; Maximum = 3	Ideal = 3 points; Moderate = 2 points; Weak = 1 point; No = 0 points
Effectiveness	Minimum = 0; Maximum = 3	Ideal = 3 points; Moderate = 2 points; Weak = 1 point; No = 0 points
Evaluation design	Minimum = 0; Maximum = 1	Experimental/quasi-experimental evaluation = 1 point; Pre/post comparison = 0.5 points; Other = 0 points
Top program by implementer	Not applicable	Used to promote diversity in implementers (used as a tiebreaker)

2.3.3   Final selection

Our recommendation to the TWG was to select the top eight programs ranked using the Defined Category Approach. All eight of the programs met the Ideal selection criterion for scale, while half met the Ideal criterion for effectiveness (with the other half meeting the Moderate criterion). After receiving approval from the CGD, Bill and Melinda Gates Foundation, and TWG, we identified the eight programs to be included in the study (**Table 4**). In alphabetical order by country, they are:

Ghana: Partnership for Education: Learning (Ghana Learning)

India: Read India

India: Scaling-up Early Reading Intervention (SERI)

Kenya: Tusome Early Grade Reading Activity (Tusome)

Nigeria: Northern Education Initiative Plus (NEI+)

Pakistan: Pakistan Reading Project (PRP)

Senegal: Lecture Pour Tous

Tanzania: Education Quality Improvement Program in Tanzania (EQUIP-T).

All selected programs agreed (and were excited to) participate in Learning at Scale. Our work began by examining these programs in depth and attempting to understand what made them so successful at scale. Additional details on program scale and effectiveness appear in the individual program descriptions in **Section 5.2.**

Table 4.   Top eight programs based on the Defined Category Approach, in alphabetical order

Program	Country	Lead implementer	Donor	Scale	Effectiveness[3]	Evaluation design
EQUIP-T (2014–2020)	Tanzania	Cambridge Education/ Mott MacDonald	United Kingdom's Department for International Development (DFID)	5,100+ schools in nine regions (63 districts)	0.5 SD effect size for Kiswahili reading ability (~9 cwpm increase in treatment schools)	Quasi-experimental design with matched control; external; 4 years
Ghana Learning (2014–2020)	Ghana	FHI 360	USAID	7,200+ schools in 100 districts	0.52 SD effect size for ORF in grades 1–2[4] across 11 Ghanaian languages (~6 cwpm increase) at midline[5] >30% reduction in zero scores	Quasi-experimental design with matched control; external; 1 year
Lecture Pour Tous (2016–2021)	Senegal	Chemonics International	USAID	All 4,000 schools in six regions	No effect size available but 13–18 cwpm increases in grade 2 at midline across 3 languages (Wolof, Pulaar, and Seereer) 29 percentage point increase in ORF benchmark	No control group (pre/post only)[6]; internal; 2 years

---

[3] Effectiveness is primarily reported in terms of improvements in cwpm, but this designation was for program selection purposes only. These results should not be used to directly compare results across programs, nor should this choice be seen as an endorsement of oral reading fluency as the primary measure of program effectiveness. For more information on using common metrics for reading performance across contexts, see the new Global Proficiency Framework in the United States Agency for International Development's (USAID's) EducationLinks newsletter: https://www.edu-links.org/resources/global-proficiency-framework-reading-and-mathematics#:~:text=The%20Global%20Proficiency%20Framework%20(GPF,nine%20in%20reading%20and%20mathematics.

[4] Grade: For ease of reference, we have standardized on the term "grade" throughout the report instead of "standard," "primary," "class," etc.

[5] Midline results were used at the time of selection (November 2019). Endline results have since been published, showing an 8.6 cwpm gain for treatment students at the end of grade 2.

[6] No control group was available for Lecture Pour Tous due to full program coverage in each target region and an absence of appropriate comparison schools teaching in national languages.

Program	Country	Lead implementer	Donor	Scale	Effectiveness[3]	Evaluation design
NEI+ (2015–2021)	Nigeria	Creative Associates	USAID	7,000–8,000 schools (10 districts per state)	ORF effect sizes of 0.2–0.7 SD in grades 2–3 at midline (~2–13-cwpm increase) in Hausa	No control group (pre/post only); internal; 2 years
PRP (2013–2020)	Pakistan	International Rescue Committee (IRC)	USAID	Seven provinces (~24,000 schools)	0.22–0.23 SD effect sizes for ORF in grades 1–2 in Urdu (~6 cwpm average gain) at endline (2017). Gains of up to 27 cwpm by province	Quasi-experimental design; internal; 2 years
Read India (2015–present)	India	Pratham	Multiple	250,000 schools in 10–12 states	0.15 SD effect size from an RCT in Haryana in 2015 (20 percentage point improvement in Annual Status of Education Report [ASER] top level in Hindi)	RCT (pilot); internal/external; peer-reviewed; 1 year
SERI (2015–2020)	India	Room to Read	United States Agency for International Development (USAID)	2,662 schools in four states	Grade 2 ORF effect sizes in Hindi of 1.3 SD for the partnership model (~18 cwpm increase); >20% increase in the number of readers reaching ORF benchmarks (≥45 cwpm)	Quasi-experimental design with comparison schools; internal; 2 years
Tusome (2015–2021)	Kenya	RTI	USAID	All 24,000+ primary schools	0.71–0.75 SD ORF effect sizes in grades 1–2 with 12 cwpm gain in grade 2 in Kiswahili	Pre/post only (national scale; no comparison possible); external; peer-reviewed; 2 years

[1] Note that DFID ceased to exist when its functions were merged into a new Foreign, Commonwealth and Development Office (FCDO) in mid-2020.
[2] Pratham has been implementing programs in India for 25 years, but the more formally structured government partnership programs began in 2007–2008. Timing for their program is state-specific.

2.4 Research Questions

The aim of the Learning at Scale study is to address three overarching research questions. The first two are focused on understanding the components of effective instruction, and the third is targeted toward understanding the system support that led to effective instruction.

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?

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 officials, and program staff. We also conducted secondary analyses of available impact data. Analyses of data across programs are used in the remaining sections of this report to illuminate the principles and intermediate goals of effective instruction, the methods programs can use to achieve these aims, how governments can scale up programs, and how system supports and program methods differ by context.

3   RESEARCH DESIGN

3.1 Learning at Scale Research Methods

In order to address the research questions listed in **Section 2.5**, we broke them down further into a set of hypotheses. Each hypothesis referred to a relationship between two actors in the system (e.g., ministry and district or teacher and student). Before developing the hypotheses further, we first introduce the theory of change guiding the hypotheses.

Figure 1 shows a 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

Figure 1.   Learning at Scale Generalized Theory of Change
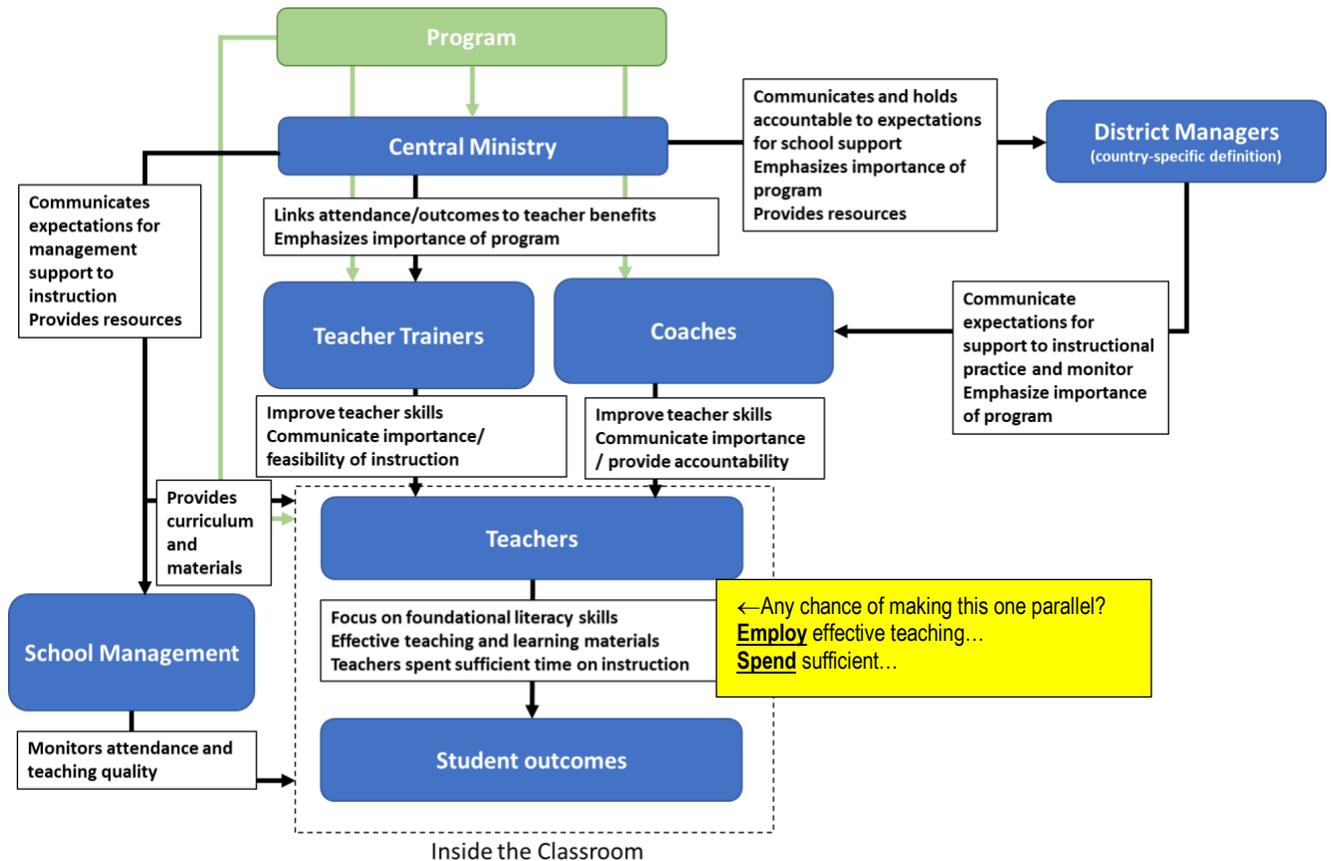
instantiations. For example, the principle of "central ministry emphasizes the importance of the program" will apply across all countries, but each country may use a different method (e.g., issuing circulars, revising job descriptions) to do this. In India and Pakistan, the state ministry fulfilled many of the functions ascribed to the central ministry in **Figure 1**. The aim

of the Learning at Scale research program was to test this theory of change in each program.

The programs we examined were funded by donors through implementing partners, rather than being initiated by the host government. The theory of change recognizes that each program may improve learning through the system (e.g., by working with the central ministry) or partly in parallel with the system (e.g., by providing curriculum materials directly to schools). The green arrows represent ways in which the program may work in parallel to the education system rather than through it.

3.1.1 Hypotheses

The design of research methods is driven by a core set of hypotheses about how systems achieve Learning at Scale. The hypotheses were identified by the research team, based on the literature. **Tables 5 through 8** set out the hypotheses at different levels, from classroom instruction to systems.

Table 5.   Instruction hypotheses

Hypothesis
1. Instruction shows students the relationship between sounds and letters (i.e., phonics),
a. Which relies on teachers using the program materials,
b. And relies on students having a chance to practice reading skills,
c. And requires teachers to demonstrate for students what they should do.
2. Teachers know the progress of more students because of increased interactions or informal monitoring and adjust their instruction accordingly.
3. Students have a chance to interact with print more than they did prior to the program.
4. Teachers use group or pair work.
5. Teachers adapt to student behaviors.
6. Teachers maximize instructional time.
7. Teachers demonstrate to students what they expect them to do.
8. Teachers are motivated because they see how the instruction leads to positive student outcomes.

Table 6.   Training hypotheses

Hypothesis
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 on how to do the lesson (from training and materials).
5. Prioritization: A realistic amount is expected of teachers during training/coaching.
6. The training improves teachers' procedural knowledge of effective teaching practices.
7. The training improves teachers' self-efficacy in implementing effective teaching practices (that they may have already known).
8. Teachers are inspired by the program.

Table 7.   Coaching hypotheses

Hypothesis
1. Coaches/meetings improve/reinforce teachers' knowledge and skills.
2. Coaches provide support/motivation/problem-solving to teachers.
3. Coaches provide accountability to teachers.
4. Coaches are provided basic resources – per diem, transport.
5. The program or government gives incentives. Resources are structured in a way to incentivize better coaching and more visits.
6. The program or government makes coaches accountable.
7. Coaches' job description reflects coaching functions.
8. Coaches believe in the program goals/purpose.
9. Coach is provided with training.
10. Coach is provided with tools to help observe and give feedback.
11. Program or government provides support on how to coach better.

Table 8.   System hypotheses

Hypotheses
1. Counterparts are informed about the program.
2. Counterparts play substantive roles in implementation.
3. Expectations for system counterparts are specified.
4. System communicates expectations for districts/schools/teachers/students.
5. System monitors performance relative to stated expectations.
6. Managers see subordinate parts of the system as having agency to solve/address problems.
7. Necessary inputs/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/developed by the program.
10. System actors can speak honestly about challenges faced in implementation.
11. Alternative hypothesis: Program (not ministry) leads communication with districts, schools.

3.1.2 Research methods

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

1.   Quantitative methods to estimate the counterfactual

In addition to the impact evaluations reported above in Table 4, we aimed to understand causal effects of programs using counterfactual reasoning. In programs where a control group was available, we used quasi-experimental methods to achieve this goal. In each program, a matching control group was selected and school-level data were compared between treatment and control groups. Examples of school-level data sources are classroom observations, student assessments, and teacher and head teacher interviews. This design allowed us to assess the counterfactual for key causal mechanisms. For example, interviews with teachers and program technical staff may indicate that a program was successful by increasing the focus on phonics in the classroom; the quasi-experimental design can tell us whether program schools did indeed have a greater focus on phonics compared to control schools. The design also allowed us to examine the relationship between teaching practices and students' achievement and to compare this relationship between treatment and control schools. For example, if the use of phonics in the classroom was more predictive of student outcomes in treatment schools than in control schools, this finding would support the hypothesis that phonics teaching was a causal mechanism by which the program had its effect.

2. Theory-based methods

The aim of theory-based methods was 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, we wanted to know whether—and how—the actions of the central ministry led to district managers performing their duties (i.e., monitoring coaches).
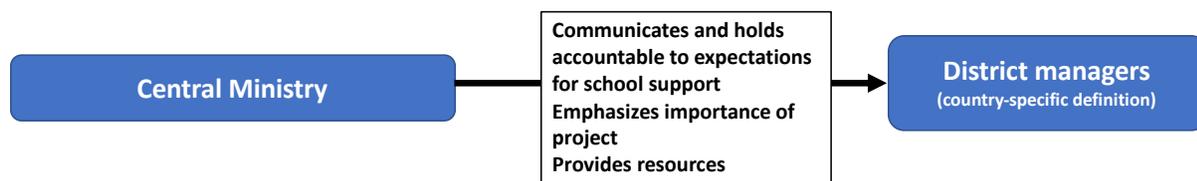
We investigated this link by questioning the central ministry about the actions they took to communicate with district managers, hold them accountable, support them, and provide them with resources. We also questioned district managers about their interaction with the central ministry. In both sets of 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 influenced 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 7**), 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 contradictory. For example, we had two hypotheses about how expectations were communicated to districts, schools, teachers, and students: by the government system (systems hypothesis #4 in **Table 8**) or by the program (hypothesis #11). In this case, stronger evidence for one hypothesis implied weaker evidence for the other.

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.

3. **Cross-case comparison**

We examined findings across programs. 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 analytical methods

The above methods resulted in several different types of questions being included in the data collection instruments, with corresponding methods for analysis.

1. **Quantitative description of programs (Quantitative data only)**

Quantitative surveys at the school and with coaches and trainers were used to assess program activities. This method applied to variables that did not apply to the control group, such as the percentage of teachers attending program training.

2. **Quantitative comparison between program and control schools (Quantitative data only)**

Where comparable data were collected, the difference in activities between program and control schools was analyzed using regression models, controlling for covariates such as class size. Data in this category came from classroom observations (e.g., time devoted to teaching reading in a class); some of the responses from interviews with teachers (e.g., confidence in teaching reading); and interview responses from head teachers, coaches, trainers, and meeting facilitators.

3.2 Instruments

1. **Questions that imply causal relationships (Quantitative and qualitative data only)**

Both qualitative and quantitative data were collected to assess respondent perspectives on the key elements of the program that led to change. The types of questions asked in this category included asking respondents to describe:

Action they took that led to changes further down the causal chain (e.g., things that a minister did that led to behavior change in districts, or things that teachers did that led to learning among students)

Aspects of the program that led to improved learning

Aspects of the program that were different from before the program started

Aspects of the program that were different from other programs

Sections 3.2.1 and **3.2.2** that follow discuss the tool development in more detail.

3.2.1 Instruction tool development

The tool-development process for this study consisted of four main steps:

1. Review of current literature (separately for instructional and system change components); see full list in the References section at the end of the report

2. Review of tools that have been used in previous research and programs; see list below in this subsection

3. Development of new tools

4. Piloting of tools.

This section provides an overview of this process for all tools created and piloted under the Learning at Scale activity.

Purpose

The theory of change underpinning the instruction component of this study is based on the Guskey (2002) model of teacher behavior change. The main theory states that teachers who receive training and support for implementing a new instructional approach will change their behavior based on the impact they see in their students. In other words, teachers who see a positive impact are more likely to change their behavior. This model supports the idea that there are some basic key components to teacher behavior change that most programs should include in various ways. Each of these components also has an impact on how likely—or unlikely—a teacher is to attempt to implement the new instructional approach or practices. Additionally, these components build on each other.

- **Curricula and materials.** Curricula and materials that target students' ability to read so they can understand other texts later will support improvements in learning outcomes. Using well-made materials may lead to improved student outcomes while also reducing the teachers' work to prepare their lessons. The combination of positive impacts on students with less work for teachers makes it more likely that teachers will continue to use the curriculum and materials.

- **Teacher training.** A successful teacher training will give teachers enough knowledge, skills, and motivation to feel confident in attempting to implement at least some part of the new curriculum and instructional practices using the supplied teaching and learning materials.

- **Post-training support.** Once teachers are trained, they receive support, either in the classroom or in community-of-practice meetings, that includes feedback and opportunities to reflect and ask questions as they begin to implement the new curriculum and instructional approaches they learned in the training. Providing support to improve implementation, reflect on practice, and compare student impact is more likely to lead to teacher behavior change. This type of support also fosters accountability for teachers who otherwise might choose not to implement the curriculum and instructional approaches that they learned, which may contribute to clear improvements in student outcomes.

Existing tools reviewed for instruction instruments

The existing tools that were reviewed as a part of the tool-development process are listed below. Full citations for the relevant resources that are noted along with each instrument appear in the list of references at the end of the report.

Classroom observation instruments

Classroom Assessment Scoring System (CLASS) Kindergarten–Grade 3 (K–3)
>Araujo et al. (2016)

Stallings Observation System
>Bruns et al. (2017)

Teacher Instructional Practices and Processes System (TIPPS)
>Seidman et al. (2018)

Teach
>Molina et al. (2018)

Early Language and Literacy Classroom Observation (ELLCO) Toolkit K–3
- Classroom Observation (1 of 3)
- Literacy Environment Checklist (2 of 3)
- Literacy Activities Rating Scale (3 of 3)
  >ELLCO (2006)

Protocol for Language Arts Teaching Observations (PLATO)
>Stanford University home page for PLATO instruments and resources: http://platorubric.stanford.edu/index.html

Framework for Teaching
>Kane & Staiger (2012); Kane et al. (2011)

Classroom Language Arts Systematic Sampling and Instructional Coding (CLASSIC)—adapted version for Observation and Classroom Inventory Checklist (Print Environment)
>Jukes et al. (2017)

Standards-based Classroom Observation Protocol for Educators in Literacy (SCOPE-L)
>Education Development Center, Inc. (EDC) (2018)

Observation des Pratiques Enseignantes dans leur Rapport avec les Apprentissages des élèves (Observation of Teaching Practices in Relation to Student Learning [OPERA]) Grid
>Altet et al. (2017)

Teacher professional development (TPD) instruments

In-service Teacher Training Survey Instrument (ITTSI)
>Popova et al. (2016)

Targeted Professional Development Scale (TPDS)
>Main & Pendergast (2017)

Characteristics of TPD Instrument
>Soine & Lumpe (2014)

Observation Checklist for High-Quality Professional Development (HQPD)
>Gaumer Erickson et al. (2020)

Coaching Observation Checklist
>Brussow et al. (2013)

Teacher Motivation Diagnostic Tool
>Hamm-Rodríguez et al. (2018)

New tool development

As previously mentioned, this study aims to discern what the components of successful at-scale programs are and how teachers experience these components in a way that leads to behavior change. With those ends in mind, after the research team's review of the data collection tools and literature, it made the most sense for the study to develop new tools grounded in the reviewed materials. The team agreed that each of the reviewed tools was useful for the intended purposes in some ways but that no tool fit all of the needs of the research, taking into account both the research questions and the planned study methods.

The newly developed tools needed to be valid and reliable, not only within a given context (such as a single country or region) but also across contexts. Based on the research methods and analysis plan, the new tools also needed to be both more structured than any of the reviewed tools, so that they would answer the research questions; and simpler, so that large teams of data collectors could be trained with reliability across various contexts, with as little bias toward any particular culture as possible. Although several of the tools reviewed include useful constructs, they are also complicated to train on, and many require a new adaptation of the tool for each context.

Finally, none of the reviewed tools seemed able to fully answer the research questions in a way that was appropriate for this study. This study will require data spanning the implementation of multiple components of each program and across various sources to get a more complete picture, and we did not find enough interview protocols to support this concern.

Using the subquestions as a guide, the team decided on a set of 16 data collection tools. **Table 9** shows the subquestions developed and the tools devised to respond to the questions.

Table 9.   Data collection tools by research question and subquestion

Category	Subquestion	Tool(s)
Research Question 1: What classroom ingredients (e.g., teaching practices, classroom environment) lead to learning in programs that are effective at scale? ▪ Do these ingredients differ by context (e.g., large versus small classes, high-capacity versus lower-capacity teachers)?		
Instruction	Is teaching practice X more common in effective versus ineffective classrooms or schools in each program?	Classroom Observation
	What instructional practices do effective programs have in common?	Classroom Observation
Research Question 2: What methods of training and support lead to teachers adopting effective classroom practices? ▪ Do the methods of training and support differ by context (e.g., high versus low budget)?		
TPD	What are the characteristics of TPD and teacher support in effective programs?	▪ Program Survey ▪ TPD Observation ▪ Coaching Observation ▪ Teacher Meetings Observation ▪ Interviews with Trainers, Coaches, and Facilitators
	What instructional practices do effective programs have in common?	▪ Program Survey ▪ TPD Observation ▪ Coaching Observation ▪ Teacher Meetings Observation ▪ Interviews with Trainers, Coaches, and Facilitators
Teacher perceptions of the program impact	For each teacher: What has been the biggest change in students' learning? What led to that change? What are you doing differently from before the program started? What are you doing differently in this program than in a previous (ineffective) program? Why?	Teacher Interviews on: ▪ Instruction ▪ Training ▪ Coaching ▪ Teacher Meetings

Two types of tools: Observation and interview

The instructional assessment tools summarized in **Table 9** can be assigned to two broad categories: *observation* instruments and *interview* instruments. Below, we describe the fundamental characteristics of each.

Observations

Observations focus on capturing what is really happening in a particular setting. For this reason, the TPD observation tools, for example, are meant to look for concrete practices and activities that are known to support teacher learning and behavior change during training and support sessions. The training and teacher meeting observation tools also

gather data on the proportion of time spent on each activity, such as lecture, practice, discussion, and modeling. This emphasis responds to the findings from the literature review, which confirmed that adults learn better if they are practicing new skills rather than just hearing about them or watching someone else demonstrate them. Also, a 2018 RTI study on teacher training (Piper et al., 2019) showed that the proportion of time spent on these four methods differed between programs that had shown success and programs that had not.

The observation tools are designed to be administered by data collectors with a variety of levels of experience in the classroom, including those who are new to classroom observation. The response options for each item observed are binary. Observers are required to note what is occurring during three different time segments. This approach makes the instrument similar to the Stallings tool (Bruns et al., 2017) but with more concrete items and longer and fewer time segments, so it does not require the same high level of training.

Interviews

The interview protocols were designed to reveal not only what the interviewees perceive is taking place in each setting but also how they are experiencing each component of a program and what factors ultimately change teachers' behavior and improve student outcomes. The instruments contain both closed- and open-ended items; however, most items also have options from which the data collector can choose based on an interviewee's response to a question. The preestablished options reduce the recording time, which allows teams to collect more data; and they also identify elements that are common across programs that may be contributing to success more reliably and efficiently than purely open-ended responses would allow.

Many questions across the interview tools ask respondents to distinguish between their experiences in the current program and their experiences in prior programs, or even before any program existed.

3.2.2 System tool development

Similar to the instruction tool-development process, after reviewing the literature and existing tools, the research team decided the best course of action would be to develop new tools for measuring system change, informed by the body of others' work but tailored to suit this study's research questions, theory of change (see **Section 3.1**), and context. It was important for these tools to fit the theory of change, to be simple enough for administration with fidelity across various countries and contexts, and to be informed by prior work in the field.

As noted above, the Learning at Scale theory of change starts with the four building blocks of a coherent instructional program:

5. Teacher training (TPD)

6. Ongoing support and coaching of teachers

7. Curriculum and materials

8. Assessment (formative assessment used by teachers; summative assessment used by districts or a broader system).

The questions incorporated into the instruments were designed to explore the ways in which the education system influences these four areas, as outlined in **Table 10.** The areas of interest were informed by both the core functions of effective systems in **Table 11** and Pritchett's (2015) design elements of accountability (delegation, finance, information, and motivation). For example, there are questions related to setting expectations (delegation), allocation of resources (finance), use of data (information), and recognition of teachers (motivation).

Table 10.   Ways an education system can support an instructional program

Focus area	System's direct support role	Broader system issues
Teacher training	**Set expectations (delegation)** ▪ Set expectations for and monitor teacher participation ▪ Set expectations for and monitor teacher performance post-training ▪ Design training in alignment with curriculum and learning goals **Provide basic inputs and targeted support (finance)** ▪ Deliver training to all teachers—planning, management, provision of trainers, funding, making payments ▪ Ensure quality of training—who trains and what mix of skills and motivation they have, as well as quality of training delivery and whether and how it is monitored **Motivation** ▪ Recognize teacher participation ▪ Recognize teacher performance post-training	▪ Teacher hiring, promotion, movement, and pay policies and practices ▪ Policies regarding allocation and use of TPD days ▪ Policies regarding recognition, credit, and reward (promotion) for attending continuous professional development (CPD) ▪ Development of strong school leaders ▪ Alignment between in-service and pre-service teacher training ▪ Alignment across goals, curriculum, instruction, and assessment
Curriculum and materials	**Set expectations (delegation)** ▪ Set expectations for what materials schools should be receiving, when, and how often **Provide basic inputs and targeted support (finance)** ▪ Design, develop, and produce materials	▪ Policy regarding "official" materials to be used in schools ▪ Corruption; national procurement policies (e.g., decentralized versus centralized) ▪ Maturity of local publishing and printing sectors ▪ Effective resource allocation (budget management)

Focus area	System's direct support role	Broader system issues
	▪ Distribute materials **Monitoring and use of data (information)** ▪ Maintain systems for tracking materials as they are delivered	▪ Policies regarding materials provision and replacement
Instructional support and coaching	**Set expectations (delegation)** ▪ Set expectations and accountability for coach performance (e.g., visits to schools, collection of data) ▪ Employ coaches as Ministry of Education (MOE)[7] staff, or to work in conjunction with MOE staff, or to be managed by MOE ▪ Determine job descriptions for coaches and process for selecting coaches **Provide basic inputs and targeted support (finance)** ▪ Ensure that coaches have the training and resources (e.g., tools, materials) to perform their jobs ▪ Finance coaching-related costs and determine how coaches receive reimbursement (ideally, incentivize the desired distribution and frequency of coach visits) **Monitoring and use of data (information)** ▪ Supervise coaches and manage their performance ▪ Set expectations and maintain systems for use of data from coaching	▪ Degree of decentralization and amount of authority delegated to the district or subdistrict level ▪ Amount of competing responsibilities assigned to districts and coaches ▪ Salaries sufficient to hire and retain good staff ▪ Merit-based employment and management practices; performance-based incentives
Assessment	▪ Set expectations for how teachers should use formative assessment and determine institutional norms for teacher practice regarding assessments ▪ Develop capacity of teachers and school leaders to conduct assessments and to analyze and use the data ▪ Determine standard practices in terms of who develops formative assessments ▪ Determine standard practices in terms of who develops summative assessments	▪ Whether any "stakes" are attached to assessment results (use of assessment results in accountability system or for recognition) ▪ Policies regarding publishing of assessment results (e.g., league tables) ▪ Infrastructure to support collection and storage of assessment-related materials ▪ Availability of skilled individuals to properly manage data collected at the system level ▪ Integration or links with other education databases (e.g.,

---

[7] Ministry of Education: We have standardized on this term for ease of reference when we discuss government support systems in general, with the acknowledgment that not all countries use this designation.

Focus area	System's direct support role	Broader system issues
	▪ Establish policies regarding standardized summative assessments ▪ Set expectations for how the system will use summative assessment data ▪ Incorporate formative assessments into teacher training and materials ▪ Manage and administer standardized summative assessments ▪ Build capacity and maintain systems for aggregating and using assessment data ▪ Build capacity of system players to analyze, interpret, and use data for decision making	education management information system)

As noted above, Table 11 summarizes the core functions of effective systems that were of particular interest for the Learning at Scale study. Full references for cited works appear at the end of this report.

Table 11.   Matrix of studied programs and their features

Citation	Geographic scope	System characteristics										
		Focus on student achievement (mission, goals)	Coherent approach to curriculum and instruction (standards, curricula, instruction, materials, assessment), aligned with goals	Use of evidence and data	Good communications and collaborative relationships with key actors (e.g., schools, parents, communities, ministries)	Investing in instructional leadership	Job-embedded professional development for leaders and teachers	Infrastructure (resources, policies, organizational structure) in alignment with learning goals	Accountability (internal and external) for results	Professional culture and collaborative relationships	Performance management and staff motivation
Banerji & Chavan (2016)	India	X	X	X	X			X			
Bottoms & Schmidt-Davis (2010)	United States	X	X	X	X	X	X	X	X		

		System characteristics									
Citation	Geographic scope	Focus on student achievement (mission, goals)	Coherent approach to curriculum and instruction (standards, curricula, instruction, materials, assessment), aligned with goals	Use of evidence and data	Good communications and collaborative relationships with key actors (e.g., schools, parents, communities, ministries)	Investing in instructional leadership	Job-embedded professional development for leaders and teachers	Infrastructure (resources, policies, organizational structure) in alignment with learning goals	Accountability (internal and external) for results	Professional culture and collaborative relationships	Performance management and staff motivation
Colbert & Arboleda (2016)	Colombia	X	X	X	X		X	X		X	
Crouch & DeStefano (2017)		X	X	X		X	X	X	X		X
Fleisch (2016)	Gauteng Province, South Africa	X	X	X			X	X			
Fullan & Quinn (2016)	Canada, United States	X		X		X	X	X	X	X	
Gallagher et al. (2016)	Ontario, Canada	X	X		X					X	
Handford & Leithwood (2019)	British Columbia, Canada	X	X	X	X	X	X	X	X	X	
Harris (2012)		X		X					X		
Kelly (2016)	Virginia, United States		X	X	X	X	X	X	X		
Knudson et al. (2011)	California, United States	X	X	X	X	X	X	X		X	
Leithwood (2010)	United States, Canada	X	X	X	X	X	X	X			
Leithwood & Azah (2017)	Ontario, Canada	X	X	X	X	X	X				

Citation	Geographic scope	Focus on student achievement (mission, goals)	Coherent approach to curriculum and instruction (standards, curricula, instruction, materials, assessment), aligned with goals	Use of evidence and data	Good communications and collaborative relationships with key actors (e.g., schools, parents, communities, ministries)	Investing in instructional leadership	Job-embedded professional development for leaders and teachers	Infrastructure (resources, policies, organizational structure) in alignment with learning goals	Accountability (internal and external) for results	Professional culture and collaborative relationships	Performance management and staff motivation
		System characteristics									
McAleavy et al. (2018)	London, England	X	X	X		X	X				X
Meyers et al. (2017)	United States	X	X	X	X	X	X	X	X	X	
Mourshed et al. (2010)	Global	X		X		X	X	X	X		X
Murphy & Hallinger (1988)	California, United States	X	X	X	X		X	X	X	X	
Osmond-Johnson & Campbell (2018)	Ontario, Canada	X	X	X	X	X	X	X		X	
Rincón-Gallardo (2016)	Mexico	X	X	X		X	X				
Shannon & Bylsma (2007)	United States	X	X		X	X	X		X	X	
Togneri & Anderson (2003)	United States	X	X	X		X	X	X			

3.2.3 Piloting of instruction and system tools

After the initial development of tools, the team needed to pilot the tools to ensure their validity and reliability. Kenya was selected as an ideal country for the pilot because one of the eight programs was located there and because RTI has a longstanding presence in-country with access to trained assessors and readily available schools for piloting.

The pilot exercise was conducted through the month of September 2019, with a follow-on in January 2020. The pilot consisted of the following stages.

Digitizing the tools for Tangerine®

After the tools were finalized, the team prepared Tangerine versions and administered internal quality checks in preparation for the phase 1 pilot. The digitized tools were used as the basis for all instruments administered throughout the study, using separate Tangerine groups for each data collection.

Recruitment and training of data collectors

RTI has a long history of conducting high-level education research in Kenya and has built a strong database of experienced data collectors there. The Learning at Scale team tapped into this database to select the best-performing data collectors for this exercise. Six assessors were selected for the pilot, comprising three former program staff and three experienced data collectors, all of whom had experience in classroom instruction, teacher coaching, and work with the government. The assessor team was trained for three days by two Learning at Scale team members, on both the tool administration and expectations for the pilot.

Data collection

Phase 1 pilot

The team first piloted the instruction tools in 21 public schools across Kenya, with a mix of urban and rural schools. In each school, a reading lesson was observed and interviews conducted with a grade 2 teacher, the head teacher, and the government's Curriculum Support Officer (CSO) who was serving as the school's trainer, coach, and meeting facilitator. System tools were piloted via interviews with MOE staff.

This process was conducted iteratively, with piloting and revision of instruments occurring in additional phases (see below) over the course of three weeks.

The preliminary findings from the phase 1 pilot were as follows:

9. The tools were valid and relevant for the purpose of the study. Most of the respondents could identify with the questions asked and were able to answer them reliably and with ease.

10. The administration by the assessors was easy, and the instruments appeared to function as intended.

11. There was a strong need to contextualize the tools for each country depending on program activities.

12. The tools were too long and needed to be shortened to ease the burden on respondents.

13. The government officials and program staff were excited to have their program included in the study and willingly participated.

Phase 2 pilot

A second, small pilot of the classroom observation tools was conducted in Nepal. The in-country team contextualized the tools according to instructions incorporated into the tool and then observed grade 2 classrooms in four schools. The pilot resulted in feedback to refine the instrument, including identification of items that needed further clarification, as well as recommendations for the contextualization process.

Phase 3 pilot

After revisions to the tools were completed, it was determined that the classroom observation tool should be piloted a third time because of the nuances in the tool that needed to be clarified for assessors and respondents. Key RTI researchers conducted this pilot in three schools in Kenya on three different days; analysis showed an 83% reliability score among the researchers. This level of agreement on scoring was above the required threshold, and thus, the tool was deemed reliable.

Phase 4 pilot

After the previous three pilot rounds, the tools were finalized and used in a mini-pilot in each of the program countries. This final usability test consisted of focus group discussions prior to assessor training to ensure that the tools were all contextually relevant and that the phrasing would be easily understood by the intended participants.

3.2.4 Final instruction and system tools

Based on the tool-development and piloting phases, the team ultimately finalized 31 instruments for use in this study. These instruments consisted of closed-ended interviews, open-ended interviews, observation instruments, and student performance assessments. All open-ended interviews and survey instruments were administered by Learning at Scale staff, while closed-ended interviews, observation instruments, and student assessments were administered by trained assessors in each country, as shown in **Table 12.** We created control school versions of the Teacher Interview and Head Teacher Interview to ensure that we could ask questions for which we wanted comparative results across treatment and control schools; they did not include intervention-specific questions that were appropriate only for program school interviews.

Table 12.   Overview of instruments for Learning at Scale[8]

Category	Instrument	Administered by
Instruction (Quantitative)	1. Teacher Interview	Trained Assessor
	2. Teacher Interview—Control schools	Trained Assessor
	3. Coach Interview	Trained Assessor
	4. Trainer Interview	Trained Assessor
	5. Meeting Facilitator Interview	Trained Assessor
	6. Meeting Facilitator Interview—Control schools	Trained Assessor
	7. Head Teacher Interview	Trained Assessor
	8. Head Teacher Interview—Control schools	Trained Assessor
Observation Instruments (Quantitative)	9. Class Observation Checklist	Trained Assessor
	10. Timed Classroom Observation	Trained Assessor
	11. Material Inventory	Trained Assessor
	12. Coaching Observation Instrument	Trained Assessor
Student Performance	13. Mini-Early Grade Reading Assessment (EGRA)	Trained Assessor
Instruction (Qualitative)	14. Teacher Interview	Learning at Scale Instruction Lead
	15. Trainer Interview	Learning at Scale Instruction Lead
	16. Meeting Facilitator Interview	Learning at Scale Instruction Lead
	17. Head Teacher Interview	Learning at Scale Instruction Lead
	18. Coach Interview	Learning at Scale Instruction Lead
	19. Teacher Meeting Observation	Learning at Scale Instruction Lead
System—"Central" MOE (Qualitative)	20. Key MOE Counterpart Interview	Learning at Scale System Lead
	21. MOE Teacher Training Interview	Learning at Scale System Lead
	22. MOE School Supervision Interview	Learning at Scale System Lead
	23. MOE Curriculum Interview	Learning at Scale System Lead
	24. MOE Material Development Interview	Learning at Scale System Lead

---

[8] Final instruments are available upon request.

Category	Instrument	Administered by
System—"District" Department of Education (DOE) (Qualitative)	25. District Education Manager Interview	Learning at Scale System Lead
	26. District Governing Council Interview	Learning at Scale System Lead
	27. DOE Teacher Training Interview	Learning at Scale System Lead
	28. DOE School Supervision Interview	Learning at Scale System Lead
	29. DOE Material Development Interview	Learning at Scale System Lead
Program Survey	30. Key Program Staff Interview	Learning at Scale System Lead
Donor Survey	31. Donor Interview	Learning at Scale System Lead

After finalization, we hired in-country translators to translate all tools to be administered by trained assessors (**Table 13**). After translation, these tools were reviewed by our in-country education consultants to ensure that appropriate technical language was used, where necessary. Finally, these tools were rendered in Tangerine for pre-piloting and training activities for each program.

Table 13.   Translation needs, by program

	Tusome	EQUIP-T	Ghana Learning	PRP	Lecture Pour Tous	NEI+	SERI	Read India
Training materials	English	Kiswahili	English	Urdu	French	Hausa	Hindi	Kannada
Quantitative interviews	English	Kiswahili	English	Urdu	French	Hausa	Hindi	Kannada
Mini-EGRAs	English, Kiswahili	Kiswahili	Not applicable (n/a)	n/a	Wolof, Seereer, Pulaar	Hausa	Hindi	Kannada

4 DATA COLLECTION

4.1 Training for Data Collection

4.1.1 Preparations

In preparation for the school-level data collection activity for each program, the Learning at Scale team led a six-day, in-person training of 26 assessor candidates[9]. The overall objective was to train the candidates to become fully capable of accurately and reliably administering all Learning at Scale instruments and protocols, to ensure that consistent, high-quality data would be collected across countries and programs.

The training team consisted of two RTI staff[10]: one lead trainer with extensive experience leading trainings on administering interviews and the EGRA, and one instructional lead with extensive technical expertise in early grade reading instruction and in design and administration of classroom observation tools.

Before the first training, RTI developed a manual that detailed step-by-step daily procedures. The documented training approach incorporated experiential learning and skills demonstrations (classroom practice videos, role play and peer practice, situated learning opportunities in real settings) and self-reflection and discussion sessions. At the end of each training, RTI updated the manual as needed, based on trainer and participant feedback. Adaptations included increasing the amount of time spent on practice, with particular emphasis on additional time for practice using the classroom observation tool.

The Learning 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 assessors would be required to cover all data collection components over the anticipated two-week span, 26 candidates participated in each training, allowing for

---

[9] With the exception for Senegal, where 18 assessors participated in training (assessor numbers were smaller because there were no control schools for data collection).

[10] Due to travel restrictions at the time of assessor training, RTI staff led training sessions in Nigeria remotely, via Zoom. All participants joined in person, and received additional support from consultants who participated in a three day training of trainers.

selection of the 24 trainees who scored highest on the accuracy and reliability measures administered throughout the training.

In the week before the training for each program, the lead trainer and in-country consultant conducted a focus group discussion with teachers to verify and contextualize the versions of the tools that had been translated from English into the local language(s). Additionally, one participant "scribe" was appointed at the start of each training, to document any issues that the trainees raised regarding the translations (for immediate review and revision in the tools, done by the training team).

4.1.2 Training structure

The trainings were structured around three categories of tools: interviews, observations, and an abbreviated (mini) EGRA. Specific competencies covered for each tool are described below.

Interviews

All interview questionnaires (listed below) were designed so that during data collection, assessors can explicitly follow clearly written scripts. Assessors can also clarify questions from respondents or probe where directed. Additionally, assessors are encouraged to establish a relaxed but professional rapport with interviewees. The training participants learned how to conduct all these interviews one-on-one and to record responses on tablets using Tangerine software:

- Teacher Interview
- Head Teacher
- Coach Interview
- Trainer Interview
- Meeting Facilitator Interview.

Classroom/coaching observation tools and checklists

The participants were trained in depth on all items in the observation tools and checklists. The facilitators used videos and two school visits to give the trainees opportunities to practice scoring in a realistic setting before beginning the actual data collection. The facilitators administered an assessor accuracy measure (AAM) to all the trainees, with training participants being accepted for data collection assignments only if their AAM scores were sufficiently high. The observation-related instruments were as follows:

- School Checklist
- Classroom Observation Checklist
- Timed Classroom Observation
- Coaching Observation.

<u>Targeted (mini) EGRA</u>

The attendees were trained to follow the standard EGRA protocols for administering the shortened version of the EGRA that would be used for data collection (i.e., they would follow detailed scripts for administration). During training, the participants practiced scoring mini-EGRA subtasks, and their reliability in doing so was measured against a "gold standard" (see details below) to produce AAM scores.

4.1.3 Overview of trainings

On Day 1, all participants received training on the administration of all interviews. On Day 2, the candidates were assigned to one of two groups based on their prior experience, interview AAM score, and trainer observations during Day 1. These groups focused on either the student assessment or the observation instruments. For most sessions from Day 2 onward, the lead trainer trained 13 assessors on administering the mini-EGRA, while the instructional lead trained 13 assessors on using the observation tools. All trainees participated in the two school visits and the sessions on school arrival and sampling protocols.

To measure the accuracy and reliability of the candidates' tool administration, over the course of the event, the training team administered five AAMs and two interrater reliability (IRR) tests:

- **One Interview AAM (Day 2):** All the assessor trainees observed a mock in-person interview conducted by the Learning at Scale in-country consultant and a data collection supervisor (both 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 4 and 5):** The participants watched a video of a lesson being taught and noted their observations. Their scoring was then compared to a gold standard pre-scored by the instructional lead.

- **Two EGRA AAMs (Days 4 and 5):** Trainees scored a live demonstration of a mini-EGRA, and their results were compared to a gold standard pre-scored by the lead trainer.

- **One Classroom Observation IRR Test (Day 5):** Groups of three individuals simultaneously scored the same lesson demonstration during the second school visit. Their scores were compared to evaluate the percentage of agreement among the three observers.

- **One EGRA IRR Test (Day 5):** One trainee administered the mini-EGRA with a student, while two others observed and scored. The scores were compared to determine percentage of agreement among the three trainee assessors.

Table 14 provides the average AAM scores on five measures taken during the training week, across the three countries where training was conducted. These scores represent all assessor candidates who participated in the training. As previously noted, the lowest-performing trainees were not engaged after the week of training; only the candidates with the top 24 scores proceeded to data collection.

Table 14.   Results of trainee accuracy and reliability measures

Measure	Average scores, by country				
	Tanzania	**India**	**Kenya**	**Nigeria**	**Senegal**
Interview AAM	**92%**[1]	**87%**	**92%**	**90%**	**96%**
Timed Classroom Observation, AAM 1	62%	— [2]	86%	44%	67%
Timed Classroom Observation, AAM 2	**88%**	**78%**	**79%**	**72%**	**80%**
Mini-EGRA AAM 1	92%	82%	98%	**90%**	89%
Mini-EGRA AAM 2	**98%**	**96%**	**98%**	— [3]	**93%**
Number of trainees per instrument	26 Interview 13 Classroom Observation 13 mini-EGRA	33 Interview 13 mini-EGRA 16 Classroom Observation	39 Interview 15 Classroom Observation 25 mini-EGRA	24 Interview 12 Classroom Observation 12 mini-EGRA	Interview 9 Classroom Observation 9 mini-EGRA

[1] Boldfaced values represent final reliability estimates for each measure.
[2] Due to technical and network issues, AAM 1 was not conducted in India, and the time was dedicated instead to additional practice and discussion.
[2] Due to a transportation breakdown while returning from a school visit, AAM 2 was not conducted in Nigeria on the last day of training.

The average AAM scores on the first timed mini-EGRA and the first Classroom Observation were lower, as expected, because the trainees were still learning on a sharp curve—through discussions, small-group practice, and the two practice school visits.

The research team set a minimum benchmark of 70% for assessors to be allowed to administer the classroom observation instrument, in comparison to 90% for the mini-EGRA. 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).

Table 15 is a representative agenda for a typical six-day training event.

Table 15.   Sample training agenda

Session and time	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
8:00–8:30 am	Arrival and registration	Arrival and sign-in	Arrival and sign-in	Arrival and sign-in	Arrival and sign-in	Arrival and sign-in
Morning Session 1 8:30–10:15 am	Introductions and overview of study program	Review of Day 1 and introduction to IRR, AAM, and tablets	1. Introduction to classroom observation 2. Introduction to mini-EGRA	**School Visit 1 (+ IRR test and feedback)**	**School Visit 2 (+ IRR test and feedback)**	Review
Break						
Morning Session 2 10:30 am–12:00 noon	Introduction to interviews (with consent)	Reading and practice—Trainer and Meeting Facilitator Interviews	Observation and practice: 1. Classroom observation 2. Mini-EGRA			Fieldwork and logistics for full-scale data collection
Lunch						
Afternoon Session 1 1:00–3:00 pm	Interview observation and practice—Teacher Interview	Practice; interview AAM	Observation and practice: 1. Classroom observation 2. Coaching observation	AAM #1: 1. Classroom observation 2. Mini-EGRA	AAM #2: 1. Classroom observation 2. Mini-EGRA	Supervisors only: Packing and preparation for fieldwork
Break						
Afternoon Session 2 3:15–5:00 pm	Reading and practice—Head Teacher and Coach Interviews	Preparation for classroom observation and mini-EGRA	Preparation for School Visit 1 and practice	Practice: 1. Classroom observation 2. Mini-EGRA and coaching observation	Practice: 1. Classroom observation 2. Mini-EGRA and coaching observation	

4.2 Data Collection Completed

4.2.1 Program entry visits

Once the eight programs were selected for inclusion in the Learning 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 for the Learning at Scale research team and the in-country program team. We administered extensive program surveys to in-country program leaders during each visit, gathering information beyond that obtained through our initial calls and conversations with program teams. We also met with donors, partners, research approval organizations, and potential data collection firms. Country leads were assigned to each program and were responsible for carrying out the program entry visits. The timing of these visits is shown in **Table 16.**

Table 16.   **Program entry visits, in chronological order**

Program	Country	Dates of visit
EQUIP-T	Tanzania	Week of November 4, 2019
Tusome	Kenya	Week of November 11, 2019
Ghana Learning	Ghana	Week of December 9, 2019
Lecture Pour Tous	Senegal	Week of December 16, 2019
SERI	India	Week of December 16, 2019
Read India	India	Week of December 16, 2019
NEI+	Nigeria	Week of January 6, 2020
PRP	Pakistan	Week of February 17, 2020

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.

4.2.2 Instruction and systems data collection

We originally had planned to complete all data collection activities by September 2020. However, our efforts were severely impacted by the COVID-19 pandemic, with many of our planned data collection activities put on hold. The status of each data collection as of April 2022 are presented in **Table 17.**

Table 17.   Status of data collection as of April 2022, in chronological order

Country	Program	Instruction data	Systems data
Tanzania	EQUIP-T	Completed February 2020	Completed February 2020
India	SERI	Completed March 2020	Completed March 2020
Kenya	Tusome	Completed March 2020	Completed March 2020
Pakistan	PRP	Cancelled due to COVID-19 and program closeout	Completed February 2020
Ghana	Learning	Cancelled due to COVID-19 and project closeout	Completed December 2021
Senegal	Lecture Pour Tous	Completed November 2021	Completed December 2021
Nigeria	NEI+	Completed March 2021	Completed June 2021
India	Read India	Cancelled due to delays in next-round implementation	Completed March 2023

Because primary data collection gathered only a portion of the data used in this report, we were still able to conduct analyses and present results across certain components for all programs. Included analyses are shown by program in **Table 18**.

Table 18.   Program data included in final report, alphabetical by country

Country	Program	High-level analysis	Primary Instruction Data	Primary Systems Data
Ghana	Learning	Included		Included
India	Pratham	Included		Included
India	SERI	Included	Included	Included
Kenya	Tusome	Included	Included	Included
Nigeria	NEI+	Included	Included	Included
Pakistan	PRP	Included		Included
Senegal	Lecture Pour Tous	Included	Included	Included
Tanzania	EQUIP-T	Included	Included	Included

Instruction: Quantitative sample

Quantitative data were collected by hired assessors using structured questionnaires rendered on tablets using Tangerine software.

By design, we planned to collect data in 60 treatment schools and 30 comparison schools in each program (as available[11]). The samples were not designed to be representative of the entire program but instead would rely on purposive sampling.

For programs without available comparison schools, the sampling process was as follows:

1. **Begin by selecting 3 "regions"[12] or higher-level administrative units.** Regional selection was informed by program suggestions and prioritized higher-performing regions to make it feasible to collect data from schools faithfully implementing the program.

2. **Select 2 "districts" within each selected region.** Programs were asked to name two to three recommended districts per region. They also provided a list of any districts to avoid.

3. **Randomly select 10 schools within each selected district** (plus two to three replacements).

4. **Randomly select 16 students within each selected school: 8 girls and 8 boys.**

Table 19 illustrates the sample-selection process for situations in which there were no available control schools.

Table 19.   Example of intervention-only school selection

Intervention only (60 schools)					
Region 1		Region 2		Region 3	
District A	District B	District C	District D	District E	District F
10 schools	10 schools	10 schools	10 schools	10 schools	10 schools

For programs with available comparison schools, the sampling process was modified slightly. Although we still selected three regions and two districts, they were divided among intervention and comparison samples. With this approach, we still obtained 60 intervention schools but also included 30 comparison schools. The comparison schools were matched with intervention schools to the extent possible (see **Table 20**).

---

[11] Some programs were unable to identify appropriate comparison schools. For example, Tusome had no available control schools because it is a national-scale program.

[12] Regions, districts: For clarity, we have used these terms generically across the program countries when we discuss them as a group.

Table 20.   Example of intervention and comparison school selection

Intervention and comparison (90 schools)					
Intervention (60 schools)				Comparison (30 schools; matched)	
Region 1		Region 2		Comparison Region 1	
District A	District B	District C	District D	Comparison District A	Comparison District B
15 schools	15 schools	15 schools	15 schools	15 schools	15 schools

Ultimately, quantitative instructional data was collected from five programs: SERI, Tusome, EQUIP-T, NEI+, and Lecture Pour Tous. The program-specific sampling approach was as follows:

- **EQUIP-T** – Two treatment regions and one comparison region were selected. Given the way that education implementation works in Tanzania, the comparison region received some donor-funded support. Within each region, two districts were selected. Within each district, 15 schools were selected. Comparison schools were matched to intervention schools using 2017 school-average exam scores.

- **SERI** – We selected one treatment district and one comparison district. Using data provided by local government officials, we chose four subdistricts (i.e., blocks) in the treatment district and two blocks in the control district. Within each block, 15 schools were selected. Comparison schools were matched to intervention schools using student–teacher ratios.

- **Tusome** – Because Tusome is a national program, no comparison schools were available for selection. We drew this sample by selecting six counties (following the procedures noted in Table 19 above) and then randomly selecting 10 schools in each county.

- **NEI+** - All data for NEI+ were collected in Bauchi State. Initially, three treatment local government areas (LGAs) were selected, based on input from the program team. Next, two comparison LGAs were selected based on their geographical proximity and similarities to the selected treatment LGAs. Within each treatment LGA, 20 schools were randomly selected; within each comparison LGA 15 schools were randomly selected. This constituted the final sample of 90 schools (60 treatment; 30 comparison).

- **Lecture Pour Tous** – Due to the scope of the original program (i.e. implementing in all schools in target regions), there was no viable option for comparison schools in Senegal. Based on discussions with the Lecture Pour Tous team, as well as the program's midline impact evaluation partner (EdIntersect), the final sample for this study was drawn from two of the program regions that best represented the program's full implementation while providing for sufficient variability in language of instruction. Ten schools were randomly selected in each of the three departments from the two selected regions. This constituted the final sample of 60 schools.

For each selected school, the expectation was to conduct one head teacher interview, to interview one teacher (grade 2), to observe one grade 2 reading lesson (from the selected grade 2 teacher) and to administer 16 mini-EGRAs (to randomly selected students from the

selected grade 2 teacher's classroom). For Tusome, two lessons were observed (one in English and one in Kiswahili).

Additionally, our subcontracted data collection firms were responsible for organizing coach interviews, coaching visit observations, trainer interviews, and meeting facilitator interviews at intervention schools, depending on program design (e.g., there are no coaches in the EQUIP-T model, so this interview was not a part of the program's data collection). Because the participants for these interviews tended to be based at the district level, it was not possible to conduct each interview at every school. Our goal was therefore to maximize the number of interviews and observations, based on each program's structure. The final sample counts for each program are provided in **Table 21***.*

Table 21. **Final sample sizes for treatment and control school-level data collection**

Program	Treatment or control	Teacher interview	Reading Lesson Observation (1 per school)	Head teacher interview	Coach interview	Coach observation	Trainer interview	Teacher meeting-facilitator interview	Student reading assessment
EQUIP-T	Treatment	59	59	59	n/a	n/a	n/a	31	944
	Control	30	30	30	n/a	n/a	n/a	n/a	479
SERI	Treatment	59	60	57	22	12	10	8	885
	Control	29	30	30	n/a	n/a	n/a	n/a	456
Tusome	Treatment	59	59 (Eng) 58 (Kis)	60	46	26	46	42	952
NEI+	Treatment	59	60	61	19	17	13	n/a	835
	Control	30	29	29	5	4	n/a	n/a	402
Lecture Pour Tous	Treatment	58	60	51	40	33	15	n/a	925

n/a = Not applicable.

Although the final sample sizes across instruments were close to the planned numbers, several instances occurred in which a teacher or head teacher was unavailable during the data collection period. For example, the SERI data collection took place just before the nationwide Holi holidays (Hindu festival of colors), so several teachers and head teachers had already left for vacation. In Tanzania, some schools were difficult to reach because of

washed-out or muddy roads due to rainy season, so one treatment school was not visited. In Kenya, one teacher was absent and the subsequent shutdown due to COVID-19 prevented the team from visiting a replacement school. At the student level, we had planned to assess 960 students in treatment schools and 480 students in control schools. However, some selected schools had fewer students in attendance on the day of data collection, so the final numbers were slightly shy of the intended sample size.

Instruction: Qualitative sample

All qualitative interviews were conducted by a member of the Learning at Scale team, with a translator where necessary. Interviews were guided by a structured questionnaire which aimed to elicit evidence for and against the hypotheses (**Section 3.1.1**).

EQUIP-T – Tanzania

Instruction interviews for EQUIP-T took place February 12–14, 2020, in seven schools in four districts: two districts in each of Simiyu and Dodoma Regions (**Table 22**). All interviews took place at the schools.

Table 22.   Interview summary: EQUIP-T

Region and district	Grade 1 teacher	Grade 2 teacher	Head teacher	Meeting facilitator
Simuyu Region				
Maswa District		1	1	1
Bariadi Rural District	1	2	2	1
Dodoma Region				
Mpwapwa District		2	1	
Kongwa District		1		1

SERI – India

Instruction-related interviews for SERI were collected March 2–7, 2020, in two districts—Balodabazar and Bhatapara—at six schools from the larger sample. The schools were a mix of town and rural locations close enough to travel there from a base point and return in the same day. At each school, a teacher and a head teacher were interviewed. Two of the head teachers were former teachers who had used the SERI program. Three coaches (i.e., meeting facilitators) and five trainers were interviewed either at a school, at the District Education Office, or at the Block Education Office. The trainer and meeting facilitator had worked in intervention schools but were stationed in Raipur District, outside the intervention district, when interviewed.

Table 23.   Interview summary: SERI

District	Teacher	Trainer	Head teacher	Meeting facilitator
Balodabazar Block	3	4	3	1
Bhatapara Block	3		3	
Raipur District (Palari Town)		1		2

Tusome – Kenya

Instruction-related interviews for Tusome were collected March 9–12, 2020, in two counties on either side of the country. Two schools were visited in western Kenya in Elgevo-Marakwet County, near the town of Iten. Three schools were visited in eastern Kenya in Nyeri County, near Nyeri town. These five were rural schools outside of any major town but close enough to the team's lodgings to visit and return in one day. At each school, two teachers were interviewed, and at all but one school, the head teacher or deputy head teacher was interviewed. As indicated earlier, CSOs played the role of trainer, coach, and meeting facilitator for the Tusome program. These officers are assigned to a zone (i.e., a geographic cluster of schools) within a county. During this data collection period, two CSOs were available to be interviewed.

Table 24.   Interview summary: Tusome

Zone	County	Grade 1 teacher	Grade 2 teacher	Head teacher	Trainer, coach, meeting facilitator
Kapchemutwa	Elgeyo-Marakwet	1	1	1	1
Kessup	Elgeyo-Marakwet	1	1	1	1
Mathira North	Nyeri	1	1		
Mathira West	Nyeri		2	1	
Municipality North	Nyeri	1	1	1 (deputy)	

Systems: Qualitative sample

EQUIP-T – Tanzania

System-level interviews were conducted in Tanzania during February 9–21, 2020. At that time, we interviewed eight central government officials in Dodoma as well as in Dar es Salaam. We also interviewed 10 district-level officials in four districts. We were able to interview DFID's two relevant education advisors and the two most senior leaders in Cambridge Education's Tanzania office. The Learning at Scale team selected the interview

participants to represent a range of geographical coverage and roles within the education system. A detailed list of system-level interviews is presented in **Table 25.**

Table 25.   System-level interviews in Tanzania (EQUIP-T)

Level	Department or position
Donor	1. DFID Education Adviser 2. DFID Education Officer
Program	3. Cambridge Education Deputy National Coordinator 4. Deputy National Coordinator
Central government (Dodoma and Dar es Salaam)	5. Permanent Secretary, Ministry of Education, Science and Technology 6. Deputy Permanent Secretary, Ministry of Education, Science and Technology 7. Deputy Permanent Secretary, Director of Tanzania President's Office for Regional Administration and Local Government (PO-RALG) 8. National Coordinator EQUIP-T, PO-RALG 9. Previous Director, Tanzania Institute of Education (TIE) 10. Technical officer, TIE 11. Assistant Director, Primary Education 12. Assistant Director, Monitoring and Evaluation
Bahi District	13. Quality Assurance Officer 14. District Education Officer
Chamwino District	15. Quality Assurance Officer 16. Academic Officer 17. District Education Officer
Dodoma District	18. Quality Assurance Officer 19. Academic Officer 20. District Education Officer
Kongwa District	21. Quality Assurance Officer 22. Academic Officer

<u>Tusome – Kenya</u>

In Kenya, system-level interviews were conducted March 9–13, 2020, and October 12–16, 2020. During that time, we interviewed 16 ministry officials in Nairobi across four counties (Elgeyo-Marakwet, Nyamira, Nyeri, and Siaya), USAID's Contracting Officer's Representative for Tusome, and Tusome's current and former Chiefs of Party. Ministry interview participants were selected jointly by the Learning at Scale team and Tusome staff, to ensure a range of geographical coverage and roles within the ministry.

Central ministry counterparts accounted for three of the system interviews, while county and subcounty personnel accounted for 13 interviews (**Table 26**).

Table 26.   System-level interviews in Kenya (Tusome)

Level	Department or position
Central ministry (Nairobi)	1. Tusome National Coordinator (2016–2017) 2. Tusome National Coordinator (current) 3. Director General
Elgeyo-Marakwet County	4. Quality Assurance Officer 5. Subcounty Officer 6. County Support Officer
Nyamira County	7. County Liaison Officer 8. Subcounty Officer 9. Deputy County Director, Teachers' Service Commission (TSC)
Nyeri County	10. County Director, Education 11. County Project Coordinator
Siaya County	12. County Director, TSC 13. County Director, Education 14. Quality Assurance 15. County Liaison Officer 16. County Support Officer

SERI – India

System-level interviews in India (refer to **Table 27**) began with a phone interview with the donor on January 29, 2020. In-person interviews took place in Chhattisgarh State March 3–6, 2020. Interviews in Raipur, the state capital, involved representatives of the state government, Room to Read, the state office of the United Nations Children's Fund (UNICEF), a donor for the SERI program in Rajnandgaon District, and education officials for Raipur District, where Room to Read's direct implementation model had been implemented prior to SERI. Interviews in Baloda Bazar were with district officials involved in the SERI program in that district, which was the focus of our research on SERI. Phone interviews also took place with officials responsible for districts where SERI was implemented in the states of Madhya Pradesh, Uttar Pradesh, and Uttarakhand. These phone interviews took place March 18–20, 2020.

Table 27.   System-level interviews in India (SERI)

Level	Location	Department or position
National	New Delhi	1. Senior Education Specialist, USAID

Level	Location	Department or position
Chhattisgarh State	State Ministry; program and donor offices	2. State Pedagogy Coordinator 3. Assistant Professor, State Council for Education, Research and Training 4. Lecturer, State Council for Education, Research and Training 5. Education Specialist, UNICEF (donor for SERI program in Rajnandgaon District) 6. State Manager, Room to Read 7. Senior Program Officer, Room to Read
	Baloda Bazar District	8. District Mission Coordinator 9. District Education Officer 10. Assistant Project Coordinators x 2 11. Former Assistant Project Coordinator and SERI district focal point 12. Block Resource Coordinators x 2
	Raipur District	13. Principal, District Institute of Education and Training, covering Raipur and Baloda Bazar 14. Lecturer, District Institute of Education and Training
Madhya Pradesh State	Barwani District	15. District Project Coordinator
Uttar Pradesh State	Varanasi District	16. District Coordinator, Training
Uttarakhand State	Champawat District	17. Senior Lecturer, District Institute of Education and Training and SERI district focal person
	Dehradun District	18. Additional State Project Director, *Sarva Shiksha Abhiyan* (state Education for All campaign for universal primary education)

<u>Pakistan Reading Project</u>

System-level interviews were conducted in Pakistan from February 20 to March 4, 2020. During this time, interviewers met with 18 ministry officials across four provinces (Islamabad Capital Territory [ICT], Azad Jammu and Kashmir [AJK] Province, Khyber Pakhtunkhwa [KP] Province, and Sindh), as well as with the USAID Agreement Officer's Representative for PRP, and a range of program staff across three provinces (ICT, AJK, and Sindh). Ministry interview participants were selected jointly by the Learning at Scale team and PRP staff, to ensure a range of geographical coverage and roles within the ministry. The final sample of interview participants for PRP is displayed in **Table 28.**

Table 28.   System-level interviews in Pakistan (PRP)

Level	Department or position
Federal Directorate of Education	1. Curriculum 2. Training and Coordination

Level	Department or position
Islamabad Capital Territory (ICT)	3. District Education Officer – Tarnol 4. PRP Program Leadership 5. PRP Agreement Officer's Representative – USAID
Khyber Pakhtunkhwa (KP) Province	6. Secretary of Education 7. Provincial Institute of Teacher Education 8. Directorate of Curriculum and Teacher Education 9. Chief Planning Office 10. District Education Officer – Dera Ismail Khan subdivision 11. District Education Officer – Karak District 12. District Education Officer – Upper Dir District
Azad Jammu and Kashmir (AJK) Province	13. Secretary of Education
Sindh Province	14. Secretary of Education 15. Provincial Institute of Teacher Education 16. Sindh Textbook Board 17. Directorate of Curriculum, Assessment and Research 18. Taluka Education Officer – Shikarpur District 19. District Education Officer – Karachi Central District 20. Taluka Education Officer – Umarkot District 21. PRP Head of Office – Sindh Province

<u>Ghana Learning</u>

System-level interviews were remotely conducted in Ghana from October 27 to December 14, 2021. During this time, interviewers met with 6 ministry officials at the national level, as well as with one district level official. Previous interviews were also completed with USAID: Ghana mission representatives and a range of program staff. Ministry interview participants were selected jointly by the Learning at Scale team and Ghana Learning staff, to ensure a range of roles within the ministry among people familiar with the program. The final sample of interview participants for PRP is displayed in **Table 29.**

Table 29.   System-level interviews in Ghana Learning

Level	Department or position
National	1. Coordinator of Private Schools (GES) 2. Former Acting Director for Basic Education; Director for Early Childhood Education 3. Former Director of National Teaching Council (NTC) 4. Former Executive Secretary of NaCCA 5. Former Executive Secretary of NIB

Level	Department or position
Donor	6. Donor Partners Coordinator
District	7. District Director of Education (DDE)

Northern Education Initiative Plus (NEI+)

System-level interviews in Nigeria were conducted remotely from May to June 2021. During this time, interviewers met with 10 government officials in Bauchi State, including officials from 3 LGEAs. Previous interviews were also conducted with project staff and USAID COR at the start of the project.

Table 30.   System-level interviews in NEI+

Level	Department or position
Bauchi State	• (Acting) Permanent Secretary, MOE
	• (Retired) Permanent Secretary, State Universal Basic Education Board (SUBEB)
	• Head of EMIS and Research, MOE
	• Director Planning, Research and Statistics, MOE
	• Director Planning, State Ministry of Budget and Planning
	• Permanent Secretary, Bauchi State Agency for Nomadic Education
	• Chief Lecturer, College of Education, Azare
Local Government Education Authority	• Education Secretary, Ninji LGEA
	• Education Secretary, Gamawa LGEA
	• Education Secretary, Ningi LGEA

Lecture Pour Tous

System-level interviews at the regional and district level were conducted in Senegal between November 15 and 19, 2022. During this time, interviewers met with 5 officials across two regions (Kaolack and Fatick). Interviews at the national level were conducted over Zoom in December 2022. Ministry interview participants were selected by the Learning at Scale team under advisement by Lecture Pour Tous and with the support of an in-country consultant. The final sample of interview participants for PRP is displayed in **Table 31.**

Table 31.   System-level interviews in Lecture Pour Tous

Level	Department or position
National	1. Inspectors in the DEE (Elementary Education Directorate) (2 person) 2. Staff of DFC (Training and Communication Directorate) 3. Head of Communications Department, DFC 4. Director of INEADE
Inspection d'Academie de Fatick	5. District Focal Person, Fatick 6. Director, Regional Teacher Training Center
Inspection d'Academie de Kaolack	7. Head of School Inspection, Kaolack 8. Program Focal Person, Kaolack 9. School Inspection Officer), Kaolack

Read India

System-level interviews at the state and district level were conducted in Karnataka, India between March 6 and 10, 2023. During this time, the country lead for India met with three district education officials, one program coach, one program staff member and two State level education officials (one now retired). Interview participants were selected by the Learning at Scale team under advisement by Pratham and with the support of an in-country consultant.

5   FINDINGS

5.1 Program Selection: Similarities

The first findings for this study resulted from our search and selection of programs for inclusion. Our final list of eight programs had several unintended similarities. For example, although we initially constrained the choice to programs that at least had a literacy component, we were open to those including other subject areas as well. However, few programs with rigorous results and at large scale also had results from other subject areas. Similarly, whereas the ultimate decision was to focus on lower primary programs, there were few possibilities with rigorous evidence at scale at the secondary or preprimary levels. Additionally, the search for programs was exhaustive, but the final list of potential programs did not include any lusophone (Portuguese-speaking) or Arabic-speaking programs, and included only one francophone program. We will have the opportunity to revisit this selection through additional Learning at Scale activities in 2021.

Interestingly, two characteristics united most of the interventions selected for inclusion in the first phase of this work. First, none of the interventions could be characterized as government only, with a government managing and implementing and funding the program on its own. This does not mean that there were no exciting government program interventions that were worth studying. Rather, such programs are seldom rigorously evaluated (although many do have repeated cross-sectional data using national or cross-national assessments), which was an important part of our selection criteria. We remain interested in examining programs that are run, managed, and implemented by governments and will be examining this set of options during the Learning at Scale expansion. We also feel a need to focus additional effort on understanding how best the sector could encourage governments engaged in these interventions to conduct more rigorous evaluations.

The second primary characteristic that united the interventions was their funding source. Six of the programs were funded at least in part by USAID, one was funded by DFID, and one had multiple donors. It is worth noting that despite careful examination of programs sponsored by the World Bank, the Global Partnership for Education (GPE), UNICEF, and several bilateral donors, none of their funded programs fit the Learning at Scale criteria. Often these programs fell short because of the smaller scale of their interventions, or because of the lack of rigorous impact evaluation data, or because the evaluation data did not reveal large-scale impacts on learning. One of the main focuses of the Learning at Scale study is to provide guidance on what is required to develop the types of effective large-scale interventions that are included in this work.

5.2 Program Overviews

5.2.1 EQUIP-T: Education Quality Improvement Program in Tanzania

Introduction

EQUIP-T was a six-year, £80 million program (2014–2020) funded by DFID.[13] This Government of Tanzania-led program, implemented by Cambridge Education, had five major components, several of which are beyond the scope of Learning at Scale's analysis. These included (1) improved access to high-quality education, (2) strengthened school leadership and management, (3) strengthened district planning and management, (4) stronger community participation and demand for accountability in education, and (5) improved learning and dissemination. As a whole-school development program, EQUIP-T had a broader focus than many of the other Learning at Scale partners. The initiatives in this program went beyond literacy and numeracy classes and targeted all teachers across all classes and subjects.

---

[13] Note that DFID's metamorphosis into FCDO occurred in mid-2020, after EQUIP-T's closeout.

EQUIP-T's geographic reach in Tanzania depended on the regional divisions of donor support among then-DFID, USAID, UNICEF, and GPE. Based on that geographic distribution, EQUIP-T began implementation in five regions that were classified as disadvantaged. The program expanded to 9 regions out of Tanzania's 31 total regions in 2018. The full EQUIP-T program was implemented in 5,196 primary schools, benefited more than 3.2 million students and 55,000 teachers, and supported 63 Local Government Authorities (LGAs), which are the equivalent of districts.

Evaluation

Given the regional distribution of the interventions in Tanzania, EQUIP-T's external evaluator, Oxford Policy Management (OPM), was unable to set up a typical RCT with assignment to treatment and control groups. Instead, the EQUIP-T evaluation strategy was to use a quasi-experimental design comprising propensity score matching and difference-in-differences analysis. Using the propensity score matching technique to select treatment and control schools, EQUIP-T's external evaluation involved collecting three rounds of large-scale evaluation data. OPM conducted the baseline evaluation in 2014, the midline evaluation in 2016, and the endline evaluation in 2018, although some of the resulting analysis reports were not publicly released until 2019. The evaluation was unique in that it incorporated both quantitative and qualitative studies in a mixed-methods approach. The regional basis of the literacy and numeracy treatment and control groups also introduced the potential for some spillover effects from the GPE Literacy and Numeracy Education Support (LANES) program or the USAID Tusome Pamoja program in the EQUIP-T control regions.

Figure 2 presents the impact graph for Kiswahili literacy from the analysis report for OPM's endline external evaluation (Rawle et al., 2019). It shows how learning outcomes changed over time across the distribution. The analysis focused on the proportional reduction of those at the lowest performance band from 39% to 16%. On the other end of the distribution, the percentage reaching the basic grade 2 level—i.e., bands 2E (yellow) and 2A (orange) combined—increased from 36% to 50% of the assessed population. Notably, the proportion in the highest achievement band increased from 12% at baseline to 18% at endline, close to the targeted 21% but lower than the midline results (Rawle et al., 2019).

Figure 2.   Summary of EQUIP-T impact on Kiswahili literacy, grade 3
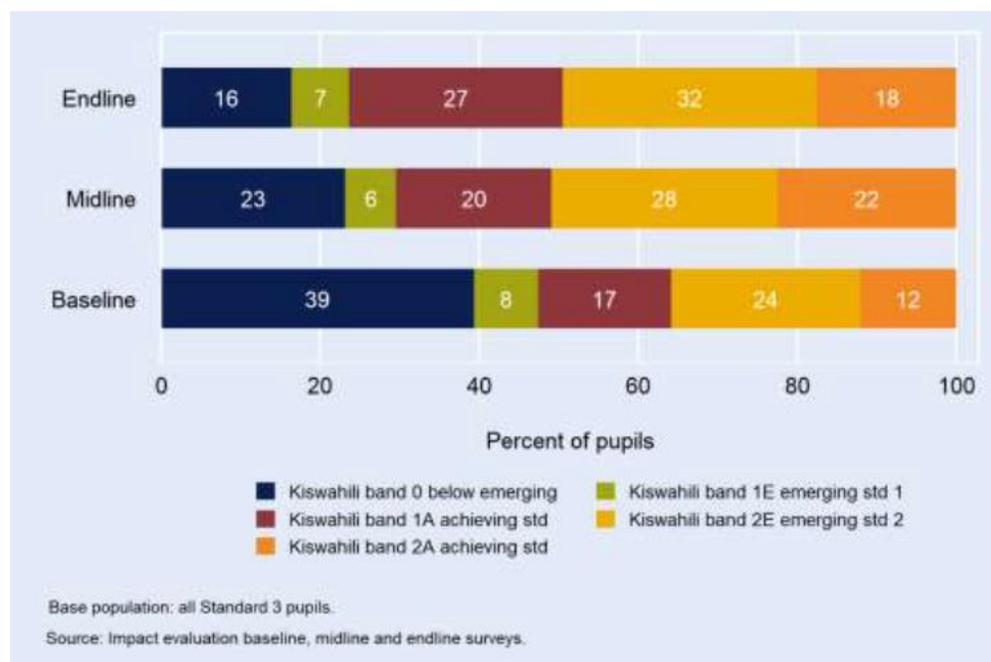
Source: Rawle et al., 2019, p. 25. Std 1, std 2 = grades 1, 2.

The impact of EQUIP-T on numeracy was less substantial (refer to **Figure 3**), with reductions in those at the lowest band from 13% at baseline to 9% at endline[14]. This change was accompanied by the proportion of students at the highest band increasing from 4% to 9%. The external evaluation noted that the gains from baseline to midline in numeracy were more than 0.2 SD but that the overall gains from midline to endline were not statistically significant.

---

[14] These smaller numeracy increases are likely due in part to the fact that numeracy modules were implemented later in the program (and that the endline dates were not revised to align with program extension dates).

Figure 3.   Summary of EQUIP-T impact on numeracy, grade 3
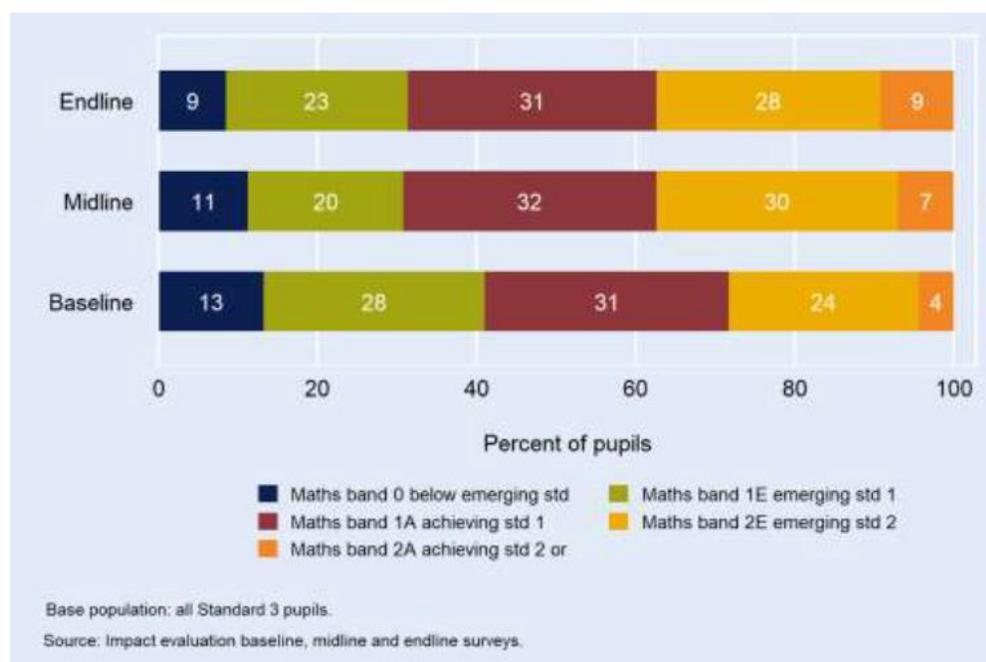
Source: Rawle et al., 2019, p. 26. Std 1, std 2 = grades 1, 2.

The final program evaluation measured the impact of EQUIP-T from baseline to endline and found it to be 0.5 SD for average Kiswahili scores and 0.3 SD for average numeracy scores. These results did not meet the benchmarks indicated in the program logical framework but did constitute significant impacts on learning. It is important to note that these impacts were achieved during a time of significant increases in class sizes due to the government's fee-free education policy. Grade 3 enrollments increased from 136,782 to 203,173 between baseline and endline, with average class sizes increasing from 59.9 to 82.2.

Program model

The EQUIP-T program was originally designed as a whole-school improvement program, but in 2015 it was adapted to better align with the Government of Tanzania's efforts to improve learning outcomes, with some interventions looking to increase learning in the "3Rs" (reading, writing, arithmetic); it also was related to the broader government goal for Tanzania to become a middle-income country by 2025. Moreover, EQUIP-T included substantial interventions at the systems level, such as budget support to the government, which none of the other interventions studied by Learning at Scale utilized. In addition, as noted in the introduction to this section, EQUIP-T had components that supported school management, district education management, construction, and community engagement in the education sector.

EQUIP-T's interventions started at the national level, working with both the Ministry of Education, Science and Technology and PO-RALG. At the subnational level, EQUIP-T had

regional teams working in the five (and eventually nine) program regions, and at the LGA level below the regions. Team members at the regional and LGA levels combined to work with the national-level teams, as well as with larger education interventions that were not limited to learning in literacy and numeracy, to create substantial overall interventions in these locations. **Figure 4** illustrates how EQUIP-T had visualized the various components working together to achieve the main goal—better learning outcomes, especially for girls—after EQUIP-T's expansion in 2019.

Figure 4.   Updated EQUIP-T theory of change
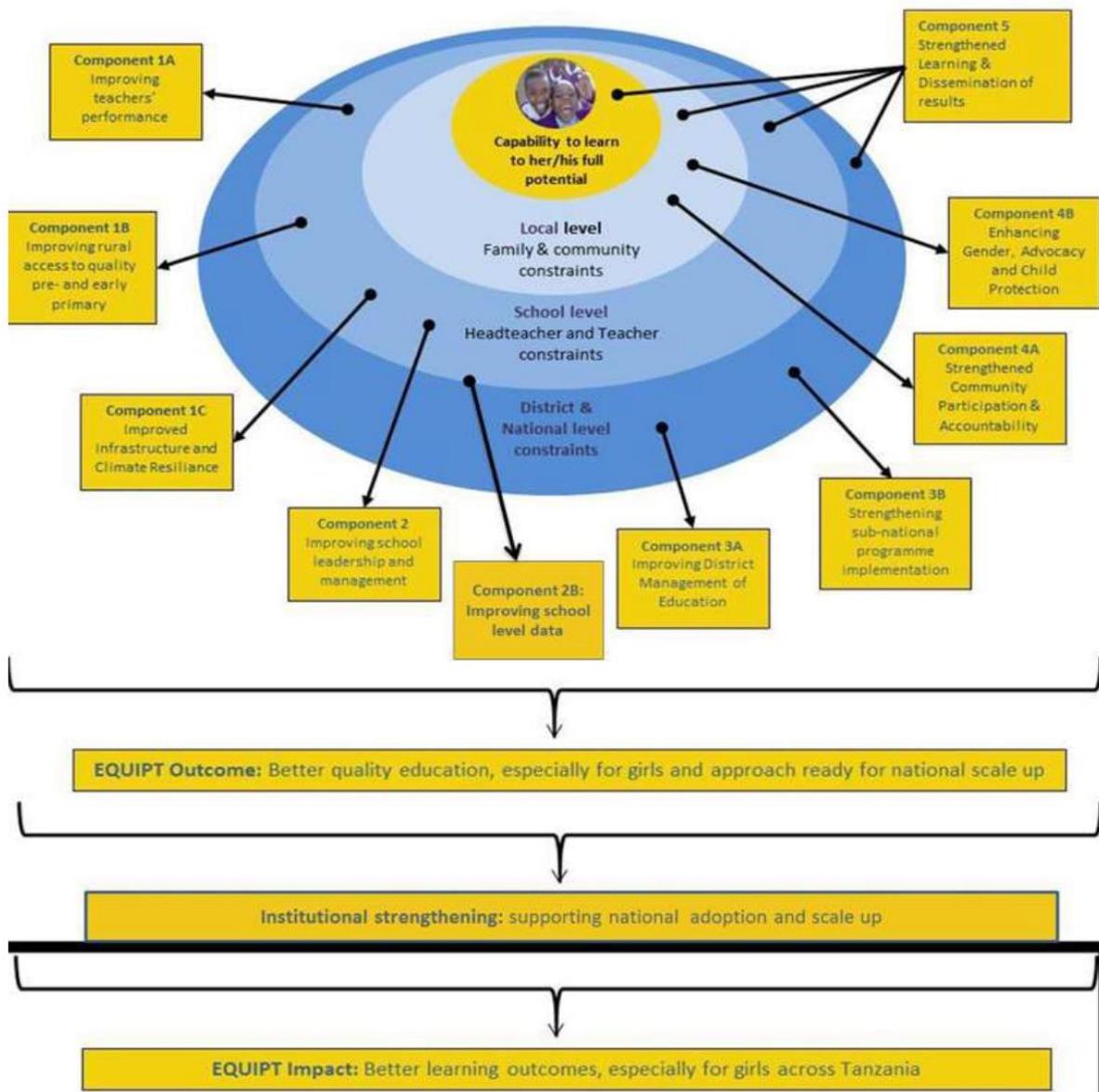
Source: Rawle et al., 2019, p. 131.

<u>Materials</u>

EQUIP-T facilitated the development of instructional materials for Kiswahili literacy and numeracy, focusing particularly on the lower grades. Most relevant for the purposes of the Learning at Scale study, EQUIP-T developed 13 literacy and numeracy modules covering the various domains of learning, which were delivered to teachers at the school level through a cascaded training modality (described below). In addition, EQUIP-T developed supplementary readers; teacher read-aloud books, also called "big books"; a literacy toolkit for each school; a numeracy toolkit for each school; short videos for both literacy and numeracy; and other instructional aids that were delivered to schools at intervals during the program's life span. EQUIP-T made the materials publicly available at http://www.equip-t.org/resources/. The midline and endline evaluations showed only limited use of the big books and supplementary readers, however, given the lack of time in the official timetable for reading practice. Teacher-made materials were an important part of the program. EQUIP-T provided materials and training for teachers to create pictures for vocabulary lessons, to design letter cards, and to teach with flipcharts.

<u>Teacher training</u>

Fundamental to EQUIP-T's teacher training model was the assumption that when teachers are trained, they will improve their instructional behavior, and that the easiest location in which to undertake that training is at schools. The program trained school-based in-service education and training (INSET) coordinators as school-based facilitators for CPD. Ward Education Officers (WEOs) and academic teachers were also part of the school-based supporting system to incentivize teachers to actively participate in school-based CPD. .

Note that in the EQUIP-T program, Communities of Learning (COLs) served as a key element of the teacher training as well as teacher support structures (see below). Unlike many other Learning at Scale programs, most teachers received the EQUIP-T training content at school-level COL meetings, led by the academic teachers who had received the module content at the previous levels of the cascade.

<u>Teacher support</u>

EQUIP-T's teacher support system primarily emphasized implementing school-level COLs led by academic teachers. The key non-school-based officers, the WEOs, set up the COLs, reinforced their importance, and encouraged head teachers and academic teachers to manage them. EQUIP-T supported the WEOs by supplying motorbikes to reduce transportation costs, and the low ratio of WEOs to schools meant that it was practical for the WEOs to visit schools frequently. Note that the WEOs' chief role was to be a champion of COLs implemented at the school by academic teachers rather than managing COLs directly or observing and coaching teachers themselves.

EQUIP-T's key instructional improvement strategy focused on the intersection of teacher training and teacher support. EQUIP-T developed the 13 literacy and numeracy modules

described above for groups of teachers to use at school-level COL meetings. After the teachers completed the existing modules, the participants were tasked with developing their own topics for discussion at subsequent COL meetings. This expectation fared better in some schools than in others.

Pedagogical approach

EQUIP-T's interest in learning outcomes was timely in Tanzania given that in 2015, the government—at the Presidential level—had begun to emphasize improving learning in the 3Rs; EQUIP-T, therefore, was in place at an opportune time to support this new priority. EQUIP-T was different from the other Learning at Scale programs in terms of its pedagogical approach. The other seven programs can be classified as having structured pedagogy; EQUIP-T is the only one of the interventions that did not. EQUIP-T's theory of change gave specific pedagogical methods used in classrooms a lower profile and was not organized around changing the instructional core that teachers were applying. For example, no particular book was used every day in a given subject, and EQUIP-T initially did not have structured teachers' guides (although the numeracy materials changed over time in this regard). There were two aims: (1) to help teachers interpret the new syllabus in the government's competency-based curriculum (CBC) that emphasized the 3Rs; and (2) to make space for teachers, academic teachers, and head teachers to meet in COLs to discuss content from modules that offered somewhat more general pedagogical improvements than were common in many of the other Learning at Scale programs. That said, EQUIP-T was associated with the use of more phonics in literacy classrooms and more active-learning materials in numeracy classrooms. There was a greater focus on student participation and student practice of skills in all classes.

Systems

EQUIP-T's budgetary processes were unique compared to the other seven studied Learning at Scale programs. Only EQUIP-T gave money directly to the government structures with which they were working. Although it complicated the implementation, this arrangement created a higher degree of access to LGA decision makers and other Tanzanian government leaders than typically is found in education programs. The access, in turn, was important for pushing the other priorities of EQUIP-T.

Another key element of EQUIP-T's work at the system level was that each of the program's technical tasks had an institutional home within the government, so that local-level implementation issues could be solved (at least in part) from the outset. Examples are EQUIP-T's provision of motorbikes to WEOs, which allowed these key education officers to be fully available for EQUIP-T implementation; and provision of fuel and office writing materials to the WEOs.

5.2.2 USAID Partnership for Education: Ghana Learning

Introduction

The early grade reading program under the USAID Partnership for Education: Learning activity ("Learning"), was a five-year activity that was conceptualized, designed, and implemented in Ghana by FHI 360 during a shortened time frame from 2016 to 2019.[15] The program supported Ghana's Ministry of Education and the Ghana Education Service, as well as a range of educational institutions, to improve reading performance for early grade students in public primary schools. Learning worked in 11 Ghanaian languages of instruction (LOIs) in 100 districts across all 10 regions of the country.[16] Ultimately, the program trained more than 51,000 teachers, head teachers, and curriculum leads, and reached 707,843 students in kindergarten through grade 2.

Evaluation

The Learning impact evaluation was conducted externally (by Social Impact), using a quasi-experimental design (Social Impact, 2019). Baseline data were collected in 2017, with midline (Social Impact, 2018) and endline data collected in 2018 and 2019, respectively. The evaluation employed a difference-in-differences model, estimating growth in treatment schools against growth in comparison schools, using repeated cross-sectional data from 470 schools (235 treatment and 235 comparison), 940 teachers, and 9,400 students. All students (grades 1 and 2) were assessed using an EGRA in a local Ghanaian LOI, as well as in English.

The endline evaluation showed an overall impact of 8.7 cwpm (averaged across Ghanaian LOIs), for treatment over comparison students in grade 2, after two years of intervention. This gain equated to an effect size of approximately 1.2 SD. Grade 2 zero scores were reduced by 47 percentage points. For grade 1 students, the impact on Ghanaian LOI was a smaller but statistically significant 3.1 cwpm, with a 38 percentage point reduction in zero scores.

Program model

The theory underpinning Learning was that materials, training, and coaching, along with continuous monitoring, would lead to instructional changes and improved learning outcomes for students from the second level of kindergarten (KG2) to grade 2. Instruction was designed as a phonics-based Simple View of Reading approach, with scripted lesson plans. Coaching consisted of two parts: district-level coaching from District Teacher Support Teams

---

[15] The contract for this program was signed in 2014, but the program was completely redesigned in 2016, so the technical work and implementation of what is conventionally known as "Learning" ran from 2016 to 2019.

[16] English instruction was an initial focus of the program but was dropped with the 2016 redesign.

(DTSTs)[17] and in-school coaching from head teachers and/or curriculum leads. There were also teacher meetings (led by head teachers) at the school level, meant as an opportunity for teachers to discuss any issues they were having in the classroom. See **Figure 5** for the complete theory of change.

---

[17] DTSTs are Ghana Education Service staff who are mandated to visit schools as a part of their regular job function.

Figure 5. Ghana Learning theory of change (Part I)

	Inputs	Assumptions	Outputs	Assumptions	Outcomes	Assumptions	Goal Impact
Training and Materials	Training of MTs consisting of DTST in EGR and use of program-developed teaching and learning materials Teacher guides with scripted lessons in the 11 official Ghanaian languages Pupil books in the 11 Ghanaian languages Classroom kits including alphabet strips, flashcards, posters, and readers Development of ASER tool for continuous pupil assessment	MTs successfully train HTs, CLs, and teachers on reading instructional model Program materials are designed, printed, and distributed in sufficient numbers and according to plan	Teachers receive five days of initial in-service training and termly refresher trainings on phonics-based instruction from MTs Teachers are provided with teacher guides with scripted lesson plans Pupils are provided with learning/activity books Classrooms are provided with supplementary materials and books Teachers are trained on continuous assessment using ASER	Training content is accurately and fully covered using a ToT/cascade model Teachers understand, absorb, and retain the training content and are motivated to put it into practice Teachers use scripted lesson plans as prescribed Pupils regularly access learning books and supplementary readers Materials are properly cared for and stored Trained teachers do not transfer or leave teaching profession Teachers are able to teach in the language of scripted lesson plans in teacher guides Teachers regularly assess pupils	Teachers demonstrate mastery of phonics-based approach to teaching reading in the Ghanaian LOI Classroom time-on-task improves for pupils and teachers Pupils are more likely to engage with books and learning materials, both inside and outside the classroom Teachers adapt instructional approach to pupils' needs through continuous assessment	Materials and practices are an improvement upon prior materials and practices Class environment is conducive to absorbing lessons Pupils are punctual and attend class with enough regularity to benefit from structured pedagogy Teachers teach reading lessons with prescribed frequency and duration Pupils understand the LOI and materials Classroom-level interventions can lead to improved reading regardless of home environment	Improved reading outcomes in LOI for P1 and P2 pupils

Figure 5, continued: Ghana Learning theory of change (Part II)

	Inputs	Assumptions	Outputs	Assumptions	Outcomes	Assumptions	Goal Impact
Coaching and Monitoring	HT and CL receive training on instructional leadership and coaching from MTs Trained DTST-LSs on providing targeted technical support Trained CSs on program monitoring Tablets and data platform for CS monitoring Facilitator guides for Head Teachers and CLs	HTs and CLs are able and willing to observe and coach teachers and facilitate SBI meetings with prescribed frequency Teachers are able and willing to participate in SBI meetings with prescribed frequency CSs are able/willing and have sufficient resources to conduct monitoring visits with prescribed frequency CSs follow monitoring protocols and record data effectively CSs are effective in identifying gaps in instruction and coaching and are able to rally DTST LSs for additional coaching Teachers come to school often enough to be monitored and coached with prescribed frequency	HTs and/or CLs observe and coach teachers on material use on a bi-weekly basis HTs and/or CLs facilitate school-based inset (SBI) meetings for one hour each week CSs conduct bi annual monitoring visits to monitor FOI of scripted lesson plans, coaching and SBI meetings CSs arrange DTST LS support visits (as needed) for weak HTs and CLs	HTs, CLs, DTST LS, and CSs understand, absorb, and retain the training content and materials and are motivated to put them into practice HT, CL, and DTST LS coaching feedback is understood and useful to teachers SBI meeting content is understood and useful to participants Regular monitoring of teachers can improve performance by informing training needs and holding teachers accountable Monitoring data is reaching intended recipients and they are able and willing to act on it	Teachers demonstrate mastery of phonics-based approach to teaching reading in the Ghanaian LOI and oral English Classroom time-on-task improves for pupils and teachers Teachers adapt instructional approach based on SBI learnings, coaching, and monitoring	Materials and practices are an improvement upon prior materials and practices Class environment is conducive to absorbing lessons Pupils are punctual and attend class with enough regularity to benefit from structured pedagogy Teachers teach reading lessons with prescribed frequency and duration Pupils understand the LOI and materials Classroom-level interventions can lead to improved reading regardless of home environment	Improved reading outcomes in LOI for P1 and P2 pupils

Source: Social Impact (2018), pp. 74–75.

Figure 5 legend:

ASER	Annual Status of Education Report
CL	Curriculum Leader
CS	Circuit Supervisor
DTST	District Teacher Support Team [member]
EGR	early grade reading
HT	head teacher
LOI	language of instruction
LS	Literacy Specialist
MT	mother tongue
P1, P2	primary grades 1, 2
SBI	school-based INSET (in-service education and training)
ToT	training of trainers

Materials

Ghana Learning developed teachers' guides with daily scripted lessons for all teachers (one for each of the three academic terms; 250–300 pages per term). Additionally, the program provided:

student books (at a 1:1 ratio, one for each term, including activities, decodable stories, etc.),

alphabet cards and alphabet charts,

teacher read-aloud books (big books),

termly assessments for teachers,

resource packets for teachers (during each termly training) with troubleshooting guides and remediation approaches,

e-learning courses on tablets for national core trainers and DTSTs, and

take-home books for grade 1 students.

All materials for the program were developed in-country (thus focusing on context-relevant stories and illustrations), and a language mapping study was carried out at the start of the program (prior to the 2016 redesign) for policy advocacy with regard to language policy and teacher deployment.

Teacher training

Learning used a cascade model for teacher training. The 120 national (core) trainers consisted of approximately 30–40 Learning staff, while the remaining trainers were ministry officials. The core trainers provided training to regional trainers, who trained district support teams. The DTST trainers (~840) all came from the Ghana Education Service, predominantly from district education offices. These district trainers ultimately were responsible for training more than 51,000 school-level staff (i.e., KG2 to grade 2 teachers, curriculum leads, and head teachers). Teachers were trained for 12 days at the start of the program, after which they received a two- to three-day refresher training every term (i.e., three trainings per year). Coaches were similarly provided with termly training.

All trainers were provided with a structured facilitator's guide. The program also supplied monitors at all training sites with daily reporting forms for documenting quality, absenteeism, venue issues, etc. Program monitors consisted of regional training and coaching coordinators, as well as FHI 360 staff.

Teacher support

Learning incorporated both in-school and external coaches into their implementation approach. Head teachers or curriculum leads served as in-school coaches, while DTSTs served as external coaches.

Head teachers and curriculum leads were given resource packs and were trained on specific skills and use of materials for teacher support. Curriculum leads tended to provide more coaching than head teachers. In-school coaching was designed to include a pre-observation with each teacher (to discuss issues/challenges), followed by a classroom observation and a post-observation discussion. These visits originally were intended to occur once per week but eventually were reduced to two times per month.

DTSTs were responsible for conducting two Learning-specific school visits per term. "Uptake visits" immediately followed training, in order to determine what participants had learned from the termly trainings (which were led by DTSTs). The second visit each term was focused on school-based coaching. DTSTs were provided with a coaching guide. During their visits, DTSTs were expected to conduct classroom observations and to lead discussions with teachers and head teachers.

In-school coaches were monitored by DTSTs, who were in turn overseen by regional program staff. Additionally, Circuit Supervisors (CSs; ministry staff responsible for visiting schools for monitoring and supervision) were monitored by Learning regional office staff.

The program also incorporated a school-based in-service education and training (SBI) component, in order to support teachers to build professional expertise among themselves. Trainings covered practice SBI meetings, and teachers were provided with SBI guides. Head teachers generally served as SBI meeting facilitators, but teachers, who were supported in the role, also had opportunities to lead. These meetings occurred two times per month (but not at regularly scheduled intervals, because they were not a part of formal school timetables).

Pedagogical approach

Learning used structured pedagogy, especially a clear scope and sequence of instruction; and explicit instructional strategies, grounded in the Simple View of Reading. The program emphasized phonics, with decoding and language comprehension seen as core instructional components. The lessons were fully scripted (i.e., everything the teacher said or did was in the teachers' guide) and they included regular instructional routines (i.e., 9–10 recurring routines).

The program also focused on increasing the amount of pair work for students, as well as promoting more student-centered approaches. Additionally, teachers were responsible for continuous assessment of learning and were provided with ASER-based assessments for their students. These assessments were part of a structured remediation program which began in Year 2 of implementation and which provided teachers with remediation activities to support struggling readers.

Systems

By design, the majority of Learning's activities were implemented by the Ministry of Education and the Ghana Education Service, with technical direction from the Learning team. This meant working through government systems, typically with oversight from program staff (particularly for components such as training and school support visits). The program aligned with government systems to provide in-school coaching and peer-learning approaches and to develop Ghanaian language content and program materials aligned with national curriculum and approved by the Ministry of Education. Data on fidelity of implementation (along with limited student outcome data) were collected by CSs (who were trained by the program and equipped with tablets). The program also issued 54,000 certificates to teachers through the National Teaching Council; in the final year of the program, the certifications fed into government records used for career development. The main exception to government implementation was the program's use of Ghanaian vendors from the private sector to oversee the distribution of materials to schools.

Learning provided national core trainers, CSs, and DTSTs with intensive capacity building in reading instruction, in order to support their roles as trainers, monitors, and school-support providers. The program also held monthly meetings with government counterparts at all levels (national, regional, and district), as well as quarterly meetings for all the regional and district directors, with the national-level representatives present. The regional and district directors were also equipped with tablets to use in visits to monitor schools, and they had access to an interactive data-visualization dashboard to review data on student and school reading performance. Capacity building of system actors—such as CSs and district-level officials—to carry out continuous monitoring and data-driven support of teachers was a major component of the program, aiming at full institutionalization of teacher support. It was a large-scale effort focused on collecting over 70,000 observations of learning outcomes in 11 languages twice a term, six times in an academic year, with follow-up data-driven discussions at each district and regional level.

Training activities for the program were funded primarily through grants to ministry bodies. The government contributed in-kind support to the program such as training venues, office space for Learning regional staff, and time for government trainers.

5.2.3 Tusome Early Grade Reading Activity – Kenya

Introduction

Launched in 2015, Tusome, derived from a Swahili word meaning "Let's read," is a six-year early grade reading program in Kenya funded by USAID and supplemented by DFID in the early years. The main goal of the program is to improve learning outcomes for over 7 million students in grades 1, 2, and 3 at national scale in Kenya (i.e., more than 22,000 public schools and 5,000 private schools across the country, as well as 1,500 alternative

basic education institutions in the urban informal settlements of Kenya; see USAID, 2020). The program, which is scheduled to end in 2021, is implemented by the Kenya MOE with technical support from RTI International (RTI International, n.d.).

The Tusome Early Grade Reading Activity grew out of the Primary Mathematics and Reading (PRIMR) Initiative, an evidence-based pilot program led by RTI International and funded by USAID and DFID, which ran from 2011 to 2014. PRIMR had meaningful impacts on learning outcomes in English, Kiswahili, and mathematics (Piper, Ralaingita et al., 2016; Piper, Zuilkowski, Dubeck et al., 2018). The PRIMR Initiative was structured as a set of RCTs to assess the effectiveness of the various early grade education interventions and their potential for a national scale-up (Piper, Zuilkowski, Kwayumba, & Oyanga, 2018). The primary interventions assessed were structured pedagogy instructional approaches (Piper & Zuilkowski, 2015) in addition to the use of a tablet-based coaching system to support policy and decision making (Piper, Zuilkowski et al., 2016).

Tusome is known for two main features: Its rigorous, evidence-based approach to program implementation (Piper, Zuilkowski, Dubeck et al., 2018); and its implementation at national scale with government leadership (Piper, Zuilkowski, Kwayumba, & Oyanga, 2018).

Evaluation

Tusome has had three phases of external evaluation. The baseline evaluation conducted in January 2016 and the midline evaluation conducted in 2017 (Freudenberger & Davis, 2017) were done by Management Systems International (MSI); the endline evaluation completed in 2019 was done by NORC at the University of Chicago.[18] As indicated in earlier sections of this report, the program evaluations had no control or comparison groups because Tusome is a national program. The effect sizes for Tusome between baseline and midline ranged from 0.40 to 1.07 SD for grade 1 and from 0.41 to 2.57 SD for grade 2, with average effect sizes around 0.7 SD for English and Kiswahili (Freudenberger & Davis, 2017). These effect sizes were higher than those of PRIMR.[19]

Program model

As described above, Tusome's design incorporates the elements that the earlier PRIMR pilot program showed would be effective at scale. The program has a four-pronged approach to improve learning outcomes in the target grades:

- Enhancing teachers' capacity to effectively deliver classroom instruction

---

[18] Tusome received three extensions between 2019 and 2021 that pushed the closeout date well past the original endline evaluation point in 2019. No additional external evaluation is planned.
[19] The endline results had not yet been finalized and released as of July 2021, but preliminary data showed smaller impacts for the Tusome program after the government decided to reduce the number of lessons per week for English and Kiswahili. We anticipate that the final Learning at Scale report will include the findings from the Tusome endline report.

- Improving schools' access to appropriate books and other learning materials
- Enhancing instructional support and supervision
- Enhancing collaboration with other literacy actors.

Tusome's design is structured to maximize the ability and skills of the MOE and its affiliated Semi-Autonomous Government Agencies (SAGAs) to sustainably improve learning outcomes, beginning during the life of Tusome and continuing beyond. **Figure 6** presents the theory of change.

Figure 6.   Tusome's theory of change
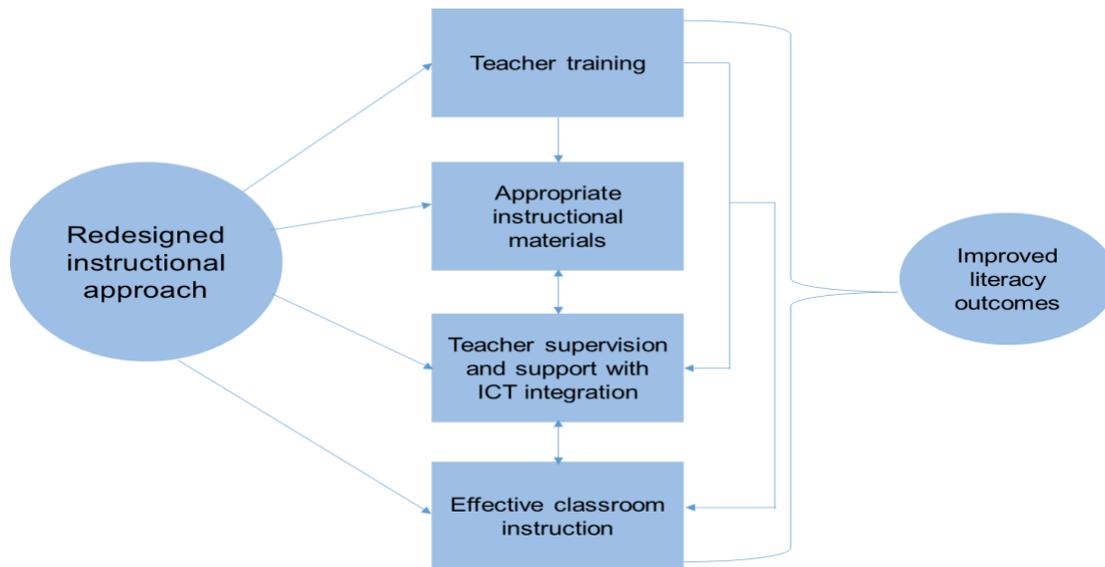
Source: Created by the authors.

Materials

Tusome developed teachers' guides with daily structured lessons for English and Kiswahili. These teachers' guides matched the student books provided for each student at a 1:1 ratio. For two years, the program also developed and distributed homework books that students used at home to undertake practice activities in the two subject areas. In addition to the core student textbooks and the related teachers' guides, Tusome developed a set of supplementary readers and leveled readers that are being distributed at a national level. The scale of Tusome's materials distribution has been substantial, with more than 26 million books distributed as of mid-2021; the external midterm evaluation showed that more than 90% of classrooms had textbooks at a 1:1 student-to-book ratio, with the rest of the classrooms having nearly 1:1 books (Freudenberger & Davis, 2017). The instructional materials for the teachers and the students were developed in alignment with the national curriculum of the country and were approved the national curriculum body, the Kenya Institute of Curriculum Development (KICD).

As noted above, English and Kiswahili instructional materials were developed for grades 1, 2, and 3. The program also facilitated capacity building for government officials in refining mathematics materials that had been developed by PRIMR and that now are being used for the Primary Education Development (PRIEDE) numeracy program funded by the World Bank and GPE.

<u>Teacher training</u>

The program uses a modified cascade model to manage teacher training at the national level (although in-person training halted completely during the COVID-19 disruptions of 2020). A cohort of master trainers, composed of high-level experts and key program staff, trains a larger group of facilitators, composed mainly of high-level government people, to lead national training-of-trainer sessions. These trainers then train the government's CSOs; through mid-2019, they also trained the program-hired Instructional Coaches who served the 1,500 private alternative basic education institutions. In the initial years of Tusome, the master trainers would fan out across the country to train the CSOs and coaches directly in order to minimize the levels of the cascade. The CSOs and coaches would, in turn, train teachers on the Tusome methodology in pairs in order to reduce the training ratios. Essential to the design of the program was an emphasis on modeling and practice at all levels of training in order to ensure that teachers left the teacher training with sufficient experience with the new Tusome methods. More recently, in line with the methods used by the PRIEDE mathematics program, Tusome has supported CSOs to identify accomplished "champion" teachers, who have demonstrated effectiveness in implementing Tusome, to assist the CSOs and Instructional Coaches in training the teachers instead of using the CSOs in pairs. This approach reduces the time that it takes to implement the national training. During each national-level training—for educators from both public and low-cost private schools—76,000 grades 1, 2, and 3 teachers and over 23,800 school head teachers receive training. Given the multiple rounds of training over Tusome's duration, nearly 140,000 individual teachers and head teachers had been trained in the methodology through the end of 2019. During the initial years of Tusome, the training rounds were five days for the first term followed by refresher trainings of two to three days in subsequent terms.[20] Trainings initially took place at the very beginning of the school terms and then moved to school holiday periods as the preferences of the government changed.

An essential part of the Tusome training is the training manuals. These are developed by the program staff working closely with their government counterparts from the National Technical Team that implements Tusome. The team works to ensure that the manuals include sufficient time for modeling and practice of each specific pedagogical skill in the program.

---

[20] The normal (non-pandemic) public school academic year in Kenya is divided into three terms: Term 1, January–April; Term 2, May–August; Term 3, September–November.

Although in-person schooling in Kenya was halted from March through December 2020 because of the pandemic, for public schools, the role of the CSOs normally is to provide continuous instructional support for teachers implementing Tusome; as noted above, for several years, the program employed Instructional Coaches to exercise a similar function for the targeted low-cost private schools. CSOs are in charge of a zone, with each zone comprising a cluster of approximately 20 public schools (Piper & Zuilkowski, 2015). Based on the findings from PRIMR, Tusome grouped the low-cost private schools and their coaches into similar geographically based cluster arrangements in order to provide consistent instructional support in classrooms (through April 2019 only). The role of the CSOs is to provide external coaching to the teachers within their schools. Each visit includes an observation of an entire lesson followed by a feedback session looking at particular elements of the observed lesson

Currently, each CSO is equipped with a digital tablet on which lesson plans, teachers' guides, and other coaching support materials are loaded. They also can access a coaching tool to support them in the process of observing a lesson taught in the classroom; it offers a scaffolded set of suggested feedback that can be given to the teacher after the lesson at the CSOs' discretion. Prior to the feedback session, the CSO reads with three students in each classroom. The observation data collected are uploaded to a cloud-based server daily, collated on a central platform, and used to populate a data-visualization dashboard. The MOE and program's senior management team use the dashboard to manage the program, to keep track of where CSOs are implementing effectively, and to assist in decision making in counties that are not implementing as expected. The live dashboard, whose updating, use, and maintenance were transitioned to the MOE during 2020 and early 2021, shows that in a typical month in 2015 to 2017, CSOs and Instructional Coaches observed more than 20,000 lessons and read with more than 60,000 students. The dashboard can be viewed here:
https://app.powerbi.com/view?r=eyJrIjoiY2NmZWJmNGYtZTgxOS00M2RhLWFjOWUtNjExND UwMmJmYzVkIiwidCI6IjJmZmMyZWRlLTRkNDQtNDk5NC04MDgyLTQ4NzM0MWZhNDNmYiIsI mMiOjF9

<u>Pedagogical approach</u>

The pedagogical approach utilized by Tusome is a bilingual balanced literacy program that includes an emphasis on the key areas recommended by a US-sponsored National Reading Panel (NICHD, 2000). The approach includes a focus on letters and their sounds, as well as vocabulary and comprehension. The content for the two languages (English and Kiswahili) is coordinated so that all letters and their sounds are taught first in Kiswahili and then in English, in order to support the transfer of literacy skills from one language to another more efficiently. English instruction begins with a focus on oral language development before

moving to the development of reading skills. The teachers' guide has fully scripted lessons at the beginning with the scripting tapering off.

Systems

Since the initial program design, the activities for government-sponsored schools have been implemented through the existing government systems, by government officials, with technical support from RTI International program staff. The program works with the MOE to involve government officials in all activities and by late 2020 had trained over 1,500 senior-level MOE staff on program coordination and various elements of program implementation. For example, county-level officers train and supervise the CSOs, who are themselves government staff.

The Tusome program is managed through the National Technical Team, similar to a program implementation unit, that is composed of education officers from across the MOE's various directorates, as well as key government bodies working on education. This National Technical Team works closely with Tusome program staff to implement the intervention at the national level. At the county level, the Tusome program has technical officers who work with the National Technical Team and the county-level officials to manage and implement the key activities of the intervention, and in many cases to use the activities of Tusome to support their broader work. These combined teams are supported to develop integrated annual work plans for each county; to monitor the implementation of the yearly work plans; to use the resulting data for decision making, including through the dashboard developed by the program and handed over to the Ministry; and to update and revise the plans each year. Essential to this process is embedding the Tusome activities into the broader work of county-level government. Tusome works with both MOE leaders and TSC leaders at the county level, which can be easier or more difficult based on the changing relationships between those portions of government and themselves as well as the Tusome team over time.

Tusome has handed over the copyright ownership of all instructional materials developed by the program, as well as digital infrastructure such as tablets and the dashboard program, as a critical step in strengthening government systems. The MOE took this ownership transition seriously and in 2019 used Government of Kenya funds to purchase and distribute revised versions of the Tusome materials for all grade 1 students in the country, in both English and Kiswahili. The work done with Tusome on this activity led the MOE to change how it managed procurements for the national CBC starting in 2019, resulting in significant savings but also 1:1 student- and teacher-to-book ratios for other subjects countrywide.

Finally, as mentioned above, the Kenya-based subcontractor to Tusome that supplied the Instructional Coaches and monitored and assisted the low-cost private schools concluded its work in April 2019. The National Council for Nomadic Education in Kenya, one of the MOE's

affiliated SAGAs, has been assigned responsibility for supporting low-cost private schools in Kenya going forward, including in implementing Tusome.

5.2.4 USAID/Senegal Lecture Pour Tous

Introduction

Lecture Pour Tous was a Ministry-led program funded by USAID, with technical and financial assistance implemented by Chemonics International. The five year program ran from 2016 to December, 2021. The program was designed to improve reading skills for students in grade 1 through grade 3 in Senegal, in six regions with direct support from the Chemonics International team, supporting government actors who serve as trainers and coaches, and a seventh region where it was implemented only by the government (with minimal technical assistance from Lecture Pour Tous). Lecture Pour Tous was implemented in three national languages (Pulaar, Seereer, and Wolof) and promoted research-based instruction, comprehensive and ongoing teacher support, increased instructional time, and community engagement. Additionally, it adopted student performance standards, ongoing assessment at multiple levels that inform iterative adjustments, and policy development to strengthen government systems to sustain the reforms Lecture Pour Tous achieves.

Evaluation

Lecture Pour Tous has been evaluated internally to understand reading growth in students, using a pre- and post-treatment design. No control or comparison group was available for Lecture Pour Tous due to full program coverage in each target region and an absence of appropriate comparison schools teaching in national languages. MSI also conducted an external evaluation at midline, but it did not address student reading performance. This evaluation was conducted in partnership with the government, as part of a capacity-strengthening activity of the program. The 2019 midline evaluation consisted of two samples (one for the original four implementation regions; one for the expanded regions that began implementation in the second year of the program). The midline results showed large increases in mean oral reading fluency scores, as well as in the proportion of students meeting ORF benchmarks. Overall, students in the program showed average gains in reading of 13–18 cwpm across languages in grade 2 (after two years of implementation). These gains were associated with reductions in zero scores, which ranged from 29 to 25 percentage points lower. There was an overall 29 percentage point increase in the proportion of students meeting the program's oral reading and reading comprehension benchmark, as well.

Program model

The Lecture Pour Tous theory of change is depicted in **Figure 7**. The program's theory of change is based on the "5 T's" (see figure) but also includes a parent and community component, which was implemented in a random sample of 20% of school communities

across six of the seven program regions. The parent and community model was shown to have a positive impact on student reading levels based on an RCT-based evaluation.

Figure 7. Lecture Pour Tous theory of change

THEORY OF CHANGE (BASED ON 5 T'S)

Tongue	Teaching in national language (L1) chosen with the school community (Pulaar, Seereer or Wolof)
Text	Provision of a variety of inclusive materials (curricular, supplementary, extra-curricular) for all with a high volume of text
Teaching	Training, tools and continuous professional development (CPD) across multiple channels to match the teacher to the local language; explicit structured, phonics approach
Time	Official reading schedule in L1 of 60mn / day + monitoring of the presence of teachers; + promotion of reading at home
Test	Evaluation framework, continuous supervision and monitoring at the classroom level, departmental level management of data
Parents/ Community	SBCC, tools and training for family EGR support, school-community grants for EGR activities

Improved reading skills

Source: Lecture Pour Tous program documentation

<u>Materials</u>

Traditionally in Senegal, teachers have to design their own lesson plans. Additionally, while the government does purchase all student textbooks, it has not been possible to ensure a 1x1 ratio, and if students lose books families have to reimburse. Lecture Pour Tous supplies teachers with a teachers' guide for the language of instruction that includes scripted lesson plans prescribing what and how to teach the content that is in the student textbook. The guides use French for all of the teacher instructions, while the directions that teachers read aloud to students are in one of the three national languages served by the program. Teachers also receive a book of stories (with vocabulary and discussion questions) to use as class read-alouds. Lecture Pour Tous gives students a textbook, a take-home book that is linked to the textbook and is designed for homework practice, as well as a home–school communication tool that uses visuals, which students take home regularly to communicate

expectations to the family. Supplemental leveled, decodable readers were made available to students for use during both scheduled class time as part of reading lessons and afterhours through classroom "reading corner" libraries. The more advanced readers include comprehension questions.

Teacher training

Lecture Pour Tous's teacher professional development consists of a mix of approaches to reinforce teachers' knowledge and give them multiple opportunities to learn and practice new teaching strategies. The training uses a cascade approach, with an emphasis at all levels on building the capacity of teachers, directors, and inspectors. Prior to the initial training, program staff and key government departments determined the organizational structure of the master training team (including roles, responsibilities, and coordination of activities). At the first level of the cascade, the master team trains the regional team (composed of inspectors and experienced school directors), which in turn trains the teachers, for the second level. During the second year of the program, when implementation began at the school level with half of the schools in four regions, teachers received 10 days of training via two in-person trainings (a seven-day training and a three-day training). In the third year of the program, the remaining schools in those four regions were added and their teachers received training.[21] The trainings have used practice and participation via simulations and videos (beginning in Year 3) to demonstrate the main Lecture Pour Tous reading methodology. At all levels of training, participants practice teaching methods using demonstration and simulations. This approach was continued for the first 2½ years of implementation at the school level, after which the model consisted of shorter, localized trainings combined with distance learning modules (and additional teacher supports), resulting from planned program design changes and the impact of COVID-19.

Additionally, the program supported a pre-service training component focused on assisting regional training institutes to develop and pilot a new teacher competency framework, as well as course modules for early grade reading and teaching in national languages.

Teacher support

Following training, teachers receive ongoing support at multiple levels. Teachers are expected to receive two coaching sessions per month; both government inspectors and school directors act as coaches and provide support, solutions, and feedback to teachers to complete the sessions. The support format has evolved since the start of the program. At one point a pilot was conducted of coaches using digital versions of the coaching and supervision tools, which had them note what they saw in the classroom. The coaches also assess five students during each visit. In addition to coaching visits and internal support,

---

[21] Teachers in Senegal generally follow cohorts of students through primary grades. Therefore, there was a new cohort of grade 1 teachers for each year of the program and teachers needed to learn the program and materials for their new grade each year.

there are both school-based teacher learning circles run by school directors and cluster-based learning circles with small groups of schools. These meetings are decentralized, so their frequency varies by location. Phones serve two roles to provide ongoing support. The program sends primarily one-way messages of support, while also including chat groups and invitations for teachers to engage in two-way communication using the Telerivet platform to complete surveys and questionnaires via short message service (SMS). Teachers are also provided subscriber identification module (SIM) cards for their mobile phones to enable peer-to-peer support via WhatsApp. The communication technology has been introduced iteratively through the years of the program. Due to challenges with the frequency with which school directors and inspectors were providing in-class coaching, the program also conducted a small randomized control trial of two additional coaching variants (distance coaching conducted by an experienced inspector over the phone versus in-person coaching by nearby experienced school directors).

Pedagogical approach

The elements of Lecture Pour Tous's reading program are grounded in the science of what works to promote beginning reading skills. Lecture Pour Tous provides instruction that develops both decoding and oral language skills to support later reading comprehension. The content taught between grades 1 and 3 is carefully planned and gradually gets more advanced to ensure mastery. In practice, prereading and letter knowledge are developed before students learn to read single-syllable words, after which they move to two-syllable words. The student textbooks repeat items with different text of the same level on a single page to promote a lot of practice, which supports accuracy. The core method is for the teacher to model an activity, then do the activity with the students, and finally monitor the activity as students do it independently in front of the whole class. This method is often referred to as "I do, we do, you do" (with the majority of time in program classrooms spent on the "you do" stage). Across all grades, students have opportunities to read, write, and discuss. The teachers' guides provide a daily lesson plan that uses a structured set of steps describing each activity and lesson scripts to guide teachers on exactly what to say to engage students in listening comprehension, reading, writing, and discussions.

Systems

Technical assistants of the Lecture Pour Tous program work with MOE at all levels of government: national, regional, and department. This collaboration begins with work planning and coordination across 10 MOE partners, with the three departments that manage elementary education, training, and evaluation having the most involvement. The tone of the collaboration is solution oriented. The assistants helped the MOE to develop reading standards that are understood and recognized at all levels (e.g., the standards appear at the front of the teachers' guide, and progress toward these standards is shared in community meetings), later updated to be policy linked to both national frameworks and the

Global Proficiency Framework for literacy, for reporting on SDG 4. To monitor their achievement, the program is helping to integrate reading assessment into existing standardized quarterly assessments, while supporting department education offices to conduct a Local Education Monitoring Approach (LEMA) with student performance measures, once per year.[22] Across the program, student-level data are shared widely, including at community dissemination events in each region that are attended by representatives from all levels of government. Lecture Pour Tous systems strengthening work also includes support to action-research and resulting research-based policy, such as a recent study on the oral vocabulary levels of early grade students in French, used to inform the upcoming policy on bilingual education.

5.2.5 Northern Education Initiative Plus (NEI+) – Nigeria

Introduction

The NEI+ activity started in 2015 with funding from USAID. The program initially was slated to end in 2020 but was extended until May 2021. The goals of the program were to strengthen the ability of Bauchi and Sokoto States in Northern Nigeria to improve the reading outcomes of over 1 million school-aged children and to increase access to basic education for 400,000 children who were out of school. The program was implemented by Creative Associates International, with Education Development Center, Florida State University, Overseas Strategic Consulting, and Value Minds as partners. Although NEI+ worked in both formal and nonformal schools, the Learning at Scale study was interested only in the activities implemented in formal schools. Over the life of the program (through closeout in 2021), the program enrolled over 269,000 students in nonformal learning centers and reached over 950,000 students in formal schools. It targeted students in about 50% of the local government education authorities (LGEAs) in each state. In target LGEAs, all schools were covered. The Initiative focused on students in grades 1–3 in formal schools, teaching reading in Hausa language from grade 1 with a transition-to-English program starting in grade 2.

Evaluation

NEI+ conducted two evaluations of the reading skills of students in Bauchi and Sokoto: a baseline EGRA conducted in 2015, before the program started its assistance; and a midline assessment in 2017 (the third year). This internal evaluation used a pre- and post-intervention assessment design, with no control group. At midline, grade 2 students were assessed in Hausa and grade 3 students in Hausa and English. Comparison of the baseline and midline results showed that on five of six EGRA subtasks, grade 2 students assessed during the midline performed better than the similar group of students assessed at baseline.

---

[22] LEMA: A monitoring approach for education that uses small sample sizes (such as schools or districts) and binary indicators to signal whether a sample meets or does not meet minimum performance standards.

The midline results also showed statistically significant reductions in zero scores on all Hausa EGRA subtasks, in both states. The improvements for students in grade 3 in Hausa were even greater, both in mean scores and in reduction of zero scores. Overall, the program produced effect size gains ranging from 0.2 to 0.7 SD in oral reading fluency across grades 2 and 3 (equating to gains of approximately 2 to 13 cwpm). The midline results also showed improvements on some EGRA subtasks in English, but the results were more modest.

Program model

Figure 8 outlines the program's theory of change. The results framework represents the NEI+ program strategy. The program's underlying developmental hypothesis was that through the strengthening of local education management systems, in conjunction with extensive community-based outreach activities, education access and quality for all children—including vulnerable children—would be improved and would result in significant improvements in reading outcomes for primary grade learners.

Figure 8. Theory of change for NEI+, Nigeria

Project Goal: To improve, equitably and sustainably, the quality of, and access to, education in Northern Nigeria

Intermediate Result 1: Government systems strengthened to increase the number of students enrolled in appropriate, relevant, and approved educational options, especially for girls and out-of-school children in target locations.

Sub IR 1.1. Increased number of educational options (formal, non-formal, NFLC) meeting school quality and safety benchmarks

Sub IR 1.2. Strengthened systematic approach to school managment and supervision

Sub IR 1.3. Institutionalized model of Non-Formal Learning Centers (NFLC) is used to ensure education for vulnerable children and youth

Sub IR 1.4. Core curriculum is adopted in non-formal schools

Sub IR 1.5. Strengthened CSO capacity to mobilize PTA/SMBC and communities around reading and access

Intermediate Result 2: Government systems strengthened to improve reading outcomes for primary grade learners in target locations

Sub IR 2.1. State and LGEA policies, timetables, and standards for reading instruction and performance improved and implemented

Sub IR 2.2. State and LGEA systems for the development, approval, and distribution of decodable readers, teachers' guides, and supplementary materials for early grade reading instruction improved and implemented

Sub IR 2.3. State and LGEA systems for providing in-service training to teachers in NFLC classrooms in the use of the evidence-based reading materials improved and implemented

Sub IR 2.4. LGEA systems for monitoring and coaching in-service teachers in early grade reading instruction improved and implemented

Sub IR 2.5. State and LGEA systems for early grade reading assessment improved and implemented.

Sub IR 2.6. State and LGEA systems for extending evidence-based reading instruction to non-traditional, non-formal schools improved and implemented.

Sub IR 2.7. State and local governments' accountability towards the public regarding reading instruction increased.

Source: Creative Associates International (2018), p. 31. IR = Intermediate Result; LGEA = local government education authority; NFLC = nonformal learning center; PTA/SMBC = parent–teacher association/school-based management committee .

Materials

NEI+ provided student books in Hausa (titled *Mu Karanta*) and in English (titled *Let's Read*), as well as teachers' guides with scripted lessons. Books were intended to be provided at a 1:1 ratio for students, although the midline external evaluation found that in every observed classroom, some students were sharing books. This shortfall was due to delays in book distribution, which were rectified soon after. The final external evaluation (in draft as of mid-2021) reported that the intended 1:1 ratio was restored. The student books provided space for students to write, draw, and conduct a mini self-assessment. In 2018, NEI+ also selected 32 Hausa and 16 English titles of supplementary readers, and these were printed and distributed to schools in 2019.

Teacher training

NEI+ used a cascade model to train teachers in targeted schools. Project staff trained master trainers, who then trained teacher trainers, who in turn trained the teachers. The master trainers and teacher trainers consisted of ministry officials from the state and federal levels and faculty from teacher training colleges. Teachers were trained on several topics, including reading instruction, continuous assessment, design and use of teacher aids, creation of safe learning environments, and gender-sensitive pedagogy. Teachers received between two and three rounds of training each year for a total of 29 days over five years (8 days in 2015; 12 days in 2016; 5 days in 2018; 4 days in 2019). In addition to in-service training of teachers, NEI+ taught a pre-service early grade reading course at three teacher training institutions in Bauchi and one in Sokoto.

Teacher support

NEI+ provided ongoing support for teachers through external coaches and professional learning circles. School Support Officers (SSOs), who are education officials within the LGEAs, handled the external coaching. Project staff trained the SSOs and supplied the tools to do their coaching observations. Some SSOs were given tablet-based tools for their classroom observations. The SSOs oversaw between 10 and 12 schools, grouped into a cluster. They were supported by Reading Coordinators, who were program staff; and by Principal Quality Assurance Officers, who are employed by the State Universal Basic Education Board. The Reading Coordinators and Principal Quality Assurance Officers mentored two SSOs each, conducted joint coaching visits with them each month, and guided them in their coaching.

In addition to supervising coaching by SSOs, NEI+ promoted teacher learning circles at individual schools as well as cluster learning circles that served groups of schools. These were opportunities for teachers to come together and learn from each other. NEI+ developed a *Teacher Learning Circles Guide* and trained teachers, head teachers, and SSOs on the guide, which created a structure for the learning circles.

<u>Pedagogical approach</u>

The program and materials used a phonics-based approach to teaching that was student centered. They relied on the "I do, we do, you do" gradual-release methodology, group/pair work, modeling and demonstration, monitoring, and assessment. Each teacher had a teachers' guide that was quite scripted, with timed instructional activities for each day of the week that included activities focusing on phonemic awareness, letter-sound recognition, vocabulary, fluency, and comprehension. As indicated above, literacy instruction in grade 1 in NEI+ schools was in Hausa, with a transition-to-English program in grades 2 and 3. The guide had four lessons per week and one lesson for review and catch-up.

Systems

NEI+ worked very closely with government systems to both strengthen them and improve students' reading. The main counterparts at the state level were the State Universal Basic Education Board, the State Ministry of Education, and the LGEA. At the federal level, NEI+ worked with the National Commission for Colleges of Education on pre-service teacher training and with the Nigerian Education Research and Development Council on curriculum and materials development.

At the state level, the program participated and supported several technical working groups—such as one on reading and another for monitoring, evaluation, assessment, and research—which made many of the technical decisions regarding NEI+. The government and NEI+ did joint work planning and joint funding of certain activities. NEI+ provided capacity building and support to government staff for the implementation of the program activities—teacher training, coaching, assessments, etc.

NEI+ also worked with the two state governments on policies related to reading. The Initiative succeeded in increasing the time available for reading, not just in targeted schools, but in all schools in both states. The program also had in place a memorandum of understanding with the government that called for joint financing of certain activities; the state governments used their own budgets for printing textbooks, training teachers outside of NEI+ schools, and conducting assessments such as the EGRA and the LEMA. NEI+ conducted a baseline and a midline assessment of the institutional capacity of state education agencies and the LGEAs in domains such as financial management, teacher management, data management, and policy and strategic planning. The results of these assessments informed the program's capacity-development strategy.

5.2.6 Pakistan Reading Project

Introduction

PRP was created by USAID and the Government of Pakistan in 2013 and was implemented by a consortium led by the International Rescue Committee until the program's end in 2020. The goal of PRP was to support provincial and regional departments of education in

Pakistan, in an effort to improve the quality of education for grade 1 and 2 students throughout the country. During the seven years of program implementation, PRP benefited approximately 1.7 million students and over 27,000 teachers across 69 districts in the provinces and regions of Balochistan, Khyber Pakhtunkhwa (KP), Gilgit-Baltistan (GB), Sindh, Azad Jammu and Kashmir (AJK), Islamabad Capital Territory (ICT), and the Newly Merged Districts (NMD; formerly known as Federally Administered Tribal Areas, or FATA). It also carried out pre-service work with 110 universities and teacher training institutes (TTIs) across the country (note: Punjab was included only in the pre-service work). At its peak, PRP had more than 1,100 employees.

Based on a 2010 amendment to the national Constitution, nearly all responsibilities for the education system in Pakistan were decentralized to the provincial level. Each province has its own curriculum, textbooks, TTIs, etc., with the exception of GB and NMD, which do not have their own curriculum or textbook boards. NMD follows KP's curriculum, and GB follows the federal or Punjab boards. Accordingly, despite the near national coverage of PRP, the program was often seen as having to run seven programs simultaneously. Additionally, the governance structure of the program was unique, with multiple partners, each leading program implementation in different provinces (i.e., IRC as the prime and lead implementer in ICT and KP, World Learning in AJK and GB, Creative Associates in Sindh, and the Institute of Rural Management in Balochistan and NMD).

The program's initial design concentrated on English and Urdu for grades 1–5 students. However, a contract modification in the early stages changed to a local language-of-instruction approach (Urdu and Sindhi) for grades 1 and 2. Accordingly, materials were developed in seven languages (Urdu, Sindhi [two dialects, in Sindh and Balochistan], Pashto [two dialects, in KP and Balochistan], and Balochi and Brahuvi [in Balochistan]). However, the program was primarily delivered in Urdu and Sindhi, while other language programs consisted of small-scale pilots. The program used a cohort-based, staggered implementation approach, with each phase lasting approximately two academic years.

Evaluation

Evidence for the impact of the PRP on student performance came from two separate evaluations—one internal and one external. PRP's internal evaluation consisted of a quasi-experimental approach that compared baseline (2016), midline (2017), and endline (2018) EGRA results of students in Cohorts 1 and 2 for Urdu- and Sindhi-speaking schools, with results from students in Cohort 3 schools (which had not yet received the intervention). Results showed that PRP interventions had increased students' reading ability in both languages among the treatment groups.

Based on the internal evaluation, difference-in-differences analyses indicated that PRP had a positive and statistically significant effect on the oral reading fluency scores of second graders.

Specifically, PRP led grade 2 students to read, on average, 8.43 more cwpm than students in the control group. These gains equated to a medium effect size of approximately 0.27 SD (with effect sizes as large as 0.61 SD in the top-performing province). Lastly, there was a 10.7 percentage point reduction in the proportion of Cohort 2 second graders who obtained zero scores in oral reading, as compared with a 4.4 percentage point drop among second graders in the control group.

The external evaluation (led by MSI) was conducted separately for each province and did not produce any program-level estimates of impact. The design was pre- and post-intervention only, examining changes in performance from the intervention group from baseline in 2013 to endline in 2017. Results showed that students participating in PRP school-level interventions performed better on the midline EGRA than students in "light treatment" who were able to benefit from PRP policy and advocacy activities but who did not receive PRP's in-school interventions. However, the evaluation found a wide range of gains across provinces, from increases as small as 2.9 cwpm to as large as 26.9 cwpm, as shown in **Table 32**.

Table 32.   External evaluation: Oral reading fluency gain scores for grade 3 students, by province

Province	Language	cwpm, by time point		Gain score (cwpm)
		Baseline	Midline	
Azad Jammu Kashmir	Urdu	28.1	42.1	14.0
Balochistan	Urdu	28.3	35.0	6.7
Gilgit-Baltistan	Urdu	26.4	29.3	2.9
Islamabad Capital Territory	Urdu	18.0	24.1	6.1
Khyber Pakhtunkhwa	Urdu	31.8	54.6	22.8
Sindh	Urdu	30.3	57.2	26.9
Sindh	Sindhi	32.0	45.0	13.0

Source of data: 2017 Early Grade Reading Assessment reports for Azad Jammu and Kashmir, Balochistan, Gilgit-Baltistan, Islamabad Capital Territory, Khyber Pakhtunkhwa, and Sindh (six separate reports).

PRP also conducted two additional internal evaluations: (1) an RCT on the impact of various ingredients of professional development on teachers' instructional practices and student learning outcomes (which found coaching to be the most cost-effective professional development ingredient)[23]; and (2) an RCT examining the impact of providing classroom libraries.

---

[23] PRP study on cost-effectiveness of professional development ingredients can be found here: https://rescue.app.box.com/s/7buuxo9l2a4nuqstq0kl6mwklwigk64e

Program model

Broadly speaking, PRP had three main components that were central to the program's theory of change (see **Figure 9**):

1. Educators and school environment: improved environment for reading in classrooms (supplementary reading materials, in-service CPD, pre-service TPD, and use of technology for CPD);

2. Policies and systems: improved policies and systems for reading; and

3. Communities: improved community-based support for reading.

Figure 9.   Pakistan Reading Project: Theory of change
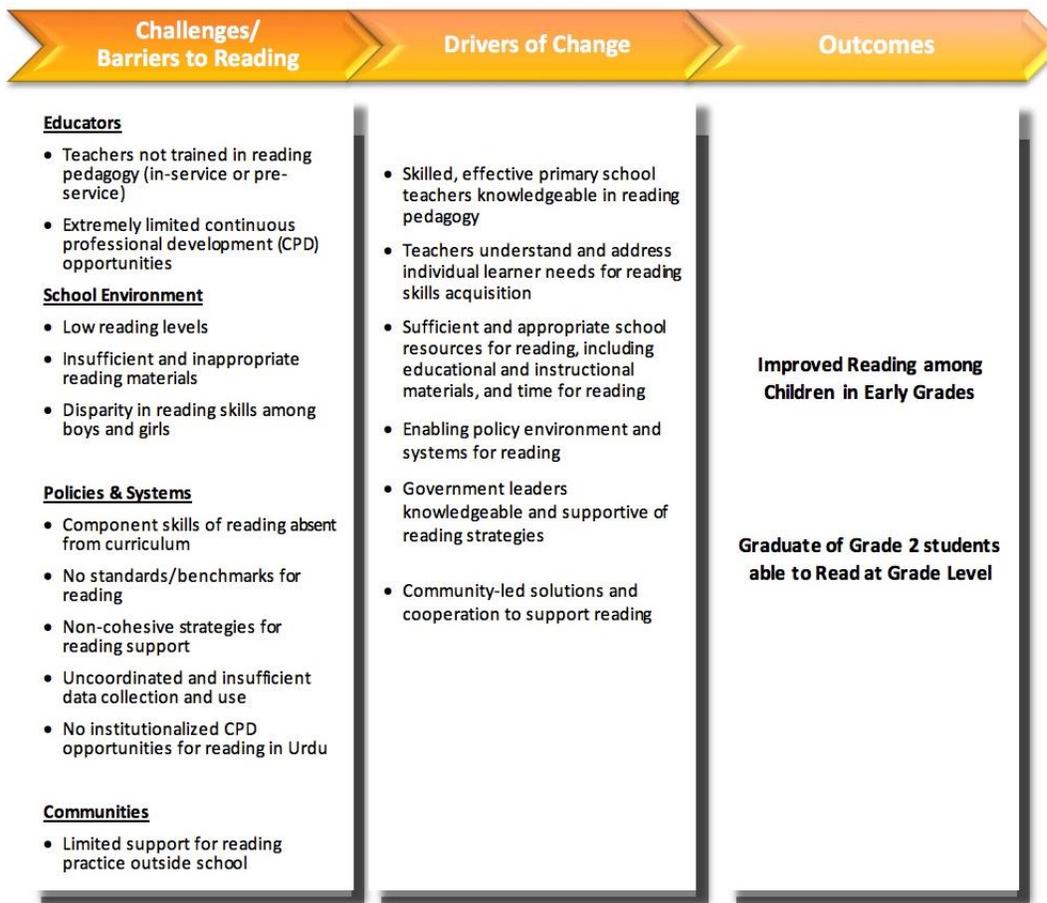
Source: International Rescue Committee (unpublished PRP documents).

Materials

In consultation with the Government of Pakistan, PRP created a wide range of "reading learning materials" for students (in a total of seven languages—five languages and two dialects), as follows: (1) daily reading lesson plans, (2) workbooks, (3) teacher read-alouds or big books, (4) flash cards, (5) syllable charts (for grade 1), (6) leveled readers (for grade 2), and (7) developmentally appropriate storybooks for classroom-based corner libraries.

Teachers were provided with fully loaded digital tablets that included structured lesson plans, as well as audio lessons for Urdu and Sindhi. These tablets also allowed for online and offline content delivery and access. Prior to the program, few supplementary materials existed, textbooks did not follow a components-of-reading approach, no Urdu leveled readers existed in the country, and languages of instruction were only Urdu and Sindhi in public schools.

Teacher training

The teacher training approach under PRP followed a cascade model, including both short- and long-term training. PRP first trained provincial program staff (at a centralized training). Provincial staff then trained district staff—School Support Associates (SSAs)—who in turn trained mentors. Mentors and SSAs were then responsible for training teachers. Trainers at all levels benefited from the same structured training materials and training preparation.

Teachers received an initial five-day face-to-face training at the start of their cohort's intervention. This orientation training encompassed an introduction to PRP's approach, the program support system, reading assessment tools, and grade 1 materials. At the start of the second year, teachers received a three-day face-to-face refresher training, during which they were introduced to grade 2 materials. Additionally, head teachers received three days of training in the first year of the program for Cohorts 1 and 2, and three additional days of training in for Cohort 3. PRP also trained school support staff and mentors with six or seven days of training per year.

Lastly, PRP worked with the Higher Education Commission and faculty from pre-service TTIs to develop five reading integration courses and five reading specialization courses, along with orientation sessions for TTI faculty on these new courses. Through this pre-service work, PRP awarded scholarships to 3,161 honors students who were working on an associate degree or a bachelor's degree in education, and trained 744 faculty from 110 TTIs across Pakistan to implement these reading courses.

Teacher support

PRP's ongoing teacher support was built around two main components: school support visits and teacher inquiry groups (TIGs). Monthly TIG meetings typically included approximately 6 to 10 teachers per group and were designed to impart long-term peer-to-peer learning to deepen teachers' understanding of teaching reading.

Teachers were expected to receive five to six school support visits each year from trained PRP staff serving as SSAs (also known as coaches), as well as mentors (who were government staff, such as senior teachers and supervisors). During these visits, teachers received on-the-job support and guidance. The visits were structured for pre-, during-, and post-visit protocols, using a specially designed tool, focusing on the use of PRP materials,

the use of skills learned in training, and the administration of classroom assessments (with a few students being assessed as a part of each visit).

Lastly, teachers gained access to virtual mentoring videos and audio lessons through program-purchased tablets, and through frequent reinforcement text messages sent via SMS.

<u>Pedagogical approach</u>

The PRP model for teaching reading centered on teachers' use of guided lesson plans (with scripted lessons), using a components-of-reading approach to instruction. Teachers were also trained to use more participatory approaches in the classroom that allowed students to engage and to direct activities, and to regularly employ supplemental resources (e.g., alphabet charts, objects, visual aids). PRP teachers used modeling techniques for teaching, while the program focused on ensuring that teachers continuously monitored student learning, provided meaningful feedback to students, and adjusted instruction according to student needs (e.g., through the use of questioning, checklists, and formative and summative assessments).

Systems

As noted above, the second component of PRP focused on improved policies and systems for reading. PRP began with a needs assessment and gap analysis to identify key policies on which to focus. Ultimately, PRP supported provincial and regional governments in revising the government "scheme of studies" to include reading instruction time (35–45 minutes for reading each day), revising the language curriculum to reflect updated reading standards, integrating PRP reading materials into new government language textbooks, integrating components of the PRP teacher training model into the government's CPD model for teacher training, and developing reading test item banks to be made available for government assessment bodies and teachers.

The program ensured adherence to public sector governance structures and focused on sustainability and government capacity building at all levels. Government staff led teacher trainings, TIGs, and school support visits. In addition to training personnel for these roles, PRP built the capacity of government staff for development and administration of EGRAs, as well as content development for curriculum, standards, and materials. However, the program ultimately was responsible for implementation oversight, including PRP's development of a monitoring and evaluation system (which consisted of data that were monitored by provincial and district-level program staff but were shared regularly with government counterparts).

5.2.7 Read India Program

Introduction

Read India is Pratham's flagship program, aimed at improving students' basic literacy and numeracy skills. Using the Teaching-at-the-Right-Level (TaRL) approach, the program is delivered in one of two ways: (1) directly by trained Pratham instructors or (2) in partnership with state and/or district governments and through government teachers. The focus of this report is the Read India program implemented in partnership with the State Government of Karnataka—a program known as "Odu Karnataka" ("Read Karnataka" in the Kannada language). Through Odu Karnataka, Pratham has implemented this partnership model of TaRL since 2016 and by 2020 had reached 22,173 schools in 20 districts, serving 564,166 students.

Evaluation

Students in all Pratham TaRL programs are assessed using instruments similar to ASER assessment tools. This assessment requires students to demonstrate reading proficiency at several levels: letter, word, paragraph (in a grade 1 textbook or equivalent level text), and story (in a grade 2 textbook or equivalent level text). The aim of the assessment is to determine the highest level at which the student is able to demonstrate reading proficiency.

The Read India (government partnership model) program has participated in one impact evaluation—in Haryana in 2015 (Banerjee et al., 2016). At baseline, 34% of students in grades 3 to 5 were able to read a paragraph or a story. After the 60-day program, 47.6% of students could read at this level in the control group, and 53.1% in the intervention group, a 5.5 percentage point increase in the number of students succeeding at paragraph or story reading attributable to the program.

Since 2015, the Pratham partnership model has evolved, and the Haryana evaluation is no longer representative of its current impact. No subsequent evaluations have been conducted of the government partnership model, but assessment data from the Odu Karnataka program can be used to estimate progress, and data from the annual ASER survey can serve for estimating the counterfactual.

Data from the Odu Karnataka program in 2019–2020 (Pratham, 2020), collected from 15,784 schools where both baseline and endline data were available, revealed significant improvement in students' reading skills:

For grade 4, 25% of students could read a grade 2 level story at baseline. After a 60-day intervention, 58% of students could read at this level, a 33 percentage point improvement.

For grade 5, 32% of students could read a grade 2 level story at baseline. After a 60-day intervention, 63% of students could read at this level, a 31 percentage point improvement.

In addition, similar results were achieved in previous years but in fewer schools.

Odu Karnataka data were compared with those from regular government schools (data analyzed for this report came from ASER Centre/Pratham, drawing on ASER data from 2012 to 2018). The average percentage point increases in the proportions of students reading a grade 2 level story were 13.2% in grade 4 (range 8.5% to 16.9% across the years) and 12.3% in grade 5 (range 7.9% to 16% across the years). Thus, the proportion of students who learned to read at the story level during a 60-day Odu Karnataka program was between two and three times the mean proportion who learned to read at that level in a typical year of government primary school.

Program model

The key to all Pratham TaRL programs is the organization of students into groups based on learning levels determined by the ASER assessment. In the Read India model, students from grades 3–5 are first combined by assigned grade, then redivided into three groups by achievement level for the purposes of the program. The Odu Karnataka program combined only grades 4 and 5 because of a separate state program (known as Nali Kali) running in grades 1–3.

Students in Read India programs typically take part in an intensive course of literacy and numeracy instruction targeted at their level, using engaging activities. The program in Karnataka took two hours of the regular school day, typically taking the place of the mathematics or Kannada lessons in the timetable. The Odu Karnataka program intervention was conducted over 60 consecutive teaching school days.

Assessments took place at baseline, midline (30 days), and endline. The assessment at midline was used to reconfigure achievement groups to allow students making faster progress to move up a group. These assessments were conducted by the government school's teachers, and data were submitted to the cluster resource coordinator (responsible for 10–15 schools). The cluster resource coordinator, in turn, submitted the data to the block level (i.e., 100–150 schools), where an official would enter the data into a web-based portal developed by Pratham. Each teacher had a Learning Progress Sheet on which to mark the level of the student at each assessment cycle (baseline, midline, endline), by which they could track the progress of each student across assessment cycles as well as grouping students at similar learning levels after each assessment cycle.

Materials

The materials provided by the Read India program differed from state to state. Under Odu Karnataka, all students had chalk and access to a board on which to write and draw to express themselves. Each class was also provided with letter cards; syllable charts; pictures for story-making activities; and short, locally relevant, engaging stories. Teachers were provided with guides. The materials contrasted with government textbooks, which were typically beyond the students' ability and less engaging.

Teacher training

Training began from the top down. First, the District Resource Group participated in three to five days of hands-on training run by Pratham. Each person attending the Pratham training then ran practice classes for 15–20 days, with baseline and endline assessments. The practice classes are a critical and unique part of Pratham's approach; this practical experience reinforced what the trainees learned in their professional development and enabled them to become master trainers. At the next level, block and cluster resource coordinators went through a similar process of professional development followed by 15–20 days' practice classes. The training for this next level was led by master trainers and supported by Pratham. Each cluster resource coordinator then trained teachers in four- to six-day sessions. This final stage of the training was the first to be run entirely by government staff. Pratham staff were present when available, but played a limited role.

Trainings were designed to be "lean" so that they could be delivered by just one person. Training covered three main areas: (1) goals of the program and assessment of progress toward them; (2) grouping of students, teaching methodology, and materials, with separate sessions for mathematics and language; and (3) assessment of progress and review of approach (a different training for teachers and mentors). Teachers received refresher training of four to five days annually. For new partnerships, the refresher training reinforced the basics. For mature partnerships, subsequent training could add new elements to the instructional model. Mentors received refresher training of two days per year. State governments have designated budgets for teacher training, which covered the cost of the Odu Karnataka training.

Teacher support

Cluster resource coordinators (coaches) supported 10–15 teachers each with three or four visits over the 60-day program cycle. Coaches helped teachers diagnose and address problems. To perform this function, it was critical that coaches had experience delivering the program themselves. Coaches observed the implementation of the program—for example, by checking whether the class was running, group activities took place, the lowest-achieving group was making progress, teaching and learning materials were being used appropriately, and the right activities were taking place. Coaches also fulfilled a monitoring function, completing a four-item binary observation checklist. A guidebook was provided at training sessions for coaches to follow. There were challenges in using the data to identify areas for support, however. Pratham staff and the coordinators visited schools together to help develop the capacity for data-driven support.

In addition to their Odu Karnataka program duties, cluster resource coordinators had to conduct administrative work for the government, including monitoring sanitation and the midday meal program, and collecting a lot of administrative data.

Block resource coordinators met monthly with all the cluster resource coordinators in their block to provide support. The meetings were timed to allow for review of midline and endline data, although in practice, data frequently were not reviewed during block meetings as planned.

<u>Pedagogical approach</u>

Again, the first step in the TaRL approach is reorganizing students from grade levels into groups with similar learning levels. The details of the reorganization depend on the number of teachers in the school and are ultimately the decision of the head teacher. In Karnataka, the ideal practice determined was to divide grades 4 and 5 into two groups, based on learning level, each run by a separate teacher. If only one teacher was available, learning groups were formed within the classroom. Pratham recommended no more than 50 students in a class, with the ideal number being 30–35 students. If classes were large, head teachers had the option of hiring a parateacher—an assistant teacher to help with instruction. In addition, Pratham offered teachers special pedagogical approaches to use with large classes.

The pedagogical focus was on improving foundational skills in reading and arithmetic. Instruction was focused on children unable to read at a grade 1 level and used a child-friendly approach called Combined Activities for Maximized Learning (CAMaL).[24] The approach involves listening, speaking, observing, reading, writing, and playing games, with the teacher serving as facilitator rather than instructor. The organization of the classroom followed a similar pattern in each lesson. The class began with whole-group activities that engaged all students, then students split off into their groups for activities based on the groups' learning level. The role of the instructor was to move around the groups and make sure everyone could do the activity. Finally, students did individual activities to ensure that each understood the lesson.

Lessons were not scripted. Teachers were given guidelines on how to create session plans, such as the progression of classroom organization and the aim to maximize time spent on activities. Within these guidelines, teachers decided the objective of the lesson and the type and duration of activities it would include.

Systems

Pratham's model for engaging the state education system has evolved over the years. To ensure government ownership, the program in each district begins with the resource group and cascades down through the system from there. Staff, from district down to teachers, are trained in the program and have experience implementing it. The system is usually supported by two or three Pratham staff in each district, depending on the size of the Pratham team in the state.

---

[24] CAMaL is used interchangeably with TaRL in this context.

In some cases, the partnership is initiated by the state government and can lead to rapid statewide implementation. This was the case in Uttar Pradesh in the Graded Learning Program in 2018–2019. In Karnataka, the partnership was initiated by the government, but they first wanted to understand the program logistics and impact before scaling up. Therefore, the program in Karnataka began with three pilot districts and scaled up in a gradual manner from 2016 to 2020. In either case, the program, with the exception of a small number of Pratham staff in each district, is funded by the government.

Pratham does not directly target change in central government policy through the Read India program. The central government is responsible for policy and planning of education, but states have much autonomy in implementation. Read India is implemented at the state or sub-state level. The aim of the program is to strengthen practice, which may subsequently influence national and state policy. Pratham also aims to influence the government through sharing of information, from the monitoring of learning outcomes in Read India and through the nationwide ASER survey.

5.2.8 Scaling-up Early Reading Intervention (SERI) – India

Introduction

Room to Read began operating library programs in India in 2003, and in 2009 expanded the programs to include literacy instruction. The SERI program design and implementation strategy was based on Room to Read's global literacy model, which has been implemented and refined across Asia and Africa in the past decade. Room to Read has implemented its comprehensive literacy program in India since 2009, with quasi-experimental impact evaluations that consistently showed positive results. In order to scale up these programs, Room to Read entered into an agreement with USAID in 2015 to implement the SERI program in government primary schools in four states, starting with Chhattisgarh and Uttarakhand in 2015, and ending with Madhya Pradesh and Uttar Pradesh in 2017. The agreement was for a five-year program with the twin objectives of demonstrating (1) a robust model for improving early grade reading skills of students attending govern-ment primary schools, and (2) an innovative approach for effectively scaling up the early grade reading model within the government education system.

The program was designed to be implemented in three phases, with a gradual release of responsibility to the state government that mirrored its pedagogical approach. First, the demonstration ("I do") phase involved Room to Read directly implementing the literacy instruction and library programs in a small number of government primary schools in selected districts to demonstrate the impact of the program. Second, in the partnership ("we do") phase, Room to Read supported state governments to expand the model across all government schools in a block (subdistrict) or district. Finally, the scale-up ("you do") phase involved handing over the model to the state government for replication and scale-up in other parts of the state. The partnership ("we do") phase in Chhattisgarh is the focus of

the Learning at Scale research study. It was implemented in one of Chhattisgarh's 28 districts. Five hundred schools were enrolled in the program, organized into 64 clusters within four blocks. A similar program was conducted in Uttarakhand over the same period, and in Madhya Pradesh and Uttar Pradesh two years later.

Evaluation

An internal evaluation of the demonstration phase of SERI was conducted during 2015–2017 in Chhattisgarh and Uttarakhand (Joddar, 2018). The sample consisted of 50 program schools (25 in each state) and 50 comparison schools (25 in each state). An internal evaluation of the partnership phase was conducted from 2016 to 2018 in 74 program and 72 comparison schools, with each group split evenly between the two states. Both evaluations employed quasi-experimental designs.

The results were remarkable in that the large gains in literacy scores in the demonstration phase were replicated in the partnership phase, despite the decreased involvement of Room to Read in this second phase. The effect sizes of between 1.3 and 1.5 SD were substantial and translated to a mean fluency rate in program schools, in terms of cwpm, that was double what was found in the comparison schools at the end of grade 2 (**Table 33**).

Table 33.   Internal evaluation oral reading fluency gain scores in Chhattisgarh and Uttarakhand States, India

| Phase | Group | ORF rate (in cwpm), by time point | | | Effect size (SD) |
| | | Baseline | Midline | Endline | |
		Start of grade 1	End of grade 1	End of grade 2	
Demonstration	Program	0.5	9.4	33.3	1.4
	Comparison	0.3	2.8	15.3	
Partnership	Program	0.9	—	36.7	1.3
	Comparison	0.5	—	17.9	

The partnership model in Madhya Pradesh and Uttar Pradesh was evaluated using a different methodology (Joddar, 2019). A cross-sectional design was used with assessments taking place at the end of grade 2 for both baseline (2017) and endline (2019); see **Table 34** for results.

Table 34.   Internal evaluation oral reading fluency gain scores in Madhya Pradesh and Uttar Pradesh States, India

| State | Group | ORF rate (in cwpm), by time point | | Effect size (SD) |
| | | Baseline | Endline | |
		End of grade 2 (2017)	End of grade 2 (2019)	
Madhya Pradesh	Program	19.0	28.8	1.2
	Comparison	14.5	19.2	
Uttar Pradesh	Program	24.4	34.6	Not available
	Comparison	18.7	26.7	

The evaluation findings suggest that the program was less effective in Madhya Pradesh and Uttar Pradesh compared with Chhattisgarh and Uttarakhand. Room to Read reported two possible explanations for the reduced impact. First, the demonstration model was implemented in parallel with the partnership model in Madhya Pradesh and Uttar Pradesh. In Chhattisgarh and Uttarakhand, the demonstration model was implemented first and was instrumental in garnering support for the subsequent partnership model. Second, implementation of the partnership model was delayed by around five months in Madhya Pradesh and Uttar Pradesh. The impact was greater in Madhya Pradesh because the program took place in a district (Barwani) with low learning levels and greater room for improvement than the district (Varanasi) where the Uttar Pradesh program took place.

Program model

SERI's "we do" partnership model was based on that of the direct implementation model used in the demonstration phase. In the direct implementation model, teachers were trained to follow a teachers' guide with scripted lesson plans to provide literacy instruction. Teachers were monitored and supported by coaches who visited schools regularly (see below). Schools were also trained in setting up libraries and provided with storybooks in the language of instruction. The key difference in the partnership model was that support to teachers was less intense, library books were fewer, and both were provided by the government rather than by Room to Read.

Materials

A District Resource Group—that is, a group of district officials assembled to provide technical guidance to the program—reviewed materials from the demonstration schools and adapted them for partnership schools. Materials consisted of the following:

- Student workbooks in two parts. The first part included 10 weeks of exercises related to phonics instruction; the second part consisted of decodable readers.

- A teachers' guide with scripted lesson plans.

- Storybooks for the school library. The government provided 300 books per school in the partnership model; Room to Read provided 1,000 books per school in the demonstration model.

- Letter cards for use in class.

Outside of Room to Read programs, government schools typically have only textbooks for language lessons and—in some states—a student workbook.

Teacher training

Training in the SERI partnership program took place through a cascade model. Master trainers were trained twice a year, typically for three days each time. In some states, the government had in place designated master trainers when SERI began. In other states, master trainers were nominated from among block or cluster resource coordinators or teachers. Master trainers were provided with training materials, presentation slides, and session plans to train teachers, with a focus on practicing new skills. The number of training days varied by state. In the Chhattisgarh partnership model, teachers were trained twice a year for four days each time, compared to a total of 12 days' training in the demonstration model. Ideally, three teachers were trained from each school: a grade 1 teacher, a grade 2 teacher, and a teacher responsible for the library. At demonstration schools, head teachers received training and helped ensure that each of their schools had three trained teachers. At partnership schools, head teachers were not trained. Cluster resource coordinators were also trained on how to monitor teachers in two trainings a year, each lasting three days.

Teacher support

The coaches who supported teachers in the partnership model were the cluster resource coordinators, employed by the government. One cluster coordinator was responsible for 210–215 schools. Cluster resource coordinators initially worked alongside Room to Read's Literacy Facilitators, who built their capacity in school observation and coaching by involving them in joint school visits. At the start of the program in Chhattisgarh, there was one Literacy Facilitator per 20 schools. The ratio shifted to one Literacy Facilitator for 50 schools by the end of the program, as cluster resource coordinators were able to take on more coaching responsibilities. They visited schools approximately once every two months. By comparison, in the demonstration model, teachers were supported by literacy coaches employed by Room to Read, who were each responsible for around 7–10 schools and who visited those schools every two weeks.

Cluster resource coordinators' responsibilities on each school visit typically included monitoring teacher and student attendance, physical infrastructure, and implementation of various government programs, including the midday meal program. Under the SERI program, they were given the additional responsibilities of monitoring the school library, conducting a classroom observation, and discussing at least one area of improvement with

teachers. Cluster resource coordinators were given a monitoring form to complete, including information on book checkout and reading activities in the library. Block resource coordinators met with the cluster coordinators once a month to discuss, among other things, their experience supporting teachers.

Given competing responsibilities, cluster resource coordinators had little time available to observe classrooms. Room to Read's theory of change for teacher behavior therefore focused more on teacher training and the use of the teachers' guide.

<u>Pedagogical approach</u>

In SERI, Room to Read took an evidence-based, comprehensive approach to literacy instruction. Its pedagogical model included three components: development of oral language, orthographic expertise, and exposure to text. The program employed a scientific approach to developing reading skills and a habit of reading. Instruction focused on development of oral language, phonological awareness, phonics, fluency, vocabulary, comprehension strategies, and writing. Instruction was supported by a library and opportunities for independent reading. The aims were to explicitly instruct students in literacy skills and to expose them to a variety of texts, both decodables and graded literature, so that the literacy experience would be as complete as possible.

The pedagogical approach differed significantly from that used prior to the SERI program in Uttarakhand, which had employed a whole-language approach. The approach to literacy instruction in Chhattisgarh had been more flexible and thus was more compatible with the Room to Read model.

Systems

SERI developed a partnership model for working with government systems. This model recognized the limitations of scaling up by having a nongovernmental organization (NGO) either expand direct implementation or hand the program over entirely to the government. The partnership "we do" phase encompassed a lot of support for government institutions and structures, particularly in the early phases of the program. The focus of institutionalization was on the role of coaches. In Baloda Bazar, in Chhattisgarh, teachers were initially coached by a combination of Room to Read's Literacy Facilitators and the government-employed cluster resource coordinators. This process of mentoring, combined with initial training, built the skills of the cluster resource coordinators, who then led the monitoring process. State monitoring forms and education management information system data were also adapted to include two to three indicators from Room to Read's monitoring forms. The program was thereafter embedded in the system, with cluster resource coordinators being managed by block coordinators and in turn by the district office. The state continued to rely on Room to Read for additional training and for some analysis of monitoring data.

The scale-up ("you do") phase of the model was in operation in three districts by March 2020. The program in Rajnandgaon District was funded by the UNICEF, and in Sukma and Jashpur Districts by the state government. On reflection, Room to Read concluded that the government required more support than anticipated, and the "you do" phase was more accurately characterized as a second iteration of the "we do" phase.

The structure of the education system in Chhattisgarh

The education ministry in Chhattisgarh State has two components. The first component is administrative and is known as the Rajiv Gandhi Shiksha (Education) Mission. This is the state component that addresses the Sarva Shiksha Abhiyan, or Comprehensive Education Campaign—the Government of India's flagship program designed to help the country achieve universal elementary education. The second component is the State Council for Education Research and Training, which has a national equivalent council. The state council is responsible for the technical aspects of education, including the design of curriculum and teacher training.

At the district level, the education office is managed by the District Mission Coordinator. The administrative side of the office is the responsibility of the District Education Officer, supported by several Assistant Project Coordinators with different functions (e.g., finance, training), one of whom was the designated "nodal" officer for the SERI program. The technical side of education is led by the District Institute of Education and Training. The chain of accountability from the district office runs through block and cluster resource coordinators to schools.

The process of developing state government partnerships

Room to Read's aim in developing the partnership model was to encourage a sense of ownership from the government. Thus, it chose not to impose its own instruction materials on state governments, but instead, to co-create them. Room to Read presented its materials to the State Council for Education Research and Training in each state, which then incorporated changes. This process was time-consuming, particularly in Uttarakhand, which previously used a whole-language approach to instruction. There was also a lot of discussion about scripted lesson plans. The norm in India was for teachers to have more freedom in designing their lessons. This issue was addressed in part by referring to the success of demonstration schools and, in some cases, by taking teachers to see Room to Read schools. Also, the scripted lessons were presented as a "guide" to teachers, rather than as something that had to be followed.

In Chhattisgarh, the proposal for the SERI partnership was developed by the State Pedagogy Coordinator, a person working under the Rajiv Gandhi Shiksha (Education) Mission with 20 years of experience in the role. He was the subsequent champion and main driver of the program. After a formal agreement was reached between state and district offices, involving the head of the education function and the overall administrative head in

each office, one Assistant Project Coordinator became the designated lead for the district. He was a champion of the program and led an advocacy campaign to persuade schools to join SERI, on a block-by-block basis, with some blocks needing more persuading than others.

The state (Chhattisgarh) provided the budget for teacher training and materials. During data collection, plans were in place to provide Room to Read's *Tarang* student workbook across all 28 districts in Chhattisgarh. Room to Read funded the training of master trainers and cluster resource coordinators.

5.3 High-Level Analysis: Program Matrix

5.3.1 Introduction

This section presents a descriptive analysis of the elements of the Learning at Scale interventions presented in detail in Annex C. We wanted to know what program design elements were included in each of the eight interventions and which of these elements the programs deemed as key to their success. These data were collected from program documents, program visits, and interviews with the program teams. After we collected the descriptive data, the programs confirmed the program 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, alongside, and to change government behavior at all levels of government systems. (20 elements) Example: Program has staff at the regional level.

The eight programs were asked to identify which elements of the program were key to their program's success. To simplify the analysis, programs could note up to three key elements per program domain.

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

5.3.2 Most frequently implemented program elements

A wide variety of elements were implemented across the eight programs studied here, suggesting there are likely multiple paths to successful implementation. Twenty-one elements were reported by seven or eight of the programs, as presented in **Table 35,** listed by domain. Considering these frequently implemented program elements is useful for other programs hoping to implement effective large-scale programs. This section describes these program elements by domain.

Table 35.   Most implemented elements across domains

Domain	Element
Materials	Supplementary readers
Materials	Program materials aligned to the government curriculum
Pedagogy	Phonics-based instruction
Pedagogy	Direct instruction (explicit and systematic)
Pedagogy	Pair work
Training	Initial face-to-face training
Training	Refresher face-to-face training
Training	Teacher training emphasizes modeling/practice
Training	Nonresidential teacher training
Training	Teacher training (lowest level of cascade) done by government officers
Training	Structured training manuals
Support	Coaches have structured tools
Support	Coaches provided with program/teacher materials
Systems	Program has regional staff
Systems	Program uses monitoring data to make decisions about implementation
Systems	Program invested in capacity building at a decentralized level
Systems	Program shares achievement data with government decision makers
Systems	Program designed to align with existing education plans
Systems	Program responsible for distribution of materials
Systems	Government uses monitoring data to make decisions about implementation
Systems	Program monitors frequency of coach visits

Only two Materials elements were frequently included. While the programs showed variation in terms of the types of materials they provided to classrooms, almost all made an effort to align their materials to the current curriculum to help ensure government acceptance and to bridge from what teachers already knew to any new content and instruction. In addition, supplementary readers were frequently a part of program design.

Pedagogical design showed three elements that were frequently included. The programs focused their instruction on phonics-based approaches that emphasized teaching the parts of a word to support students' decoding skills. Direct instruction, which supports instruction that explicitly teaches skills systematically, was also ubiquitous. Pair work was also frequently used as a pedagogical method.

Several elements of Training were used consistently across the Learning at Scale programs. Program trainings were aligned with research on adult learning: the trainings focused on modeling and practice to make the training practical, and were implemented as multiple, shorter trainings to avoid overloading teachers with too much content at one time. Trainings involved government officers, not only to create a sense of government ownership but also to build officers' knowledge in the same way as teachers' knowledge. Programs used structured training manuals to increase fidelity to the key training approaches.

The domain with the fewest agreed-upon elements was Teacher Support. How a program implements teacher support depends on that program's context and resources. However, the programs agreed that coaches needed both structured tools and the same materials that the program provided to teachers. These tools can serve as a guide for coaches who are new to the instructional methods and support mechanisms, and coaches need copies of the teacher materials to understand the instructional approach and follow along with the lessons.

Finally, programs employed several Systems elements. These included the use of data, as most programs used monitoring data to make decisions about implementation, and so did their government counterparts. The Systems elements included elements related to working with government, including aligning with existing education plans, having regional staff who work with government counterparts, and building capacity at the lower levels of government. In addition, the programs, rather than the government, were responsible for distributing materials and for monitoring how often coaches made classroom visits, even if these coaches were government officers. It is difficult to infer very much about the relative frequency of common program elements included in most of the programs by domain, but it is worth emphasizing that these Systems areas were more frequent in the Learning at Scale interventions.

5.3.3 Program design

We used the basic information derived from each program to develop the results shown in **Figure 10.** 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 were key to their success, but in this subsection, we simply present data on all the possible elements. We found that the average Learning at Scale intervention had more than 60% of program elements in each of the five domains, from a low of 63% in Teacher Support and Materials to 74% of the Systems elements. Tusome's domain with the highest percentage of possible elements was Training (85%), and the lowest was Teacher Support (57%). PRP had relatively large percentages for all program-designated key element domains, with a high of 86% in Teacher Support and a low of 69% in Training. Pratham's Read India had 80% of the Pedagogy elements and only 37% of the Systems elements. EQUIP-T had 68% of the Systems elements and only 20% of the Pedagogy elements, and SERI had 69% of the Training elements and 31% of the Materials elements. Ghana Learning, like PRP, had large proportions of all the element domains, with 80% of the Pedagogy elements and 69% of the Training elements. Lecture Pour Tous was similar to Ghana Learning, with high percentages of the elements in all domains, ranging from 92% of the Materials and Training elements to a low of 70% of the Pedagogy elements. Finally, NEI+ had 90% of the Pedagogy elements and 69% of the Materials elements.

These data on element inclusion allow for a comparison between each of the eight programs and the typical intervention considered under Learning at Scale. It is clear that PRP, Ghana Learning, Lecture Pour Tous, and NEI+ were similar in that they had large portions of all the program elements across the domains. In contrast, Tusome was more focused on Training and Systems, and Read India was more focused on Pedagogy. EQUIP-T emphasized Systems, and SERI had high percentages in Training and Systems. This comparison also allows us to see how programs were similar to or different from each other and reveals that the four programs mentioned above (in Ghana, Nigeria, Pakistan, and Senegal) had key similarities. Note that this simple descriptive analysis does not allow us to determine which of these elements were essential for program implementation and it is likely that programs had elements in place that were not critical for their impacts on learning.

Figure 10.   Learning at Scale program elements
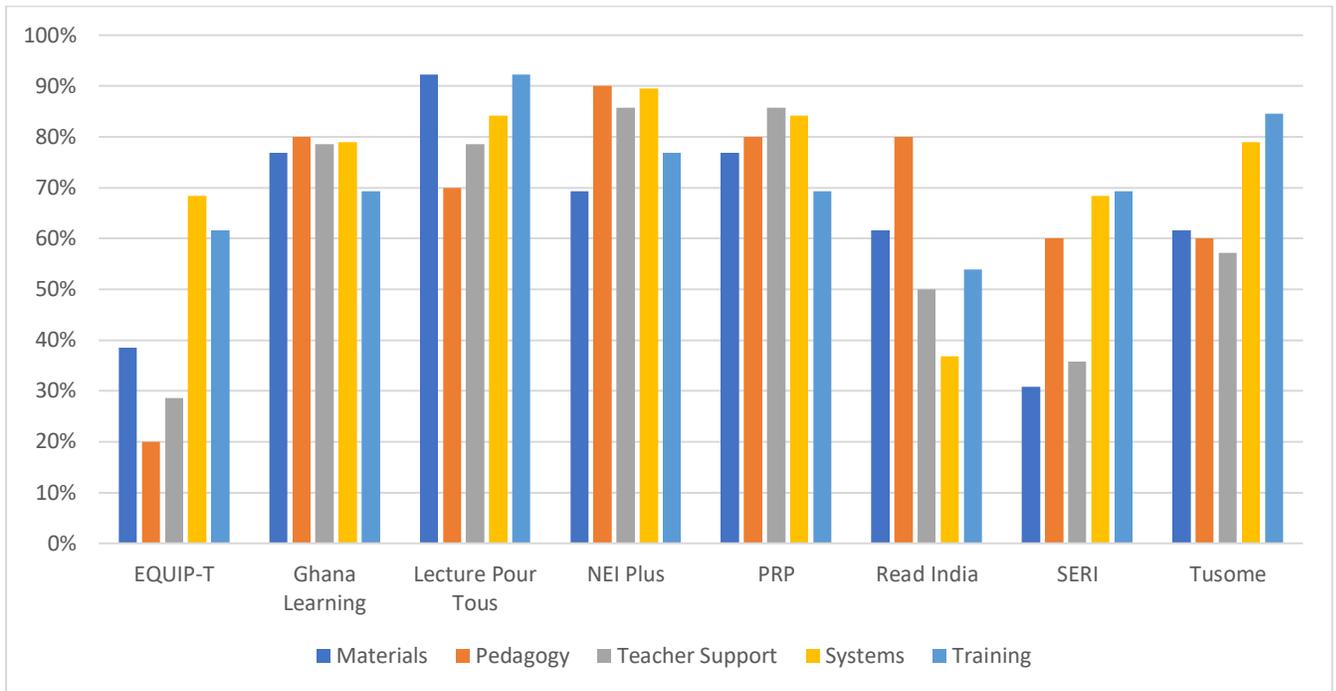
5.3.4 Key elements of Learning at Scale programs

The previous subsection focused on the various elements included in the programs. It is also important, however, to know what elements the implementers themselves identified as key to their success. **Table 36** presents the top three elements they named, by domain. This breakdown allows us to see similarities or differences in these key elements. Although each program was asked to identify up to three elements per domain that were key to success, some identified fewer than three, and some identified more.

Table 36.   Key program elements named by implementers, by domain

Category	EQUIP-T	Ghana Learning	Kenya Tusome	Lecture Pour Tous	NEI+	PRP	Read India	SERI
Materials	Materials developed with the government	Structured teachers' guides (scripted lessons)	Structured teachers' guides (scripted lessons)	Structured teachers' guides (scripted lessons)	Structured teachers' guides (scripted lessons)	Structured teachers' guides (scripted lessons)	Teaching aids (e.g., big books, letter cards, pocket charts)	Structured teachers' guides (scripted lessons)
Materials	Program materials aligned to the government curriculum	Student books for all students (at least one book per student)	Student books for all students (at least one book per student)	Student books for all students (at least one book per student)	Local-language materials	Materials developed with the government	Word walls, letter charts, and others	Student books for all students (at least one book per student)
Materials	Teaching aids (e.g., big books, letter cards, pocket charts)	Materials developed with the government	Program materials aligned to the government curriculum	Local-language materials	Textbook taken home	Consumable student books (workbooks)	Lesson plans (little to no scripting or structure)	Program materials aligned to the government curriculum
Pedagogy		Direct instruction (explicit and systematic)	Direct instruction (explicit and systematic)	Direct instruction (explicit and systematic)	Direct instruction (explicit and systematic)	Phonics-based instruction	Continuous assessment	Direct instruction (explicit and systematic)
Pedagogy		Phonics-based instruction	Phonics-based instruction	Phonics-based instruction	Increased instructional time in lessons	Increased instructional time in lessons	Group work	Increased instructional time in lessons
Pedagogy		Gradual-release model ("I do, we do, you do")	Increased instructional time in lessons	Gradual-release model ("I do, we do, you do")	Mother-tongue program	Continuous assessment	Implemented with a numeracy program	Gradual-release model ("I do, we do, you do")
Training	Teacher training emphasizes modeling/ practice	Teacher training emphasizes modeling/ practice	Teacher training emphasizes modeling/ practice	Teacher training emphasizes modeling/ practice	Teacher training emphasizes modeling/practice	Teacher training emphasizes modeling/practice	Teacher training emphasizes modeling/ practice	Structured training manuals

Category	EQUIP-T	Ghana Learning	Kenya Tusome	Lecture Pour Tous	NEI+	PRP	Read India	SERI
Training	Initial face-to-face training	Initial face-to-face training	Structured training manuals	Initial face-to-face training	Initial face-to-face training	Initial face-to-face training	Initial face-to-face training	Teacher training (lowest level of cascade) done by government officers
Training	School-based training	Refresher face-to-face training	Teacher training (lowest level of cascade) done by government officers	Training for head teachers	Training for head teachers	Structured training manuals	Refresher face-to-face training	Training of trainers done by program staff
Training						Training for head teachers		
Teacher Support	School-based community-of-practice meetings	Coaches have structured tools	Coaches have structured tools	Coaches have structured tools	Coaches have structured tools	Coaches have structured tools	School-based community-of-practice meetings	Coaches have structured tools
Teacher Support	Internal-to-school coaching	School-based community-of-practice meetings	External-to-school coaching	Coaches provided with program/ teacher materials	Coaches use tablets or other devices (e.g., smartphones)	External-to-school coaching	Coaches provided with program/ teacher materials	Coaches provided with program/ teacher materials
Teacher Support		Program supports coaches in schools	Coaches use tablets or other devices (e.g., smartphones)	Internal-to-school coaching	Virtual communities of practice (WhatsApp, SMS; not face to face)	Coaches are government staff	Coaches meet in groups/ with supervisors	Coaches are government staff
Systems	Program invested in capacity building at a decentralized level	Program develops and uses dashboard for result/data sharing	Program develops and uses dashboard for result/data sharing	Program invested in capacity building at a decentral-ized level	Program develops and uses dashboard for result/data sharing	Program invested in capacity building at a decentralized level	Program invested in capacity building at a decentralized level	Program designed to align with existing education plans

Category	EQUIP-T	Ghana Learning	Kenya Tusome	Lecture Pour Tous	NEI+	PRP	Read India	SERI
Systems	Program supports government beyond literacy instruction	Program shares achievement data with government decision makers	Program responsible for distribution of materials	Program designed to align with existing education plans	Program monitors frequency of coach visits	Program designed to align with existing education plans	Government responsible for distribution of materials	Government responsible for monitoring frequency of coach visits
Systems	Government uses monitoring data to make decisions about implemen-tation	Program staff embedded in government offices	Program monitors frequency of coach visits	Program staff embedded in government offices	Program has regional staff	Program shares achievement data with government decision makers	Program uses monitoring data to make decisions about implemen-tation	Program mobilizes additional local resources to support schools
Systems	Program sends funds to the government					Program invested in capacity building at a central level		
Systems						Program mapping/scoping exercise prior to intervention		

Table 37 shows only those 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 C,** "Full Data on Program Elements and Key Elements for Program Implementation.") To simplify our analyses, we focused on the 10 program elements that were identified by at least four programs as key to program success. The 10 elements most frequently cited as key for program success were:

1. Program's teacher training focused on modeling and practicing new skills (seven programs)

2. Program included structured teachers' guides (six programs)

3. Coaches were provided structured tools to support teachers (six programs)

4. Program used face-to-face training methods for their initial trainings (six programs)

5. Program used direct instruction pedagogical methods (five programs)

6. Student books were available at a 1:1 ratio for all students (four programs)

7. Program used a phonics-based instructional methodology (four programs)

8. Program increased the amount of instructional time in reading lessons (four programs)

9. Program built capacity at a decentralized level (four programs)

10. Program was designed to align with existing government education plans (four programs).

In total, the eight Learning at Scale programs identified 18 different program elements as key to at least three programs each (four key elements in Materials; four key elements in Pedagogy; three key elements in Teacher Support; three key elements in Systems; four key elements in Training). Those considering best-practice design in highly effective large-scale programs should consider these program elements as potentially key.

Another way to examine consistency is to identify how many of the most common key elements were cited by each program. Lecture Pour Tous (12), Ghana Learning (11), PRP (11), SERI (10), and Tusome (10) had the largest numbers of the key program elements identified by other programs. This suggests, in some ways, that these programs were more likely to be collinear to other program designs. The other three programs, NEI+ (eight), EQUIP-T (seven), and Read India (five), had quite different key elements from the others; they offer more information individually about how programs can be effective. We recommend reviewing the elements of these individual programs in more detail in Annex C.

Table 37.   Key program elements as described by Learning at Scale programs (minimum three programs each)

Category	Items	EQUIP-T	Ghana Learning	Kenya Tusome	Lecture Pour Tous	NEI+	PRP	Read India	SERI	Learning at Scale programs
Materials	Structured teachers' guides (scripted lessons)		1	1	1	1	1		1	6
Materials	Student books for all students (1:1)		1	1	1				1	4
Materials	Materials developed with the government	1	1				1			3
Materials	Program materials aligned to the government curriculum	1		1					1	3
Pedagogy	Direct instruction (explicit and systematic)		1	1	1	1			1	5
Pedagogy	Phonics-based instruction		1	1	1		1			4
Pedagogy	Increased instructional time in lessons			1		1	1		1	4
Pedagogy	Gradual-release model ("I do, we do, you do")		1		1				1	3
Teacher Support	Coaches have structured tools		1	1	1	1	1		1	6
Teacher Support	Coaches provided with program/teacher materials				1			1	1	3
Teacher Support	School-based community-of-practice meetings	1	1					1		3
Systems	Program invested in capacity building at a decentralized level	1			1		1	1		4
Systems	Program designed to align with existing education plans	1			1		1		1	4
Systems	Program develops and uses dashboard for result/data sharing		1	1		1				3
Training	Teacher training emphasizes modeling/practice	1	1	1	1	1	1	1		7

Category	Items	EQUIP-T	Ghana Learning	Kenya Tusome	Lecture Pour Tous	NEI+	PRP	Read India	SERI	Learning at Scale programs
Training	Initial face-to-face training	1	1		1	1	1	1		6
Training	Structured training manuals			1			1		1	3
Training	Training for head teachers				1	1	1			3
Totals		**7**	**11**	**10**	**12**	**8**	**11**	**5**	**10**	

In the sections that follow, we present analyses of the five program element domains.

Materials

The first domain that we examine in detail is Materials. All Learning at Scale programs developed or used some type of materials as part of their implementation. However, the programs varied drastically in terms of the types of materials they used. **Figure 11** presents both the existence of the program elements across the eight programs and the data showing whether the program elements were determined to be key. Figure 11 is sorted by the number of programs that included these elements and shows that all programs included some kind of supplemental reading materials as the most common program element. Seven of the programs had program materials aligned to the government curriculum. Six of the programs ensured 1:1 ratios of books to students. Half the programs used consumable student books, and half used textbooks.

Teacher materials were developed in most programs alongside student materials. Six programs used structured or more heavily scripted lessons. Three of those six programs also used lesson plans that had little to no scripting or structure at least in some parts. This utilization of lighter scripting in several programs is in keeping with other studies that show teachers do not necessarily need a year's worth of scripted lesson plans, especially if the lessons are predictable. Several programs also mentioned having some other kind of teaching resource, such as word walls, letter cards, or big books.

Figure 11 also shows which of the Materials were deemed to be key to the program's success. As noted above, structured teachers' guides were the most frequently mentioned key Materials element (cited by six programs), along with having student books at a 1:1 ratio (cited by four programs). Several other Materials elements were identified by two programs as key, namely that the program was aligned to the government curriculum, the materials were developed with the government, and the materials were in a local language.

Figure 11.   Materials program elements and key elements
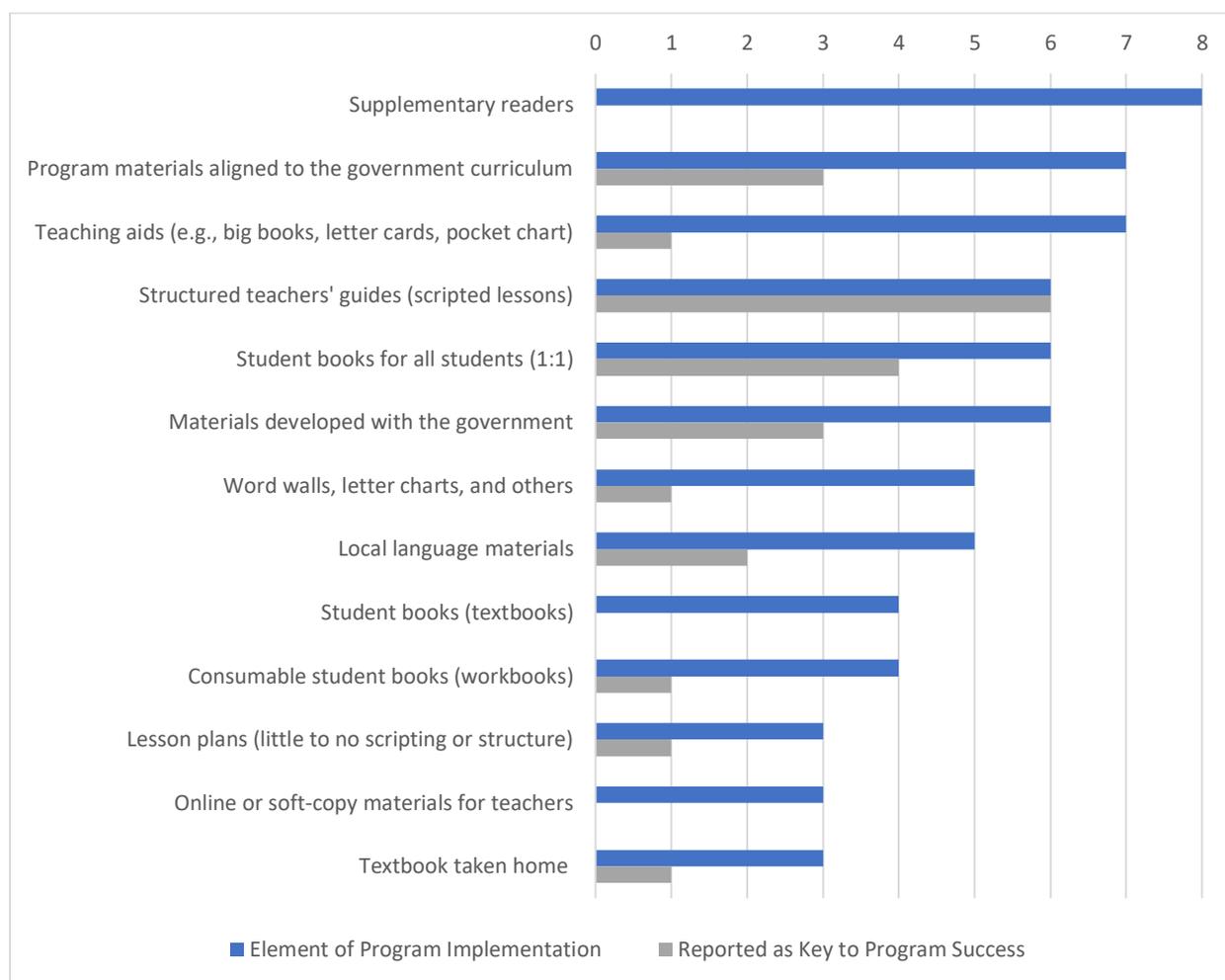
Pedagogy

For this study, we defined Pedagogy as the methods, activities, and instructional approaches used for the program's approach to classroom instruction. **Figure 12** shows the specific pedagogical elements that were most frequently cited by programs. We found that all programs used a phonics-based approach. This finding is worth emphasizing given that prior to the program, many countries typically had a language subject that focused on vocabulary, grammar, and reading connected text rather than the letters, sounds, and word-reading elements that would be the focus of a phonics approach. Most programs (seven) also said they used a direct-instruction approach, and six programs used the gradual-release model commonly called "I do, we do, you do." Both of these approaches are complementary to the phonics instructional approach and seek to teach students discreet skills explicitly and systematically to increase learning outcomes. Pair work was a part of seven programs, six programs increased instructional time, and another six cited continuous assessment as part of their intervention.

Language was another element of several programs. Five programs used mother tongue as their language of instruction, and three programs were considered bilingual programs. These language decisions are not usually made by programs themselves but can have a significant impact on how instruction is designed—and on student success.

Note that, once again, the program elements that were considered key did not match the frequency of the elements. Five programs noted that direct instruction was key to program success, while four programs cited phonics-based instruction and another four cited the increased instructional time as key. Three programs pointed to the gradual-release model, and two identified the continuous assessment of their intervention as key for success. These data showed that merely having a particular program element did not make it key; seven programs cited pair work as part of their program design, but none of them noted that it was a key program element.

Figure 12.   Pedagogy program elements and key elements
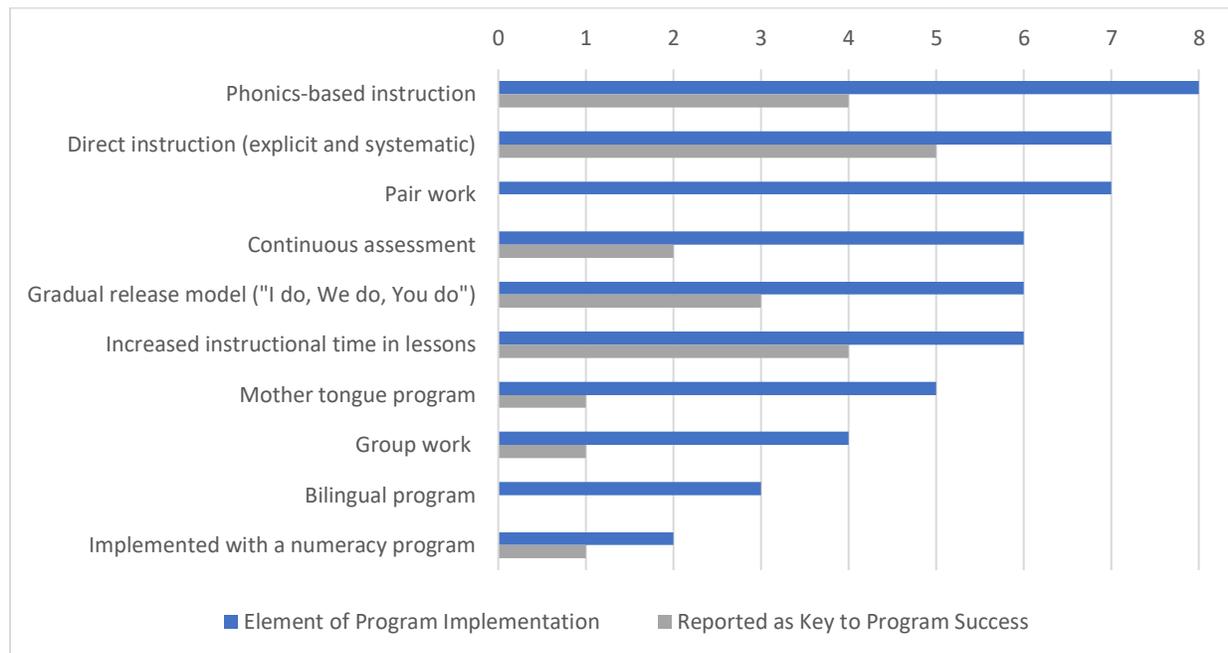
Training

In this section, we look into the program elements for Training. Note that our analysis focused primarily on in-service teacher training, given the design of the interventions, and did not include training of other government officers, because that training falls in the Systems domain.

Figure 13 shows that all programs included both initial and refresher face-to-face trainings. Nonresidential teacher training was used by seven programs, and seven programs used government officers for the lowest level of the cascade. Modeling and practice were

emphasized in the training activities of all eight programs. Another seven programs used structured training manuals. Many of these more frequent training design elements are in keeping with recommendations from the RTI teacher training study (Piper et al., 2019).

Six programs included a training for head teachers, and another six programs used program staff to train trainers of trainers. Five programs used government staff to train the trainers, and another five programs implemented school-based training. We found that pre-service training was used in four programs.

As we found elsewhere, the key elements of the programs were not always the most frequently used elements. The key elements of training emphasized by seven programs were modeling and practice, and another six programs deemed that initial face-to-face trainings were key. Three programs called out using structured training manuals and providing training for head teachers as key.

Figure 13.   Training program elements and key elements
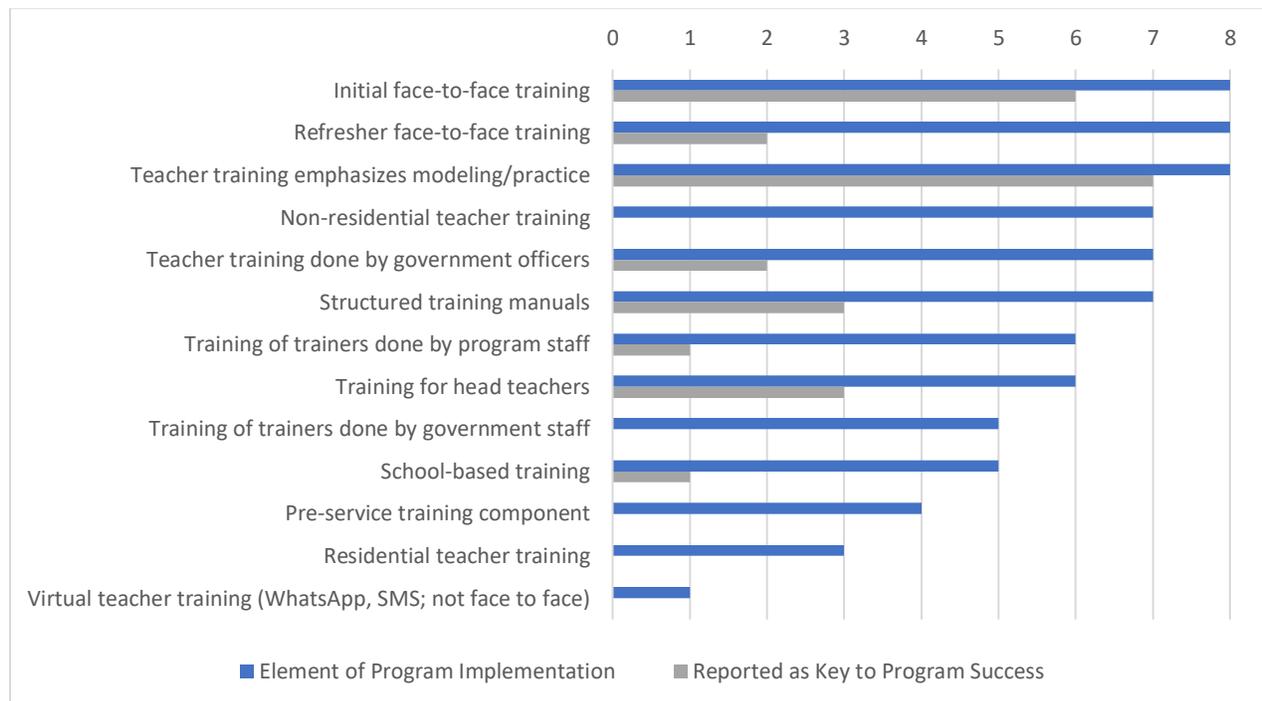
Teacher Support

Our definition of Teacher Support included both teacher coaching and communities of practice, based on the popularity of both types of teacher support in Learning at Scale programs. Programs seemed more consistent regarding these Teacher Support elements than regarding the elements in other domains. As can be seen in **Figure 14,** seven programs provided structured tools and program or teacher materials to their coaches. Six programs indicated that they used coaches who were external to the school, and the same

number of programs used government staff as coaches. Five programs said they supported their coaches, and half gave the coaches tablets to use for coaching. Only half the programs reimbursed coaches for transportation.

Communities of practice were another category of Teacher Support, used by six programs, and they were external to the schools as well. Only half the programs used some kind of structured tool for communities of practice, and three programs used some virtual technology as part of their communities of practice.

There were fewer key elements of Teacher Support that were similar across programs. Six programs noted that giving coaches structured tools was key, while three programs agreed that providing coaches with the program or teacher materials was key. Three programs used school-based communities of practice as a key element of their intervention. The other elements did not seem to be widely recognized as key, even though many programs used those program elements.

Figure 14.   Teacher Support elements and key elements
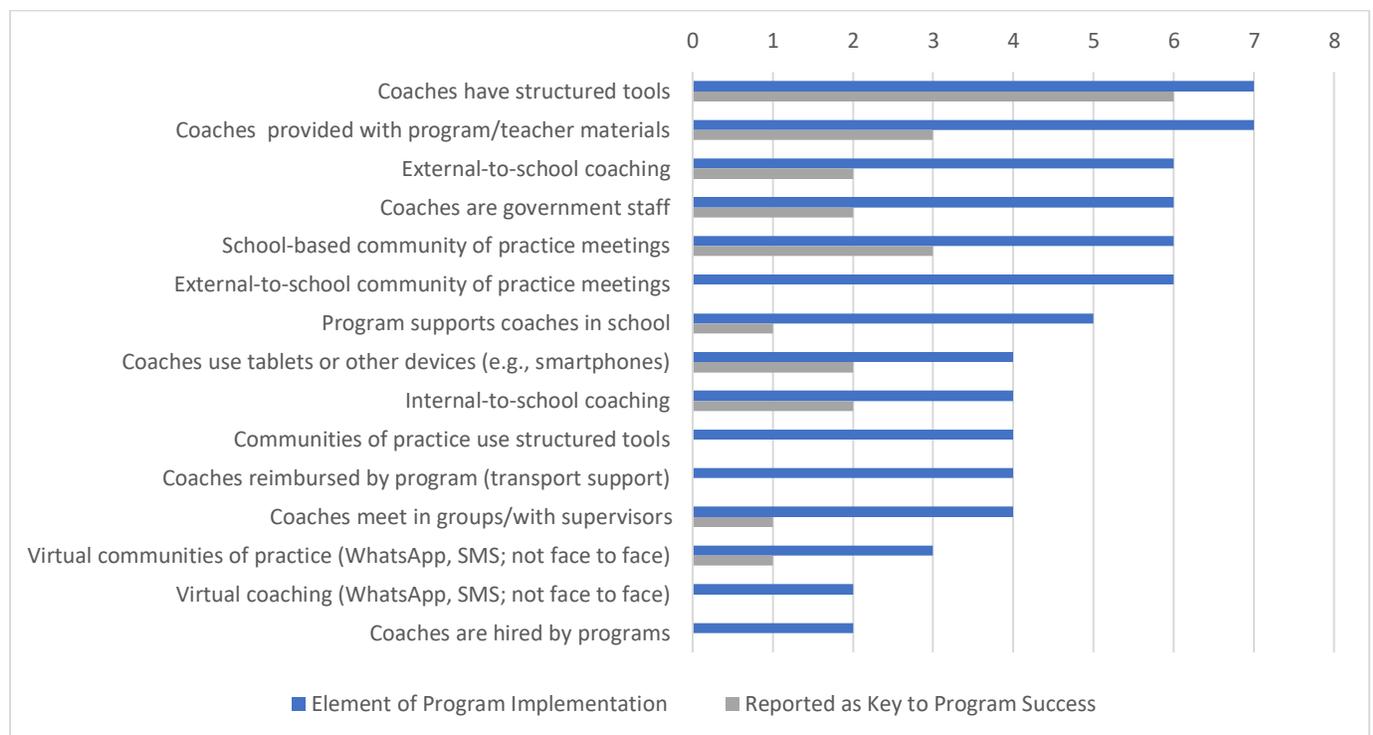
Systems

We defined Systems to include the government systems and structures as well as the program's work at scale. The Systems category was the largest and most varied domain of program elements. **Figure 15** shows that programs made a variety of design choices with respect to how the program worked with the government and how it operated in the

broader educational structure. Three program elements were consistent across all eight interventions. These were that the programs had regional staff, that the program used monitoring data to make decisions, and that the program invested in capacity building a decentralized level.

These program elements revealed activities that the programs undertook with the government to work more effectively. For example, seven programs shared achievement data with the government, seven were designed to align with existing educational plans, and seven supported the government to use monitoring data for decision making. Seven programs monitored the frequency of coach visits, and seven were responsible for book distribution. We found it quite interesting that some of these program elements were implemented by the government and others by the program as separate entities, whereas in some programs, both entities did the same tasks in tandem.

Six programs supported capacity building at the central level, and six supported the government in other areas beyond the literacy program that was the impetus for inclusion in Learning at Scale. Six programs embedded their staff in government offices with their government counterparts, and six programs ensured that government staff were responsible for conducting coaching and support visits. The program activities were split between the program and government, and in some programs, these tasks were intertwined between the two.

Given the variation in how each individual program worked with, inside of, or alongside the government, we found a wide variety of key elements in the Systems domain. Investing in decentralized capacity building was a key element for four programs, as was designing the program to align with existing educational plans. Three programs saw the program's developing and utilizing a dashboard for data management as key. Many other program elements were identified by one or two programs as key, and they should be seen as potential program design choices for future interventions, depending on how those programs will work with the government.

Figure 15.   Systems program elements and key elements
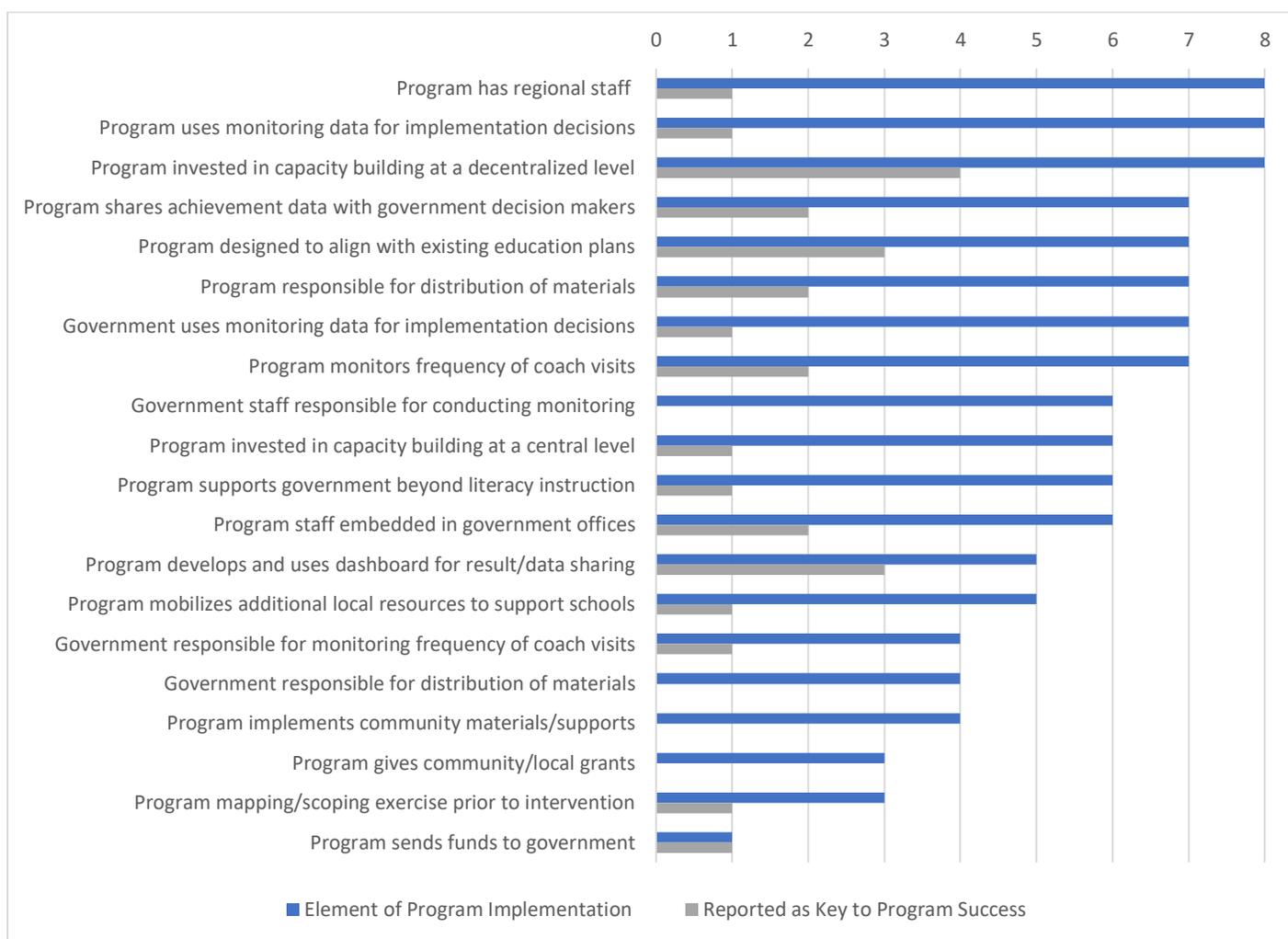
5.3.5 Summary of program matrix findings

The availability of these program design data allowed us to describe how the Learning at Scale interventions worked across the five domains of program characteristics that we identified. We found that the program design structures fit relatively well into these five areas (Materials, Instruction, Training, Teacher Support, and Systems) and that programs were designed to emphasize particular portions of these domains. A handful of programs had relatively similar designs, while others were more unique and individual in what they focused on within and across the five domains. We found 21 elements that were included in at least seven of the programs, which is useful descriptive information for program design. We also found 10 elements that were determined to be key for four or more programs and suggest that future interventions consider those 10 elements as essential for program impact. The list of essential program elements is reproduced below.

1. Program's teacher training focused on modeling and practicing new skills (seven programs)

2. Program included structured teachers' guides (six programs)

3. Coaches were provided structured tools to support teachers (six programs)

4. Program used face-to-face training methods for their initial trainings (six programs)

5. Program used direct-instruction pedagogical methods (five programs)

6. Student books were available at a 1:1 ratio for all students (four programs)

7. Program utilized a phonics-based instructional methodology (four programs)

8. Program increased the amount of instructional time in reading lessons (four programs)

9. Program provided capacity building at a decentralized level (four programs)

10. Program was designed to align with existing government education plans (four programs).

We found that some of the programs were similar in their most essential elements. Lecture Pour Tous, Ghana Learning, PRP, SERI, and Tusome all had at least 10 of the most common program elements. These interventions therefore had fundamentally similar designs. This tells us that five of the eight Learning at Scale programs were similar not only in that they were funded in part by USAID, but also in the program elements most important for program success. On the other hand, we found that NEI+, EQUIP-T, and Read India—the first two of which also were sponsored by USAID—had fewer similar program elements. Understanding these programs' designs might reveal other pathways to success at scale.

5.4 Quantitative Analyses (five programs)

5.4.1 Learning at Scale – Quantitative results

School-level data collection: Cross-program results

As described in Section 4.2.2, by March 2020, the Learning at Scale team had completed school-level data collection for the following three programs: EQUIP-T, SERI, and Tusome. Between March and December 2021, the study team completed two additional data collections: one for NEI+ and one for Lecture Pour Tous. We collected data from control schools for EQUIP-T, SERI, and NEI+, but there was no control group for Tusome or for Lecture Pour Tous[25]. In this section, we focus on treatment-only analyses to examine how results compare across the three programs. When interpreting these findings, it is critical to

---

[25] Lecture Pour Tous had no control group because the program works at scale the program was a full scale across 7 of the 14 regions in the country with no equivalent control group due to local area language and other factors.

view the data presented in this section as describing common attributes observed across successful programs – and not as a comparison between these programs. As noted above, all classroom observations were conducted during grade 2 lessons, within a three week period for each program – as such, these observations did not capture different moments along the reading curriculum progression in that country. As such, some elements that might have been observed in a different grade or at a different point in the scope and sequence were not detected during data collection. It is additionally important to note that each teacher was observed only one time.

The data used for the comparative analyses in this section are displayed in **Table 38.** For each program, assessors interviewed one grade 2 teacher and one head teacher in each sampled school. Assessors interviewed coaches, trainers, and meeting facilitators as available at the school level, which differed for each program. For EQUIP-T, we interviewed school-level teacher meeting facilitators, in addition to teachers and head teachers.

Table 38.   Sample sizes for treatment-only school-level data collection efforts

Program	Teacher Interviews	Reading Lesson Observations (1 per school)	Head Teacher Interviews	Coach Interviews	Coach Observations	Trainer Interviews	Teacher Meeting Facilitator Interviews	Student Reading Assessments
EQUIP-T	59	59	59	—	—	—	31	944
SERI	59	60	57	22	12	10	8	885
Tusome	59	59	60	46	26	46	42	952
Senegal	58	60	51	40	33	15	—	925
Nigeria NEI	59	60	61	19	17	13	—	835

1.   How has instruction improved student performance?

We designed the grade 2 lesson observations for this study to better understand the teacher practices that led to effective gains in student performance for the selected programs. Based on the timed classroom observation, we calculated the percentage of time during the observed reading lesson that the teachers and students spent on a range of activities. First, we examined the instructional focus area—the teacher's instructional goal at the time of observation. As shown in **Figure 16**, during their reading lessons, teachers in four of the five programs spent the most time focused on teaching reading (ranging from 33.8% in

EQUIP-T to 66.4% in Lecture Pour Tous[26]). NEI+ teachers spent just under one-third of the time teaching reading and roughly the same amount of time teaching oral language, which may be, in part, due to the transition to English in grade 2 (the grade being observed) in Nigeria. Tusome teachers also focused most of the remainder of their time on oral language and phonological awareness (PA), while SERI and EQUIP-T supplemented these activities with a considerable amount of time for writing activities. SERI and Lecture Pour Tous teachers spent relatively lesstime on phonological awareness. Again, it is important to note that while all observations were conducted in grade 2, classroom data was collected at different points in the school year for different programs.

Figure 16.   Percentage of observed class time spent on each instructional focus area, by program
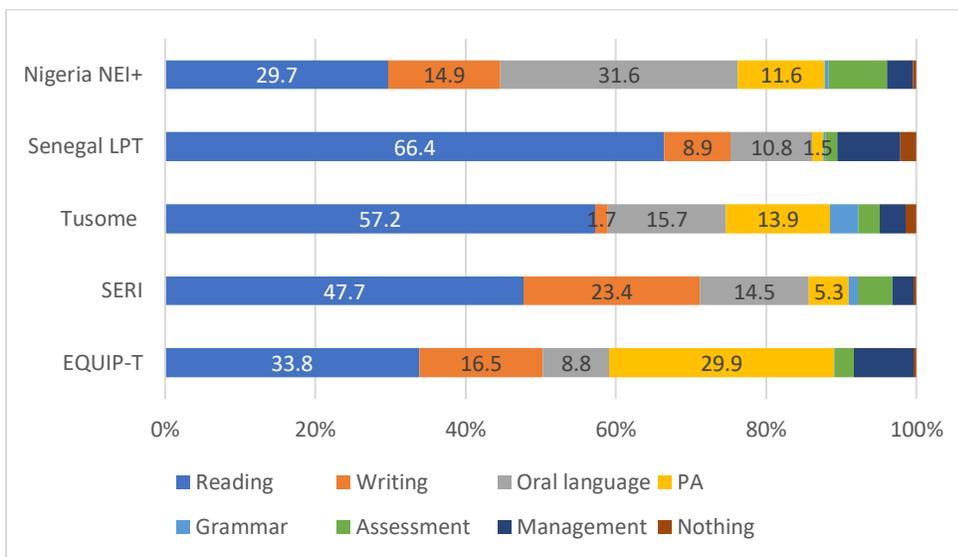
PA = phonological awareness

As shown in **Figure 17,** the amount of class time students spent actually reading varied from close to 60% in Lecture Pour Tous classrooms, to just over 40% in SERI and EQUIP-T classrooms and under 30% in Tusome and NEI+ classrooms. Students in NEI+ and Tusome classrooms spent more time listening and providing oral responses than their Lecture Pour Tous, SERI, and EQUIP-T counterparts. EQUIP-T spent substantially more time on students' writing than other classrooms.

---

[26] Assessors observed Kiswahili and English lessons in Tusome schools. Because the findings were similar across the two languages, the results in this section are streamlined by providing estimates only from Kiswahili lessons.

Figure 17.   Percentages of observed class time spent on each student activity, by program[27]
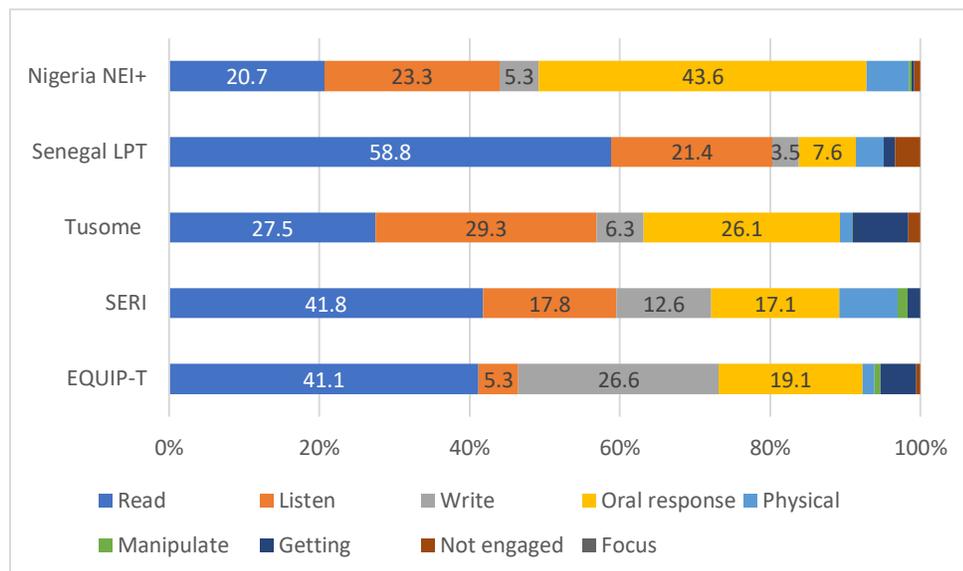
When it comes to the unit of language being taught, at the time that the classroom observations were conducted, Tusome, SERI, and NEI+ teachers dedicated more than 60% of their time to teaching words, sentences, or story reading and a smaller percentage of time on word parts, letters, and sounds, as shown in **Figure 18.** EQUIP-T and Lecture Pour Tous teachers focused more than half of the class time on the building blocks of word parts, letters, and sounds. This is likely due, in part, to where each class was in the curriculum scope and sequence at the time of data collection. However, it is clear that teachers in all five successful programs spent a significant amount of class time on the basic building blocks of reading using a phonics-based component to their instruction.

---

[27] Reading (R): The students read something silently or aloud.
Listening (L): Students are listening or waiting to answer a question. This includes raising hands.
Writing (W): Students respond by writing.
Orally (O): Students respond orally (e.g., singing a song, answering a question).
Physical (P): Students have a physical response, (e.g., raising a number of fingers on a hand; writing a word in the air)
Manipulate (M): Students respond by manipulating something with their hands. (e.g., moving blocks to represent sounds)
Getting (G): Students are getting something or waiting for the teacher (e.g., a book, school bag; lined up waiting to have their work checked).
Not Engaged (N): Students are not engaged (e.g., looking out of the window).

Figure 18.   Percentages of observed class time spent on language parts, by program
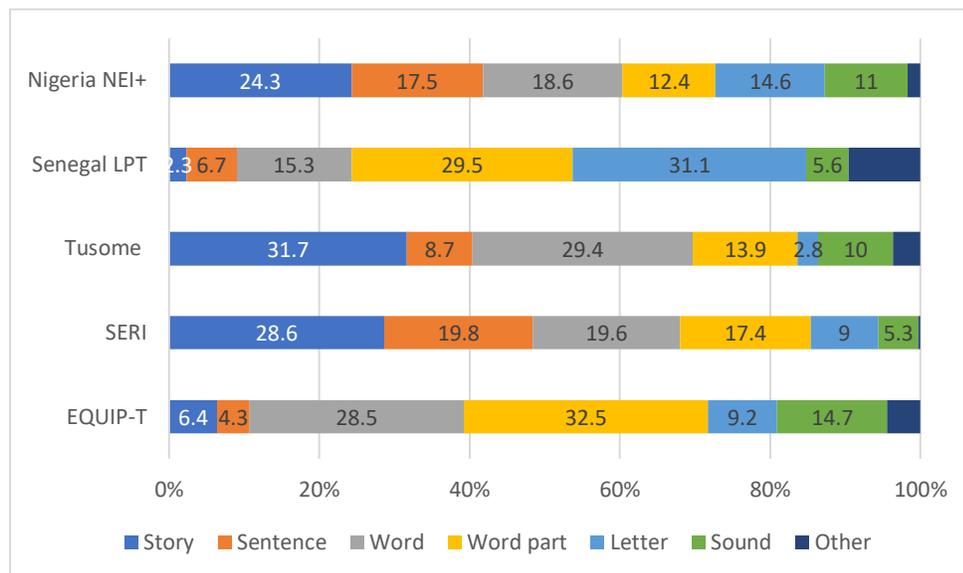
Teachers in Tusome, NEI+, SERI and Lecture Pour Tous reported that the specific focus on phonics (e.g., letters, sounds, blending) had the biggest impact on student learning, as shown in **Table 39.** In EQUIP-T classrooms, program teachers did not explicitly see the emphasis on phonics as having the biggest impact on student learning, but 39% of EQUIP-T teachers did believe that their overall new methodology or instructional approach – which included a heavy emphasis on the relationship between letter sounds and symbols - was the most important factor for change.

Table 39.   Instruction-related factors, as reported by teachers, for improved student performance, by program

What part of your instruction has had the biggest impact on student learning? (Select only one)	Tusome	EQUIP-T	SERI	Lecture Pour Tous	NEI+
More focus on letters, sounds, and/or blending	50.8%	25.4%	57.6%	63.3%	55.9%
More student centered and/or less lecture	13.6%	6.8%	3.4%	1.7%	10.2%
More pair and/or group work	6.8%	10.2%	10.2%	5.0%	6.8%
New methodology and/or instructional approach	27.1%	39.0%	18.6%	20.0%	18.6%
Involves more materials and/or activities	1.7%	11.9%	10.2%	3.3%	8.5%
Other	0	5.1%	0	3.3%	0
Program did not impact student learning	0	1.7%	0	3.3%	0

Additionally, more than two-thirds of teachers in all five programs noted that their current instruction dedicated more energy than previous approaches on letters, sounds, and

blending and that they were using a new methodology or instructional approach in the classroom (relative to what they were doing before the program), as shown in **Table 40.** It is also worth noting that the new methodology was seen as a key difference, as was the aim of placing students more at the center.

Table 40.   Instructional differences between their program and activities prior to the program, as reported by teachers

Is [program] instruction different from what you were doing before the program? (If yes, how?) (Mark all that apply)	Tusome	EQUIP-T	SERI	Lecture Pour Tous	NEI+
No difference	0	1.7%	1.7%	8.3%	1.7%
More focus on letters, sounds, and/or blending	76.3%	67.8%	83.1%	68.3%	74.6%
More student centered and/or less lecture	40.7%	55.9%	62.7%	15.0%	44.1%
More pair and/or group work	15.3%	40.7%	64.4%	10.0%	40.7%
New methodology and/or instructional approach	74.6%	67.8%	71.2%	53.3%	69.5%
Involves more materials and/or activities	16.9%	40.7%	55.9%	6.7%	30.5%
Other	8.5%	10.2%	0	10.0%	0

Lesson observations also tracked the materials teachers used during instruction. Tusome, Lecture Pour Tous, and SERI teachers spent between 40% and 50% of class time using books, with the remainder of the time primarily focused on written materials (e.g., chalkboard, letter cards) or simply speaking about language. For NEI+, teachers delivered roughly one-half of lessons orally, using books or written materials the other half of the time. For EQUIP-T, more than three-quarters of teachers' lessons used written materials, with limited book usage, as shown in **Figure 19.** This difference in approaches is explained in part by contextual factors and the difference in the availability of student textbooks across the programs. Based on post-observation classroom inventories, we found that more than 94% of students had textbooks in SERI and Tusome classrooms, while 88% of students had textbooks in Lecture Pour Tous classrooms. Conversely, only 54% had textbooks in NEI+ classrooms and just 15% of students in EQUIP-T classrooms had textbooks (as EQUIP-T also had a large focus on supplementary materials).

Figure 19.   Percentages of observed class time spent using each material type, by program
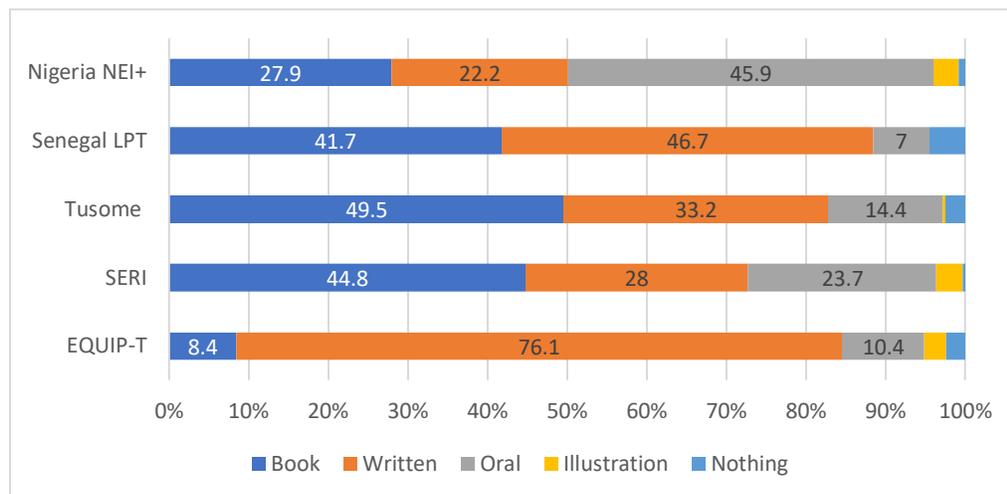
Despite variations in the time spent using books and written materials, teachers in all five programs reported that student materials were one of the most important factors for improved student learning, as shown in **Table 41.** Teacher materials (which may include lesson plans and/or guided lessons) were also frequently cited. More specifically, 70% of teachers in Lecture Pour Tous, 79.7% in NEI+, and 00% in Tusome most commonly cited textbooks as being the most important type of student material (**Table 42**). In contrast, 47.5% of teachers in EQUIP-T reported that supplementary readers were the most important student material.

Table 41.   Factors for improved student learning, as reported by teachers, by program

Has [program] helped improve student learning? (If yes, what has had the biggest impact on student learning?) (Mark all that apply)	Tusome	EQUIP-T	SERI	Lecture Pour Tous	NEI+
No improvement	0	0	1.7%	0	0
Student materials	66.1%	62.7%	74.6%	76.7%	88.1%
Teacher materials	37.3%	62.7%	71.2%	60.0%	66.1%
Lesson plans and/or guided lessons	18.6%	49.2%	83.1%	35.0%	57.6%
Teacher's instruction	40.7%	52.5%	69.5%	10.0%	23.7%
Focus on phonics	40.7%	30.5%	83.1%	20.0%	33.9%
Other	11.9%	13.6%	0	6.7%	0

Table 42.   Most useful student materials, as reported by teachers, by program

Which one of these student materials do you believe is the MOST useful? (Select only one)	Tusome	EQUIP-T	SERI	Lecture Pour Tous	NEI+
None	0	1.7%	0	0	0
Textbooks	100%	23.7%	42.4%	70.0%	79.7%
Supplementary readers	0	47.5%	5.1%	26.7%	13.6%
Student exercise books	0	5.1%	22.0%	1.7%	6.8%
Other stationery	0	3.4%	30.5%	0	0
Other	0	1.7%	0	1.7%	0
Not applicable	0	16.9%	0	0	0

In all five programs, most of the actual class time recorded during classroom observations was spent on activities involving the whole class (as opposed to group or individual work), as shown in **Figure 20.** Group work made up the smallest percentage of time, with less than 5% in four of the five programs (group work constituted 13% of observed class time under NEI+), despite the fact that between 10% and 64% of teachers reported that they were using more pair or group work than they did prior to the program (**Table 40**). Between 15% and 63% of teachers reported that their instruction was more student centered with less lecture (than before the program began), which likely accounted for the class time spent on individual student activities as opposed to group work—this was highest in Lecture Pour Tous (32.1%), followed by EQUIP-T (17.8%). This finding suggests that effective programs can primarily utilize whole-class instruction, if done well but that supplementing with other approaches is beneficial.

Figure 20.   Percentages of observed class time dedicated to various class response types, by program
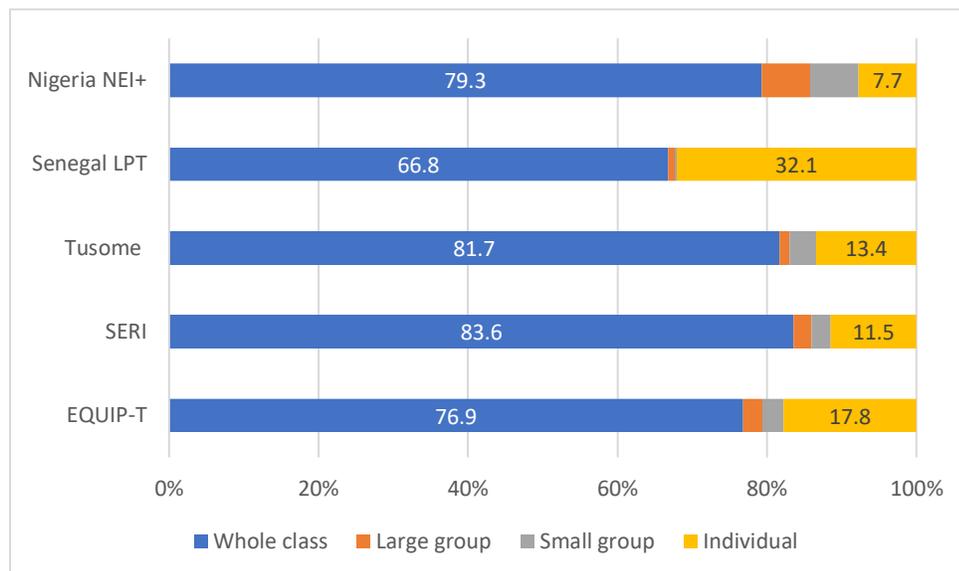
Additionally, classroom observation data indicate that these programs were effectively engaging students, with most or all students found to be "on task" (i.e., engaged in the current activity or doing something sanctioned by the teacher) between 80.9% (NEI+) and 98.3% (Lecture Pour Tous) of the time across programs (**Figure 21**). Furthermore, between approximately 20% to 50% of teachers reported that students were more engaged because of the program (**Table 43**). In addition, relatively large proportions of teachers across programs noted that their students enjoyed reading more, while approximately one-third of Tusome (27.1%), SERI (39%) and EQUIP-T (33.9%) teachers reported that their students were more confident. Also, 47.5% of teachers in EQUIP-T, 55.9% in SERI, and 39% in NEI+ reported that student attendance had improved and that students liked lessons and activities (39% for EQUIP-T, 59.3% for SERI, and 25.4% for NEI+). Nearly half of the teachers in SERI also reported that student behavior was better under the program. These findings all provide evidence of improvements in student engagement and participation in classroom activities.

Figure 21.   Percentages of observed class time students were on task, by program
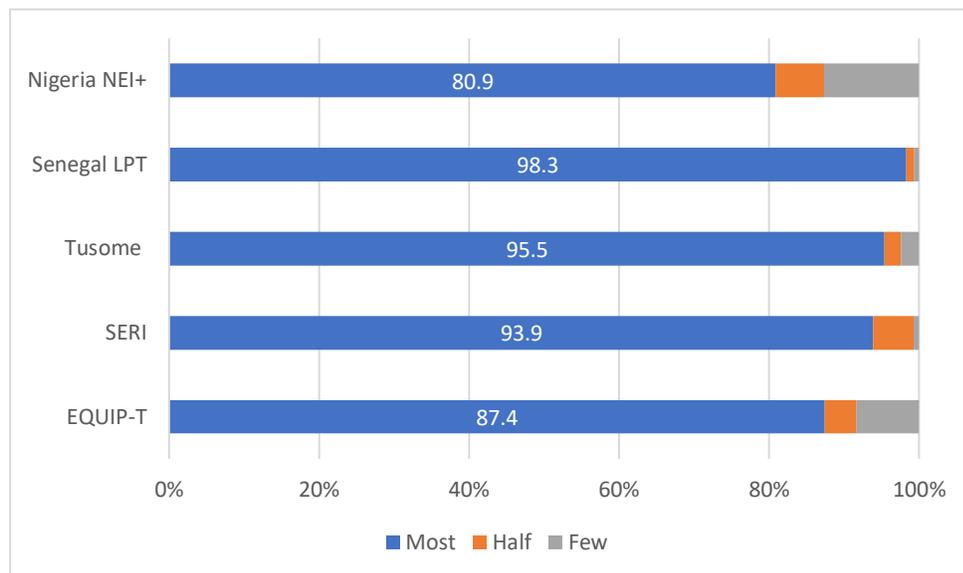
Table 43.   Changes in student engagement, as reported by teachers, by program

Has [program] impacted or changed student engagement? (If yes, how?) (Mark all that apply)	Tusome	EQUIP-T	SERI	Lecture Pour Tous	NEI+
Student attendance has improved	13.6%	47.5%	55.9%	8.3%	39.0%
Student behavior is better	10.2%	13.6%	47.5%	1.7%	25.4%
Students are more engaged	33.9%	52.5%	49.2%	21.7%	28.8%
Students like the lessons and/or activities	15.3%	39.0%	59.3%	5.0%	25.4%
Students enjoy reading more	39.0%	45.8%	61.0%	18.3%	25.4%
Students are more confident	27.1%	33.9%	39.0%	5.0%	16.9%

In addition to the classroom observation instrument, all assessors completed a post-observation checklist to determine which key activities occurred at least once during the observed lesson. The results are presented in **Table 44.** There were high levels of inconsistency in the activities observed and recorded on the post-observation checklists across programs- likely in part due to the curriculum (i.e. time dedicated to writing or grammar) and the timing of data collection (where classes were in the scope and sequence).While this underlines that there is "more than one right way of doing things", activities were consistently observed across programs in the Responsiveness category. This finding shows that teachers in these programs were most consistently conducting activities (often focused on the Simple View of Reading) and were checking for understanding and providing scaffolded responses to students.

Table 44.   Percentages of classrooms with observed activities based on the post-observation checklist, by program

Checklist Factors	Programs				
Student-Centered Dimensions	**EQUIP-T**	**SERI**	**Tusome**	**Lecture Pour Tous**	**NEI+**
4. The teacher demonstrated how to use a resource (e.g., word wall, alphabet strip)	69.5%	75.0%	37.3%	62.7%	61.0%
5. Playing with language (e.g., alliteration, clapping syllables, onomatopoeia, rhyme)	59.3%	50.0%	27.1%	25.4%	54.2%
24. Instruction included an item designed for interaction between teacher and the students or between students	76.3%	71.7%	54.2%	91.5%	45.8%
27. The content was student directed	10.2%	60.0%	11.9%	20.3%	39.0%
Demonstration					
1. The teacher said that s/he wanted students to do the activity that s/he did	69.5%	98.3%	89.8%	100.0%	91.5%
2. The teacher explained what s/he was demonstrating	25.4%	45.0%	39%	49.2%	62.7%
3. While demonstrating, the teacher "thought aloud" to show her/his thinking	20.3%	55.0%	27.1%	42.4%	49.2%
Research-Based Simple View of Reading					
6. Print awareness activities (e.g., discussing book title and cover, tracking text)	16.9%	73.3%	69.5%	30.5%	44.1%
7. Instruction of letter sounds and their symbols	59.3%	73.3%	88.1%	67.8%	88.1%
8. Word strategies for decoding (e.g., blending, chunking, look at letter)	64.4%	88.3%	88.1%	67.8%	71.2%
9. Attention to accuracy, rate, or expression while the teacher reads or students read	66.1%	80.0%	94.9%	61.0%	79.7%
10. Activities to learn the meaning of words	37.3%	66.7%	91.5%	30.5%	62.7%
11. Activities or discussions about the meaning of connected text	45.8%	80.0%	91.5%	54.2%	79.7%

Checklist Factors	Programs				
Student-Centered Dimensions	**EQUIP-T**	**SERI**	**Tusome**	**Lecture Pour Tous**	**NEI+**
Application					
12. Without the teacher, students practiced skills that were demonstrated earlier	47.5%	70.0%	78.0%	61.0%	50.8%
13. Students read letters, individual words, or connected text on their own or with others	54.2%	61.7%	88.1%	79.7%	54.2%
14. Students discussed the text read (e.g., gave an opinion, evaluated a character)	6.8%	41.7%	50.8%	15.3%	16.9%
15. Students wrote or drew (letters, words, or sentences) without copying	33.9%	70.0%	45.8%	5.1%	32.2%
22. Students read from something that was dedicated for their individual use	35.6%	96.7%	98.3%	89.8%	55.9%
Responsiveness					
16. The teacher used a visual method to check understanding (e.g., thumbs up)	86.4%	65.0%	81.4%	59.3%	81.4%
17. The teacher asked questions to check for individual understanding	91.5%	76.7%	94.9%	86.4%	94.9%
18. The teacher helped students to achieve the correct answer for an initial incorrect answer	71.2%	46.7%	72.9%	79.7%	76.3%
19. The teacher circulated, leaned in, and/or redirected in order to monitor student progress	61.0%	75%	94.9%	86.4%	76.3%
20. The teacher rephrased explanations when students did not understand	88.1%	78.3%	81.4%	78.0%	83.1%
21. The teacher included students who did not volunteer to answer	74.6%	48.3%	54.2%	74.6%	57.6%
Preparedness and Efficiency					
25. The teacher had materials ready at the start of an activity or the class	88.1%	93.3%	98.3%	79.7%	91.5%
26. The pace of the instruction was lively	72.9%	65.0%	93.2%	91.5%	84.7%

Note: Green cells signify activities that occurred in at least 75% of the observed classrooms. Question 23 was dropped from the analyses, following exploratory factor analyses.

2. What do teachers see as the most useful supports and changes resulting from their respective programs?

Although each of these programs provided teachers with a range of supports (most commonly training, but also student and teacher materials), we asked teachers which type of support they found to be most useful overall (**Figure 22**). Most teachers across four of the programs said that training was the most important support provided, with 45.8% of teachers in Tusome, 47.5% in SERI, and 61.0% in NEI+ and EQUIP-T pointing to training as essential. While many Lecture Pour Tous teachers (31.7%) also cited training as being most helpful, the majority saw teacher materials as being the single most important support they received from the program – a sentiment echoed by 20.3% of NEI+ and 22% of Tusome teachers. Aside from training and teacher materials, the only other supports selected by at least 10% of teachers across all three programs were internal coaching (10% for Lecture Pour Tous), teaching aids (16.9% for EQUIP-T, 11.7% for Lecture Pour Tous and 13.6% for SERI), and student materials (13.6% for Tusome and NEI+, and 15.3% for SERI). Despite the fact that most of the teachers in each program reported that their teaching changed because of the coaching they received, coaching was rarely selected as the most important program support.

Figure 22.   Most useful program supports reported by teachers, by program
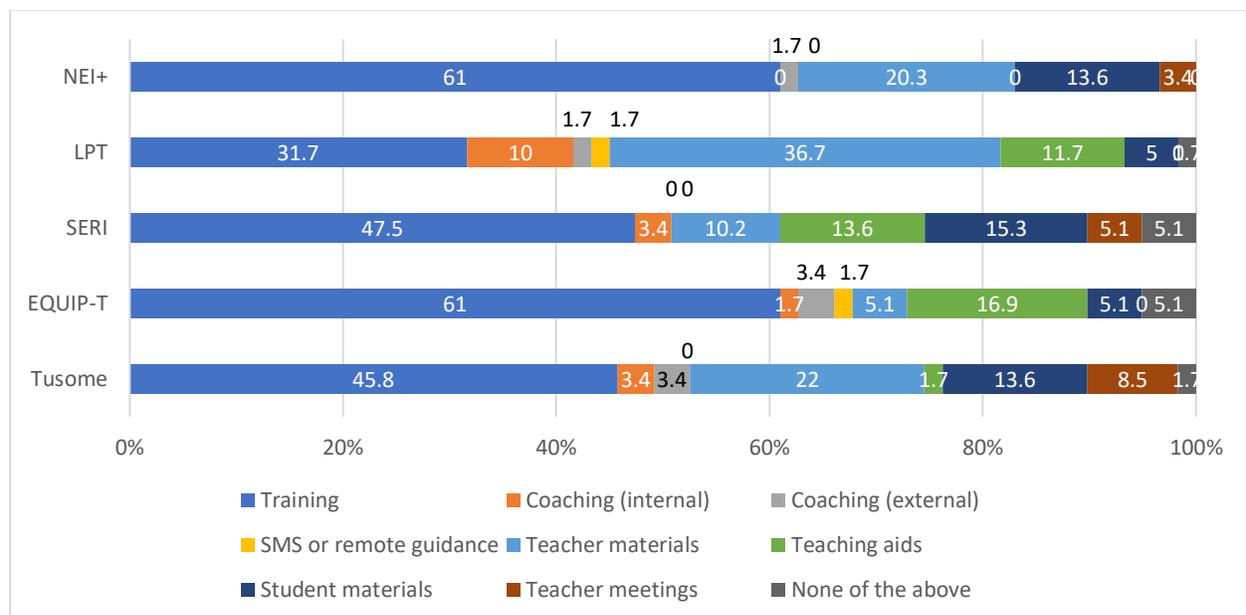
Training

Because most teachers reported that training was the most important program support, it is important to understand what made the program training sessions so effective. Teachers across all programs reported that the training sessions used more discussion (ranging from

56.9% to 88.9%) and small-group practice (ranging from 53.4% to 96.2%) than previous teacher training sessions they had attended (**Figure 23**). Approximately three-quarters of Tusome and SERI teachers and two-thirds of Lecture Pour Tous teachers also reported that the program training sessions contained more modeling than previous training sessions. A smaller percentage of teachers reported an increased use of large-group practice and lecture.

Figure 23. Percentages of teachers reporting program training methods used more frequently than in previous training sessions
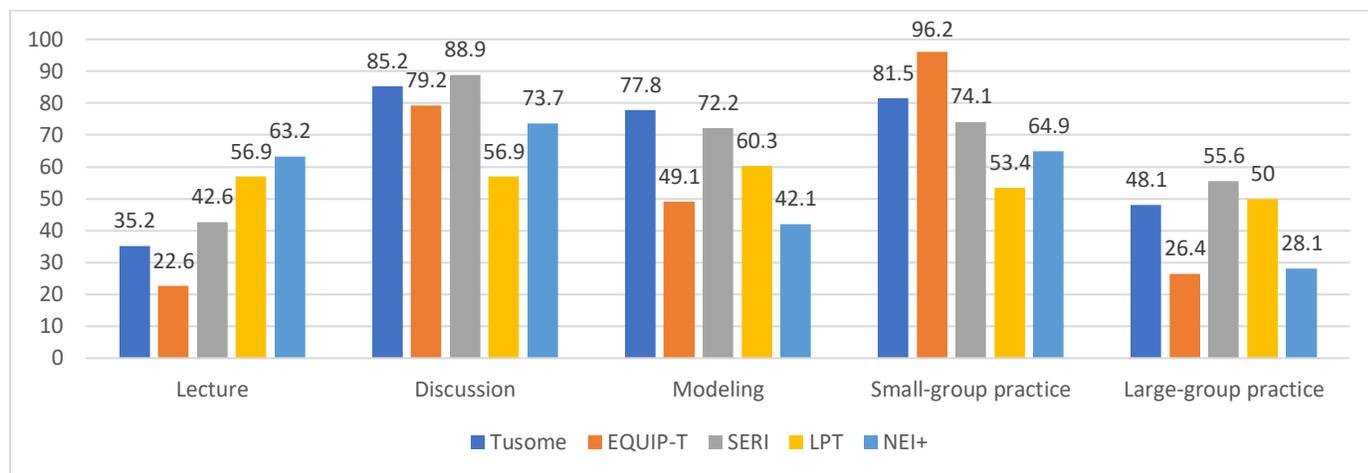
Coupled with the fact that teachers most often reported that discussion, small-group practice, and modeling were the most useful training methods (**Table 45**), it seems clear that these adjustments for program training sessions were well received and likely influenced program uptake.

Table 45. Most useful training methods, as reported by teachers, by program

Which [program] training methods did you find MOST useful? [Select only one]	Tusome	EQUIP-T	SERI	Lecture Pour Tous	NEI+
Lecture	1.9%	1.9%	7.4%	15.5%	24.6%
Discussion	16.7%	20.8%	33.3%	24.1%	26.3%
Modeling	46.3%	15.1%	29.6%	29.3%	8.8%
Small-group practice	33.3%	62.3%	14.8%	19.0%	28.1%
Large-group practice	1.9%	0	14.8%	12.1%	12.3%

In addition to training methods, we asked teachers about the training content that they received (**Table 46**). The most frequently reported training content area across all five programs was lesson plan development (reported by 52.6% to 90.7% of teachers) and

reading instruction (47.4% to 83.3%). This was followed by conducting student assessments (34.5% to 90.7%) and classroom management (31% to 85.2%).

Table 46.   Content areas for which teachers reported having received training, by program

Under [program], on which of the following did you receive training? (Mark all that apply)	Tusome	EQUIP-T	SERI	Lecture Pour Tous	NEI+
Reading instruction	83.3%	81.1%	50.0%	62.1%	47.4%
Lesson plan development	88.9%	73.6%	90.7%	67.2%	52.6%
Conducting student assessments	87.0%	81.1%	90.7%	34.5%	59.6%
Using student assessment results	55.6%	49.1%	77.8%	27.6%	45.6%
Classroom management	85.2%	79.2%	72.2%	31.0%	66.7%
Parental involvement	72.2%	49.1%	50.0%	17.2%	29.8%

As shown in **Figure 24,** there was relatively little consistency across the programs with regard to the training content areas that teachers reported as being the most important. This figure shows that 72.2% of teachers in Lecture Pour Tous, 49.1% of teachers in EQUIP-T, and 68.5% of teachers in Tusome reported that either reading instruction or lesson plan development was the most important. However, only a combined 25.9% of teachers in SERI and 19.3% of teachers in NEI+ believed that these instruction-based components were most important. Conversely, a total of 59.2% of SERI teachers and 54.4% of NEI+ teachers reported that the most important content area was related to student assessments.

Figure 24. **Most useful training content areas, as reported by teachers, by program**
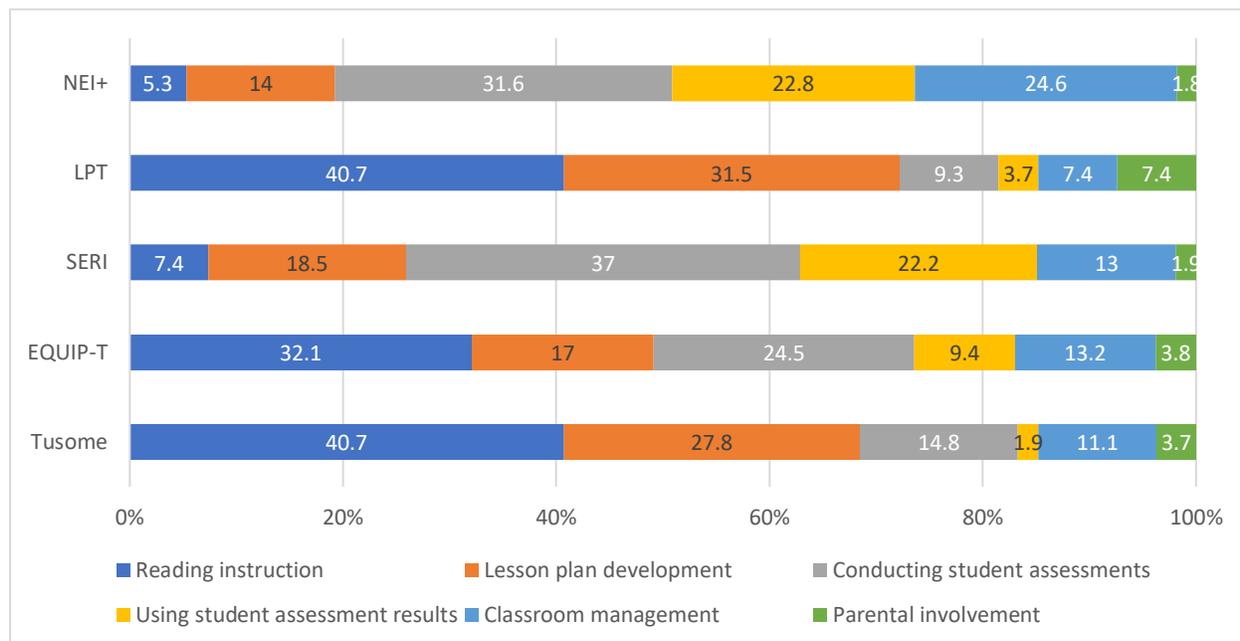
In addition to the previously mentioned "checking for understanding" that was regularly observed across classrooms (Table 44), nearly all teachers reported that they conducted assessments of students' reading ability in their classroom monthly or more frequently (**Table 47**). Furthermore, half of Tusome, NEI+, and EQUIP-T teachers said they assessed students daily, while three-quarters did so in SERI classrooms.

Table 47. **Frequency of student reading assessment, as reported by teachers, by program**

How often do you conduct assessments of students' reading ability in your classroom?	Tusome	EQUIP-T	SERI	Lecture Pour Tous	NEI+
Rarely or never	1.9%	0	0	6.9%	0
A few times per year	5.6%	3.8%	3.7%	3.4%	10.5%
Monthly	7.4%	24.5%	14.8%	3.4%	3.5%
A few times per month	27.8%	20.8%	7.4%	69.0%	33.3%
Daily	57.4%	50.9%	74.1%	17.2%	52.6%

In addition to training method and content area-specific questions, we also asked teachers what they believed the most important overall differences were between program training sessions and other teacher training sessions they had attended. The results of the most commonly reported factors are presented in **Table 48.** Although there was large variation in the responses across programs, training organization was the most highly reported for three

programs (i.e., Tusome with 44.4%; EQUIP-T with 54.7%; and SERI with 79.6%), and "more time to practice" was the most highly reported of two programs (i.e., Lecture Pour Tous with 34.5% and NEI+ with 56.1%). Other important differences in the training frequently cited by teachers across programs were "more time for discussion" and "materials are more relevant".

Table 48.   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 program training sessions and other training sessions? (Mark all that apply)	Tusome	EQUIP-T	SERI	Lecture Pour Tous	NEI+
Nothing	1.9%	3.8%	0	6.9%	0
More time to practice (individual, pair, group)	24.1%	30.2%	61.1%	34.5%	56.1%
Less lecture	13.0%	22.6%	20.4%	12.1%	10.5%
More time for discussion	29.6%	47.2%	68.5%	32.8%	38.6%
Training sessions are more frequent	20.4%	24.5%	37.0%	17.2%	38.6%
Training is better organized	44.4%	54.7%	79.6%	31.0%	52.6%
Trainers are better prepared and/or more knowledgeable	27.8%	43.4%	53.7%	8.6%	29.8%
Materials are more relevant and/or helpful	31.5%	41.5%	53.7%	36.2%	54.4%
More focus on specific reading skills	31.5%	34.0%	59.3%	10.3%	35.1%
Better allowances (transportation, per diem, etc.)	7.4%	15.1%	5.6%	6.9%	14.0%
Expectations are clear	13.0%	15.1%	14.8%	0	15.8%
Workload was manageable	1.9%	17.0%	18.5%	1.7%	12.3%
[Program] training sessions are worse	3.7%	3.8%	3.7%	1.7%	3.5%
Other	16.7%	9.4%	0	10.3%	0

Teacher and student materials

All five programs provided a range of materials to teachers and students. Unsurprisingly, teachers said that they found the teachers' guides, teaching aids, textbooks, and other student materials to be helpful in their classrooms. A more interesting question was how these materials compared with those being used prior to the five programs. For comparison, **Table 49** presents the perspectives across all give programs regarding teacher materials, and **Table 50** focuses on student materials. Generally, most teachers across all five programs reported that the materials were easier to follow, more engaging, and appropriate and/or enjoyable for students; and that the step-by-step instructions in teacher materials

were new for many teachers. Many teachers in EQUIP-T, SERI, and NEI+ also noted that teacher and student materials were better aligned to the curriculum than previously.

Table 49.   Comparison of program teacher materials with prior materials, as reported by teachers, by program

How do [program] teacher materials differ from what you were using before the program? (Mark all that apply)	Tusome	EQUIP-T	SERI	Lecture Pour Tous	NEI+
Better organized; easier to follow	74.6%	72.0%	66.1%	63.3%	89.8%
Step-by-step instructions	55.9%	40.0%	71.4%	38.3%	81.4%
Include more or better activities and examples	32.2%	36.0%	76.8%	3.3%	47.5%
Teaching aids keep students more engaged	28.8%	68.0%	75.0%	43.3%	42.4%
Manageable amount of content and materials	20.3%	16.0%	37.5%	15.0%	15.3%
Aligned with textbooks and/or curriculum	15.3%	46.0%	66.1%	6.7%	30.5%
Delivered on time	1.7%	8.0%	30.4%	1.7%	10.2%
Received sufficient quantity	13.6%	10.0%	33.9%	5.0%	10.2%
Other	10.2%	12.0%	0	10.0%	1.7%

Table 50.   Comparison of program student materials with prior materials, as reported by teachers, by program

How do [program] student materials differ from what you were using before the program? (Mark all that apply)	Tusome	EQUIP-T	SERI	Lecture Pour Tous	NEI+
Quantity (available for more or all students)	44.1%	18.4%	33.9%	40.0%	45.8%
Stories are more appropriate and enjoyable	50.8%	67.3%	94.9%	25.0%	67.8%
More attractive (e.g., illustrations, font, layout)	40.7%	57.1%	88.1%	43.3%	84.7%
Content is clearly presented; easy to follow	72.9%	55.1%	62.7%	53.3%	62.7%
Aligned with textbooks and curriculum	11.9%	36.7%	55.9%	10.0%	45.8%
Delivered on time	3.4%	4.1%	27.1%	0	8.5%
Other	13.6%	10.2%	0	8.3%	0

Coaching

External coaching was a core component of Tusome, NEI+, and SERI, while it was provided by a combination of school directors and school inspectors under Lecture Pour Tous and it

was not a main focus of EQUIP-T. As a result, findings on coaching visits (**Figure 25**) reflect the potential differences between using external, program-supported coaches and engaging actors within the system as coaches (beyond their usual job description). Visits from coaches were reported to have occurred more frequently in Tusome, NEI+, and SERI, with at least 86.5% of teachers reporting that they received visits more than once per year. A higher percentage of teachers reported receiving coaching at least once a month in SERI (57.5%) and NEI+ (72.2%) than in Tusome (26.9%). Under another question (results not shown here), the percentages of teachers who reported that coaching helped them to change their teaching was similar across NEI+ (94.4%), Tusome (88.5%), and SERI (87.5%) but was slightly lower in EQUIP-T (63.2%) and Lecture Pour Tous (45.5%).

Figure 25.   Frequency of coaching visits, as reported by teachers (percentage), by program
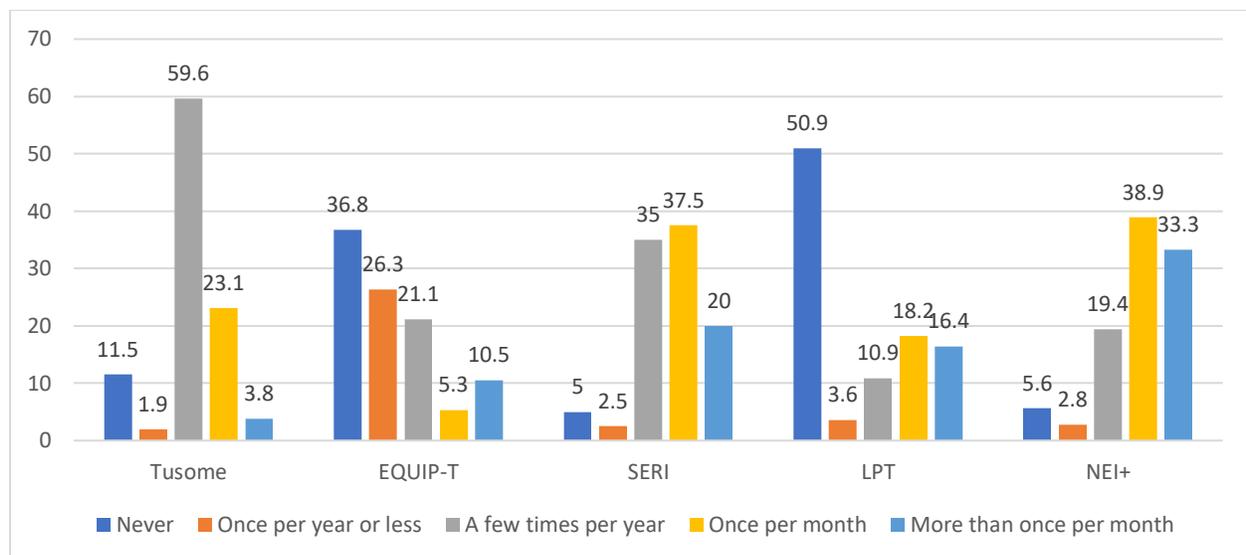
Teachers most commonly reported that the aspect of coaching that made them change their teaching was receiving guidance from coaches about how to teach (**Table 51**). Interestingly, one of the least commonly reported factors among SERI, NEI+, Lecture Pour Tous, and EQUIP-T teachers was receiving regular feedback from the coaches. This finding is further supported by the results presented in **Table 52**, which shows that most teachers reported not that the frequency of visits increased under the programs but that the quality and focus of those visits did. In another question (results not shown here), nearly 90% of teachers in SERI and Tusome cited that coaches observing their teaching was the most helpful activity during a typical coaching visit.

Table 51. **Coaching factors that impacted teaching, as reported by teachers, by program (response options NOT read aloud)**

What about the coaching made you change your teaching? (Mark all that apply)	Tusome	EQUIP-T	SERI	Lecture Pour Tous	NEI+
Regular feedback from coach	50.0%	25.0%	25.7%	32%	27%
Receiving guidance from coach on how to teach	76.1%	83.3%	91.4%	68%	100%
Asking coach questions about my teaching	37.0%	66.7%	65.7%	44%	44%
Coach modeling instruction	41.3%	41.7%	65.7%	44%	35%
Coach helps with lesson planning	13.0%	50.0%	77.1%	24%	35%
Coach helps with classroom management	26.1%	50.0%	71.4%	8%	44%
Other	8.7%	8.3%	0	0%	0%

Some of the most compelling information from teachers came in responses to a question about how their interactions with coaches differed under the program (as compared with before the program). Approximately two-thirds of teachers in three of the programs reported that program coaches were more supportive (**Table 52**). This aligns with qualitative findings on coaching (Section 5.5.3) which found a shift towards friendlier, more supportive interactions between teachers and coaches under all five programs. Other commonly cited differences included receiving more helpful feedback and receiving more suggestions on how to improve teaching.

Table 52. **Differences between coaching before and during programs, as reported by teachers (response options NOT read aloud)**

How are your interactions with the coach different than with coaches (or inspectors) before [program]? (Mark all that apply)	Tusome	EQUIP-T	SERI	Lecture Pour Tous	NEI+
More frequent visits	34.8%	8.3%	28.9%	7%	44%
Coaches are more supportive	71.7%	66.7%	68.4%	41%	47%
I have more opportunities to ask questions	19.6%	50.0%	55.3%	33%	41%
I receive more helpful feedback	41.3%	58.3%	63.2%	26%	50%
I receive more suggestions for how to improve my teaching	30.4%	66.7%	76.3%	19%	59%
Coaches are friendlier	67.4%	58.3%	44.7%	4%	6%
This is my first experience with a coach or inspector	6.5%	16.7%	15.8%	11%	29%
No differences	2.2%	0	23.7%	30%	0%

Although nearly 90% of teachers in NEI+ and over 60% of teachers in SERI and Lecture Pour Tous reported having received remote coaching and guidance (results not shown here), the same was true for less than 50% of teachers in Tusome and EQUIP-T. It is important to note once again that data from Lecture Pour Tous and NEI+ was collected after the COVID-19 outbreak, while data from SERI, Tusome, and EQUIP-T was collected in early 2020—before the outbreak. Across all five programs, teachers most commonly noted that the main value of remote support came in terms of reminding them to implement the program (40% to 85.2%). Between 50% and 100% of teachers in four of the programs also said the remote coaching made them feel supported. While a necessity during COVID, remote coaching support seemed less important in terms of actual instructional changes as a result of the coaching; instead, the advantage appeared to be related to communicating expectations and supporting teachers generally.

Teacher meetings

All five programs included some version of teacher meetings in their programming design, but these meetings were a particularly important aspect of EQUIP-T's model and theory of change. This focus is clear from the frequency of these meetings (**Figure 26**). Although a few teachers in all three programs reported that they never attended such meetings, 58.3% of teachers in EQUIP-T attended meetings more than once per month. Monthly meetings were frequently reported by teachers in Lecture Pour Tous (65.2%) SERI (55.6%), and NEI+ (39.3%).

Figure 26.   Frequency of teacher community-of-practice meeting attendance, as reported by teachers (percentage), by program
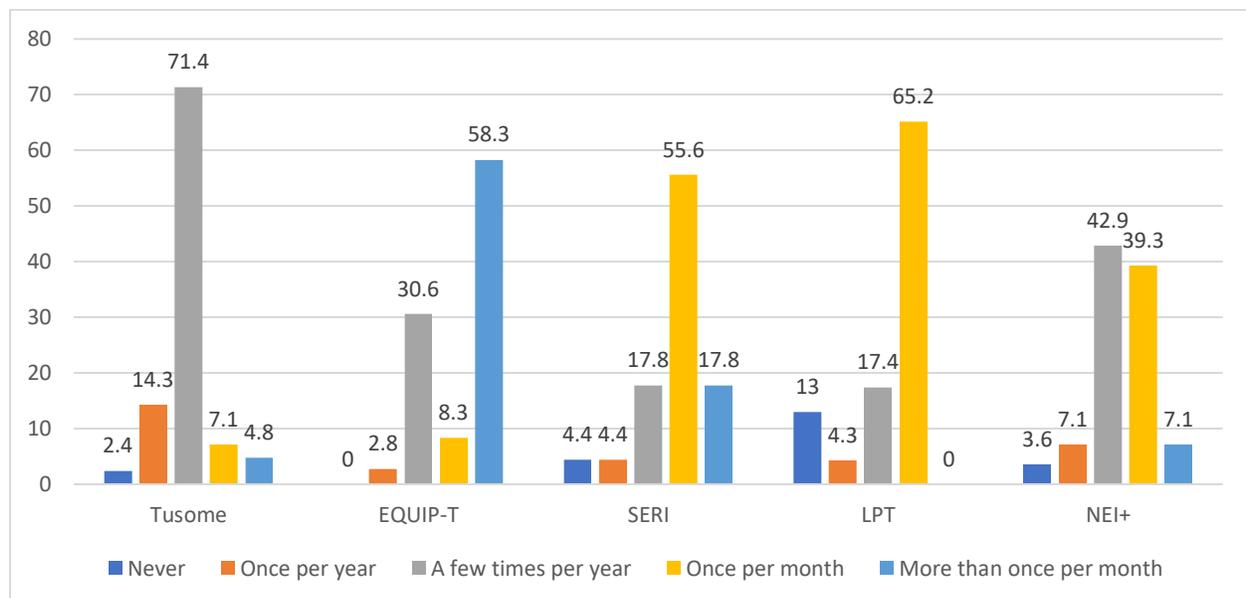
The most commonly cited aspect of these meetings that teachers across all programs found to be useful was simply having the opportunity for discussions with other teachers, with between 61% and 86% of teachers making this claim (**Table 53**). The teachers also said that feedback on how to improve and handle challenges (17.4% to 74.4%) and learning new information and/or approaches (21.7% to 81.4%) were relatively useful. Other aspects tended to be more idiosyncratic, with differing values across programs.

Table 53. **Most useful aspects of teacher meetings, as reported by teachers, by program**

What do you find to be useful from these meetings? (Mark all that apply)	Tusome	EQUIP-T	SERI	Lecture Pour Tous	NEI+
Feedback on how to improve or to handle challenges	43.9%	58.3%	74.4%	20%	59%
Discussions with other teachers	61.0%	66.7%	86.0%	80%	78%
Having someone answer my questions	14.6%	30.6%	69.8%	5%	56%
Time to practice teaching	7.3%	33.3%	79.1%	15%	59%
Learning new information and/or approaches	31.7%	55.6%	81.4%	25%	44%
Reminder, refresher of skills, information, and/or approaches	24.4%	52.8%	39.5%	15%	30%
Other	7.3%	5.6%	0	5%	0%

Approximately two-thirds of teachers in all three programs noted that they discussed their classroom instruction most of the time or always during the meetings (**Figure 27**).

Figure 27. **Frequency of discussion of own classroom instruction during teacher meetings, as reported by teachers (percentage), by program**
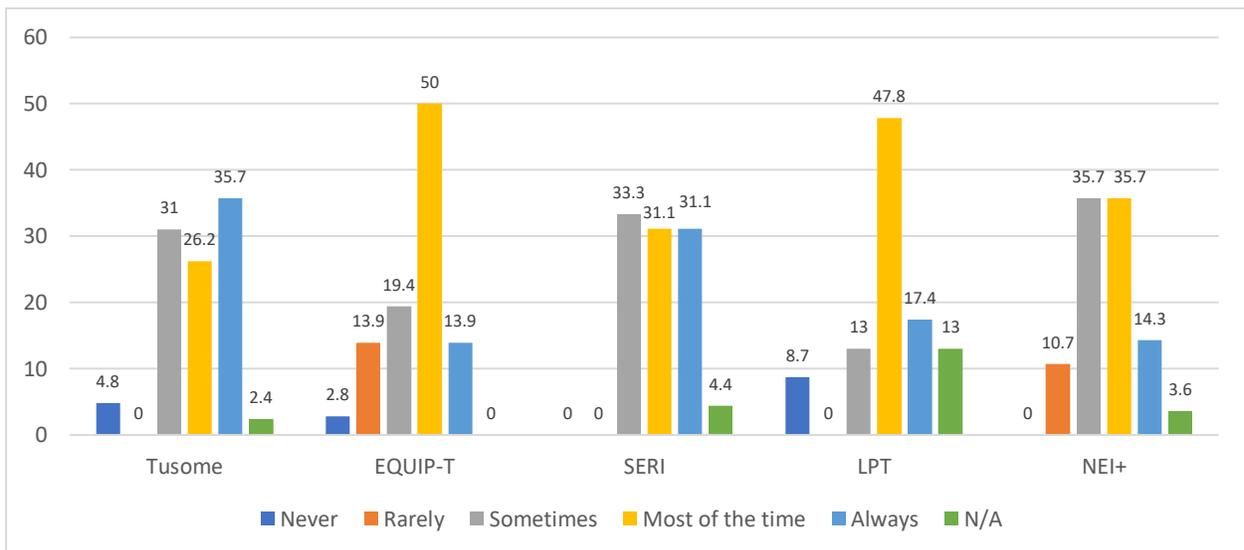
Finally, when the teachers across all three programs were asked about what improvements would make the meetings more useful to them, between 26.1% (Lecture Pour Tous) and 61.1% (EQUIP-T) reported that they would have preferred more frequent meetings (**Table 54**). In addition, 83.7% of teachers in SERI, 42.9% in NEI+, and 50.0% in EQUIP-T said that they would have benefited from more time to practice. In responses to other related questions (not shown), teachers in Tusome, SERI, and NEI+ reported that distance and transportation, as well as meeting schedules, were key obstacles to attending teacher meetings. However, 58.3% of teachers in EQUIP-T, 48.8% in SERI, and 34.8% in Lecture Pour Tous reported that there were no obstacles for attendance at all, likely because the teacher meetings were held at the school.

Table 54.   Improvements for teacher meetings, as reported by teachers, by program

What improvements do you think would make the meetings more useful to you? (Mark all that apply)	Tusome	EQUIP-T	SERI	Lecture Pour Tous	NEI+
More frequent meetings	43.9%	61.1%	39.5%	30%	56%
Larger allowances (e.g., transport and meals per diem)	63.4%	8.3%	4.7%	25%	67%
Smaller number of participants	7.3%	8.3%	55.8%	5%	11%
More time for practice	17.1%	50.0%	83.7%	25%	44%
More convenient timing/location	34.1%	19.4%	48.8%	30%	48%
More training and support for facilitators	26.8%	36.1%	41.9%	30%	48%
Nothing	2.4%	0	0	5%	0%
Other	19.5%	27.8%	0	30%	0%

Conclusion

Finally, we asked teachers across all three programs what they believed was the single most important aspect of the program for improving student learning. As shown in **Table 55,** the focus on phonics was in the top two for three of the programs and was the leading response for teachers in Tusome (22.0%) and SERI (54.2%). Across all five programs, the top three also included either student and/or teacher materials (teachers guides was the leading response for teachers in Lecture Pour Tous, while teacher and student materials tied as the leading response for teachers in NEI+). Interestingly, teachers from only one program (i.e., EQUIP-T, with 37.3%) selected training as a top aspect for what improved learning.

Table 55. **Most important program aspect for improving student learning, as reported by teachers, by program**

What would you say is the SINGLE MOST important aspect of this program for improving student learning overall?	Tusome	EQUIP-T	SERI	Lecture Pour Tous	NEI+
Focus on phonics	22.0%	16.9%	54.2%	6.7%	10.2%
Instructional change (other than phonics)	11.9%	5.1%	3.4%	15.0%	3.4%
Teachers' guide	6.8%	1.7%	10.2%	26.7%	16.9%
Teacher materials (other than teachers' guide)	6.8%	13.6%	11.9%	5.0%	20.3%
Textbooks	20.3%	8.5%	1.7%	11.7%	15.3%
Student materials (other than textbooks)	18.6%	8.5%	13.6%	13.3%	20.3%
Teacher training	8.5%	37.3%	3.4%	10.0%	11.9%
Coaching and/or mentoring (internal)	0	1.7%	0	1.7%	0
Coaching (external)	1.7%	0	0	1.7%	0
Communities of learning / practice	0	1.7%	0	1.7%	0
Parental and/or community engagement	0	0	1.7%	0	1.7%
Other	3.4%	5.1%	0	6.7%	0

1. Views from other stakeholders

Head teachers

The head teachers in all three programs reported that their program was successful in improving students' learning in their school (84.7% in EQUIP-T, 91.2% in SERI, 91.7% in Tusome, 76.5% in Lecture Pour Tous, and 59% in NEI+). Furthermore, the head teachers consistently reported that program training sessions were more effective than other in-service training that they had attended (82.5% in SERI, 88.3% in Tusome, 88.1% in EQUIP-T, 68.6% in Lecture Pour Tous, and 88.5% in NEI+).

We asked head teachers across all five programs to identify the type of information that had been communicated to them or to their schools about their respective program. The three types of information most consistently identified by the head teachers were: (1) the goals and objectives of the program, (2) the roles and responsibilities of teachers, and (3) official curriculum and materials (**Table 56**). This clear and consistent communication may have been important for program success and the ability of head teachers to reinforce the program for teachers at their schools.

Table 56.   Program information communicated to head teachers or to their schools, as reported by head teachers, by program

What type of information has been communicated to you or your school about [program]? (Mark all that apply)	Tusome	EQUIP-T	SERI	Lecture Pour Tous	NEI+
None	6.7%	3.4%	0	2.0%	1.6%
Goals and objectives of the program	58.3%	71.2%	93.0%	56.9%	68.9%
Roles and responsibilities of teachers	28.3%	69.5%	70.2%	41.2%	75.4%
Roles and responsibilities of the head teacher	5.0%	22.0%	47.4%	37.3%	32.8%
Official curriculum and materials	38.3%	33.9%	38.6%	45.1%	32.8%
Training expectations	46.7%	50.8%	22.8%	15.7%	47.5%
Expectations about student outcomes	28.3%	47.5%	42.1%	25.5%	21.3%
Policy changes regarding instruction	15.0%	20.3%	8.8%	7.8%	9.8%

We also asked head teachers to identify the changes that they had made as a result of participating in the program. The top three responses were that they did the following tasks differently: (1) monitor teachers' performance (51% to 83.1%), (2) provide more instructional support to teachers (51% to 65.0%), and (3) lead (more) teacher meetings and discussions (35.0% to 64.4%) (**Table 57**). In addition, approximately half of head teachers in both EQUIP-T and SERI noted that they now discipline teachers differently. Changing the management processes by head teachers in line with the program's objectives, particularly regarding monitoring teachers and providing more instructional support, may have been important. Note, of course, that these data cannot tell us whether head teachers actually were doing these activities more frequently.

Table 57. Changes that head teachers have made because of participating in their respective program, as reported by the head teachers, by program

What do you do differently as a head teacher as a result of [program]? (Mark all that apply)	Tusome	EQUIP-T	SERI	Lecture Pour Tous	NEI+
Nothing	1.7%	0	0	2.0%	0
Provide more instructional support to teachers	65.0%	61.0%	56.1%	51.0%	55.7%
Lead (more) teacher meetings and/or discussions	35.0%	64.4%	61.4%	58.8%	49.2%
Monitor teachers' performance	58.3%	83.1%	71.9%	51.0%	78.7%
Emphasize the importance of reading and instruction with teachers	38.3%	52.5%	57.9%	45.1%	50.8%
Discipline teachers	1.7%	42.4%	52.6%	3.9%	14.8%
Other	23.3%	28.8%	0	13.7%	1.6%

Coaches

As previously noted, coaches were not interviewed or observed in EQUIP-T because they did not utilize a coaching model. Between 67.4% (Tusome) and 90% (Lecture Pour Tous) of coaches across the four programs examined said they received training on coaching from the program; while between 25% (Lecture Pour Tous) and 91.3% (Tusome) said that they had received support from the program during their coaching visits. It is important to note that school level data were collected in Nigeria (NEI+) and Senegal (Lecture Pour Tous) in 2021, after protracted school closures and other disruptions as a result of COVID-19, which likely had an impact on teacher perspectives of supports that they received (prior to the pandemic).

Interestingly, although 100% of Tusome coaches and 85% of Lecture Pour Tous coaches reported that they observed a lesson (and 97.8% and 82.5%, respectively, held a debrief session with each individual teacher) during typical coaching visits, only approximately half of SERI teachers made the same claims. Although 73.7% of NEI+ coaches said they observe lessons during school visits, only 47.4% held debrief sessions with individual teachers. This appears to be a design difference between these programs. Across all four programs studied, more than three-quarters of coaches said that they either review lesson plans or check teachers' materials regularly. Further, more than one-half of coaches for Tusome, SERI, and Lecture Pour Tous also said they regularly check student materials.

To understand the importance using coaching data, we also asked the coaches what they typically do with the information they collect during their coaching visits (**Table 58**). Close to two-thirds of coaches in each of three programs (Tusome, Lecture Pour Tous, and NEI+)

said that they use the information during debrief meetings and discussions. Conversely, among coaches in SERI, Lecture Pour Tous, and NEI+, top responses included writing reports to the program, Ministry, or school. This highlights the competing uses for coaching data that many programs face, i.e., for immediate support within the school and for targeting support through the system.

Table 58. **Use of information collected during coaching visits, as reported by coaches in Tusome, SERI, Lecture Pour Tous, and NEI+ schools**

What do you do with the information you collect during a coaching visit? (Mark all that apply)	Tusome	SERI	Lecture Pour Tous	NEI+
Send written report to the program or Ministry	21.7%	68.2%	42.5%	47.4%
Provide written report and/or comments to the school	37.0%	54.5%	10.0%	47.4%
Use during debrief meetings and discussions	63.0%	31.8%	65.0%	68.4%
Use for follow-up visits	56.5%	36.4%	35.0%	31.6%
Feed into school improvement plans	32.6%	54.5%	25.0%	31.6%
Upload results	4.3%	4.5%	0	26.3%
Use during cluster meetings and/or training	58.7%	40.9%	12.5%	31.6%
Other	2.2%	0	0	0

One of the more compelling findings from the coach interviews was that a large majority of coaches in all four programs reported that the purpose of their coaching was to improve teaching in schools, from a low of 65.2% in Tusome to a high of 97.5% in Lecture Pour Tous (**Figure 28**). This finding aligns with the goal of multiple programs' to focus coaching on teacher support and move away from earlier models of (sometimes punitive) school inspection, and points to an important shift in attitudes for coaching models both external and working through the system.

Figure 28.   Purpose of coaching, as reported by coaches (percentage) in Tusome, SERI, Lecutre Pour Tous and NEI+ schools
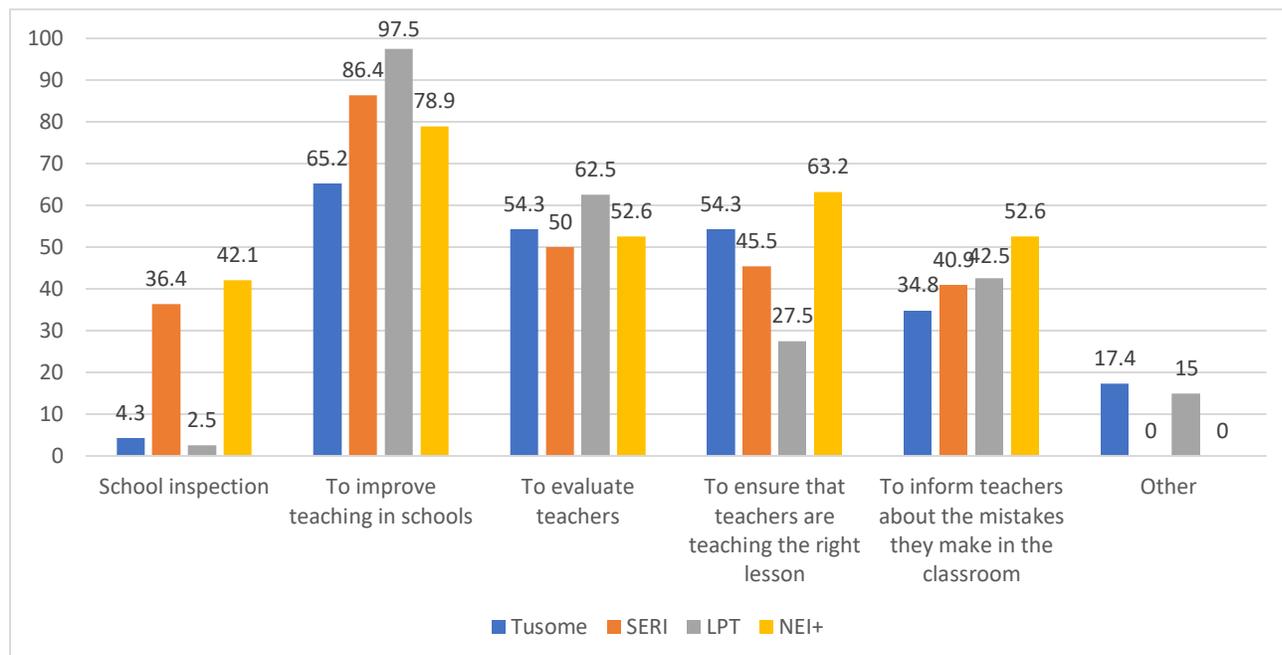
While nearly all coaches reported having received training as a coach, their views on what was most useful from the trainings varied somewhat (**Table 59**). More than 60% of Tusome, Lecture Pour Tous, and NEI+ coaches reported that the most useful aspect of the program was either how to give feedback to teachers or how to help teachers reflect on practice. However, none of the SERI coaches made these same claims. Instead, SERI coaches noted that the two most useful aspects of training were how to build rapport with teachers (54.5%) and the process of visiting schools (18.2%).

Table 59.   **Most useful aspect of training received, as reported by coaches in Tusome, SERI, Lecutre Pour Tous and NEI+ schools**

What was the most useful aspect of the training you received as a coach? [Select only one]	Tusome	SERI	Lecture Pour Tous	NEI+
How to give feedback to teachers	26.1%	0	10.0%	10.5%
How to help teachers reflect on practice	34.8%	0	55.0%	52.6%
How to build rapport with teachers	8.7%	54.5%	12.5%	15.8%
Process of visiting schools	0	18.2%	0	5.3%
How to use the observation tool	13.0%	4.5%	5.0%	5.3%
How to assess students	6.5%	9.1%	15.0%	0
How to prioritize which teachers and/or schools to visit	2.2%	4.5%	2.5%	10.5%
No training received	8.7%	9.1%	0	0

Although SERI coaches did not feel that training on providing teacher feedback was useful, they reported that this feedback was helpful for improving teaching (63.2% in **Table 60**). Similarly, 67.4% of coaches in Tusome and 73.7% in NEI+ schools said that the most helpful element was the ability to provide helpful feedback to teachers. Nearly all (90%) of coaches in Lecture Pour Tous schools said the post-observation discussion was most helpful, while approximately half of coaches in all programs also reported that regular follow-up visits and the act of being observed were useful to improve teaching.

Table 60.   **Most helpful items for improving teaching, as reported by coaches in Tusome, SERI, Lecutre Pour Tous and NEI+ schools**

What about the coaching is helpful for improving teaching? (Mark all that apply)	Tusome	SERI	Lecture Pour Tous	NEI+
Post-observation discussion	43.5%	63.2%	90.0%	42.1%
Providing teacher feedback (e.g., praise, improvement)	67.4%	63.2%	42.5%	73.7%
Being able to follow-up and visit regularly	45.7%	52.6%	42.5%	52.6%
Teachers improve when they are observed	56.5%	52.6%	45.0%	47.4%
Other	8.7%	0	5.0%	0

Based on the observations of coaching visits, it was clear that coaches in all four programs met with teachers before and after their reading lessons, but the approaches and topics of conversation varied (**Table 61**).

Table 61. Reported coaching visit practices from coaching observations, by program

Coach Observation	Tusome	SERI	Lecture Pour Tous	NEI+
Counts	*n = 26*	*n = 12*	*n = 33*	*n = 17*
1. Did the coach and teacher meet before the lesson?	84.6%	100%	90.9%	76.5%
2. Did the teacher and coach agree on goals and objectives for the lesson observation?	46.2%	91.7%	87.9%	76.5%
3. Did the teacher and coach discuss goals or areas of improvement from previous visit?	15.4%	50.0%	66.7%	64.7%
4. Did the coach examine the teachers' guide, the scheme of work, and/or the lesson plan?	65.4%	66.7%	78.8%	88.2%
5. Did the coach use a tool for observing the teacher?	96.2%	50.0%	90.9%	76.5%
6. Did the coach do any of the following during the lesson?				
Leave the classroom	15.4%	50.0%	9.1%	11.8%
Interrupt the lesson	7.7%	0	0	5.9%
Talk on the telephone and/or text	15.4%	40.0%	3.0%	5.9%
Distract students	0	10.0%	0	0
None of the above	69.2%	30.0%	72.7%	52.9%
7. Did the coach actively observe the entire lesson?	96.2%	50.0%	78.8%	58.8%
8. What materials or resources did the coach use during the lesson observation?				
None	0	10.0%	0	0
Coaching tool	19.2%	40.0%	45.5%	23.5%
Tablet	96.2%	0	3.0%	35.3%
Teachers' guide	26.9%	30.0%	45.5%	47.1%
Student book	15.4%	40.0%	24.2%	29.4%
Notebook	73.1%	0	57.6%	23.5%
9a. Coach conducted a student assessment at the end of the lesson	84.6%	58.3%	30.3%	23.5%
9b. Average number of students assessed	3	9	5	9
Section: Did the coach and teacher meet after the lesson?	96.2%	91.7%	90.9%	94.1%

Coach Observation	Tusome	SERI	Lecture Pour Tous	NEI+
10. Did the coach ask the teacher what he or she thought of the lesson?	73.1%	66.7%	75.8%	76.5%
11. Did the teacher ask questions about the lesson, activities, and how to help students?	57.7%	75.0%	69.7%	58.8%
12. Did the teacher talk about what he or she thought went well during the lesson?	65.4%	66.7%	75.8%	64.7%
13. Did the teacher talk about what he or she found to be challenging during the lesson?	65.4%	41.7%	72.7%	76.5%
14. Did the teacher talk about her or his students' engagement and learning?	53.8%	66.7%	60.6%	41.2%
15a. Did the coach give at least one area of praise?	96.0%	45.5%	70.0%	56.2%
15b. Average number of areas of praise that were given	2.5	3.0	3.4%	2.8%
16a. Did the coach give at least one area of improvement?	92.0%	54.5%	83.3%	62.5%
16b. Average number of areas of improvement that were given	3.7	3.3	2.2%	2.5%
17. Did the coach discuss strategies for improving instruction?	92.0%	63.6%	86.7%	87.5%
18. Did the coach model part of the lesson?	28.0%	45.5%	73.3%	56.2%
19. Did the coach show a video?	0	45.5%	0	12.5%
20. Did the coach refer to the lesson plan or teachers' guide in order to offer assistance?	72.0%	72.7%	83.3%	75.0%
21. Did the coach discuss how students were engaged?	92.0%	81.8%	70.0%	56.2%
22. Did the coach discuss whether students were understanding and/or learning the lesson content?	84.0%	81.8%	43.3%	62.5%
23. Did the coach and teacher discuss ways to help struggling students?	92.0%	72.7%	73.3%	81.3%
24. If the coach assessed students, did he or she discuss assessment results?	80.0%	63.6%	26.7%	31.3%
25. Did the coach discuss issues or improvements from previous visits?	44.0	72.7%	73.3%	68.7%
26. Did the coach and teacher discuss goals and progress made?	64.0%	54.5%	73.3%	87.5%
27. Did the coach and teacher agree on skills or practices to focus on moving forward?	92.0%	72.7%	86.7%	81.3%
28. Did the coach and teacher discuss next steps and/or set new goals?	84.0%	63.6%	76.7%	75.0%

Note: Green cells signify practices that occurred in at least 75% of the observed coaching visits.

Teacher meeting facilitators

Teacher meeting facilitator interviews and meeting observations were conducted only for Tusome, EQUIP-T, and SERI. The majority of meeting facilitators in all three programs reported that they received training as a meeting facilitator (78.6% in Tusome, 87.5% in SERI, and 96.8% in EQUIP-T). Furthermore, while the majority of facilitators for Tusome (73.7%) and EQUIP-T (94.2%) noted that teacher community-of-practice meetings were more helpful for teachers than previous meetings, 50% of SERI facilitators reported that previous meetings were more helpful for teachers. This may be explained in part by the fact that two-thirds of SERI facilitators reported that they were responsible for supporting too many teachers, while the same claim was made by 45.2% of facilitators in Tusome and only 12.9% in EQUIP-T. On a related note, the majority of SERI meeting facilitators (62.5%) noted that teachers typically had less than 30 minutes to practice instruction during teacher meetings; this was reported to be the case for only 31.0% of Tusome facilitators and 42.0% of EQUIP-T facilitators (**Figure 29**).

Figure 29. **Teacher practice time for instruction during teacher meetings, as reported by meeting facilitators, by program (percentage of meeting facilitators)**
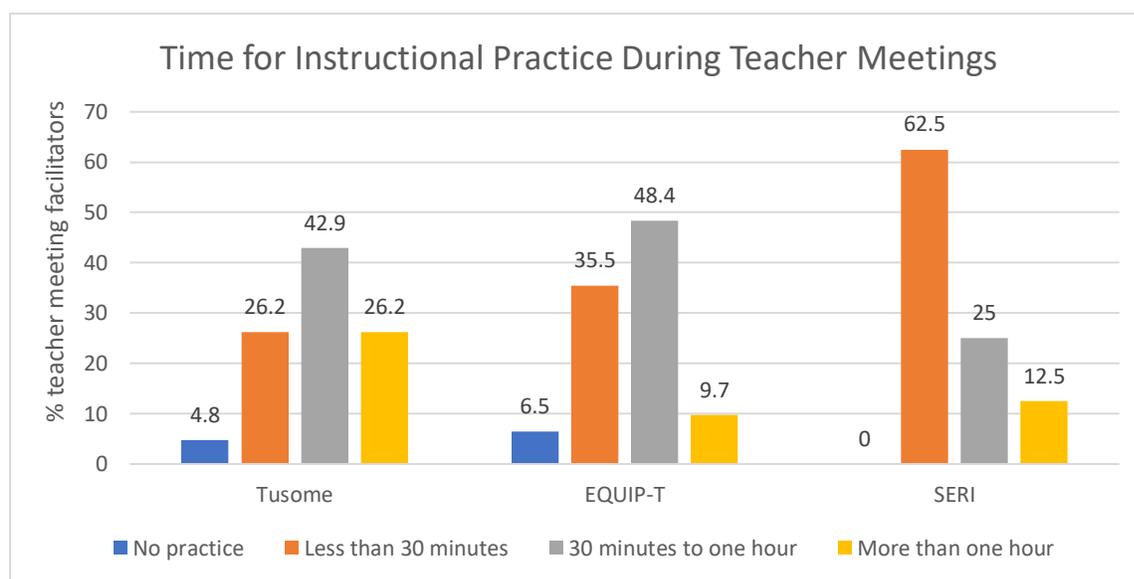
In EQUIP-T, where these meetings were a central part of the program, 90.3% of facilitators also noted that teachers attended all (or nearly all) of the meetings (while 61.9% of Tusome facilitators and only 25.0% of SERI facilitators said the same; see **Table 62**). This finding stems, in part, from the fact that meetings were held at the school and that meeting schedules in EQUIP-T were created by meeting facilitators, teachers, and head teachers. In Tusome and EQUIP-T, the meetings were held at one selected school within a cluster (i.e., not on site for all participants), and there was greater involvement from program staff and

155

government officials in creating the teacher schedules, which likely made it more difficult for teachers to attend.

Table 62.   The proportion of invited teachers who attend meetings, as reported by teacher meeting facilitators, by program

How many of the invited teachers attend these meetings (on average)?	Tusome	EQUIP-T	SERI
None	2.4%	0	0
Less than half	2.4%	0	0
Approximately half	2.4%	3.2%	62.5%
More than half	31.0%	6.5%	12.5%
All (or nearly all)	61.9%	90.3%	25.0%

Finally, we asked meeting facilitators what they felt were the most useful aspects to teachers of the teacher meetings. Meeting facilitators in EQUIP-T (83.9%) and Tusome (83.3%) most commonly noted the importance of feedback and learning from other teachers, which aligns with teacher-reported statements regarding the value of meeting with other teachers (**Table 63**). Despite the shorter opportunities to practice skills in SERI teacher meetings, the majority of facilitators (87.5%) felt that the time to practice was useful for teachers, along with being able to provide feedback and correct issues for a group of teachers (87.5%).

Table 63.   Useful aspects of teacher meetings, as reported by teacher meeting facilitators, by program

What about teacher meetings do you feel is useful for teachers? (Mark all that apply)	Tusome	EQUIP-T	SERI
Time to practice	31.0%	61.3%	87.5%
Teachers asking questions	47.6%	61.3%	62.5%
Feedback and learning from other teachers	83.3%	83.9%	37.5%
Frequency of meetings	11.9%	9.7%	12.5%
Being able to provide feedback and correct issues for a group of teachers	66.7%	51.6%	87.5%
Other	19.0%	16.1%	0

Teacher trainers

Teacher trainer interviews were conducted only by SERI, Lecture Pour Tous, NEI+, and Tusome data collectors. Nearly all interviewed trainers reported having received a training manual (100% in SERI, NEI+, and Lecture Pour Tous; and 91.3% in Tusome). When asked what made the training manual easy to use, trainers in Tusome, SERI, and NEI+ ranked the

step-by-step instructions as important (80% in SERI and 76.2% in Tusome; see **Table 64**).
SERI and Lecture Pour Tous trainers reported the importance of useful activities and
practice lessons (80% and 46.7%, respectively), while more than one-third of trainers in all
programs noted the clear explanation of concepts.

Table 64.   Factors that make the training manual easy to use, as reported by trainers, by program

What aspects of the training manual make it easy for you as a trainer? (Mark all that apply)	Tusome	SERI	NEI+	Lecture Pour Tous
Timetable	23.8%	50.0%	7.7%	20.0%
Step-by-step instructions	76.2%	80.0%	61.5%	33.3%
Useful activities and practice lessons	35.7%	80.0%	38.5%	46.7%
Clear explanations of concepts	61.9%	60.0%	38.5%	33.3%
Roles clearly defined	45.2%	30.0%	53.8%	40.0%
The entire manual is helpful	45.2%	90.0%	46.2%	60.0%
Other	9.5%	0	0	26.7%

In terms of training support, between 73.3% (Lecture Pour Tous) and 100% (SERI and
NEI+) of trainers said that program staff came to offer support during teacher trainings.
Trainers in Tusome, SERI, and NEI+ most often said that program staff would observe the
training or give feedback and advice. Trainers in SERI (80%) and Tusome (72.1%) also
noted that program support staff were very likely to help train attendees (**Table 65**).
Although program staff also tended to evaluate the trainings and have discussions with
teachers, these activities occurred to varying degrees across the programs.

Table 65.   Role of program support staff during teacher trainings, as reported by trainers, by program

What do [program support staff] do at the training? (Mark all that apply)	Tusome	SERI	NEI+	Lecture Pour Tous
Observe	76.7%	60.0%	84.6%	40.0%
Give feedback and advice	53.5%	50.0%	76.9%	26.7%
Help train	72.1%	80.0%	38.5%	26.7%
Evaluate your training	27.9%	70.0%	61.5%	20.0%
Have discussions with teachers	46.5%	90.0%	61.5%	26.7%
Other	9.3%	0	0	26.7%

Teacher trainers noted a range of factors impacting their ability to adhere to the scheduled timetable for trainings (most notably, teachers not being on time). When it came to addressing this issue, the most commonly reported action from trainers was to extend the time for the training (e.g., stay late, start early in subsequent days, shorten break times) or to continue on to get through as much as possible (**Table 66**).

Table 66. **Action taken if training activities took longer than intended, as reported by trainers, by program**

What do you do if something takes longer than the timetable allows?	Tusome	SERI	NEI+	Lecture Pour Tous
Nothing	2.2%	20.0%	30.8%	0
Skip or shorten the next activity	15.2%	20.0%	7.7%	6.7%
Keep going and get through as much as possible	43.5%	50.0%	30.8%	20.0%
Do not complete all of the training topics	2.2%	0	7.7%	13.3%
Consult a program staff member	4.3%	30.0%	15.4%	20.0%
Stay late, start early, and/or shorten break time	54.3%	50.0%	30.8%	40.0%
Other	15.2%	0	0	26.7%

In the instances when activities took less time than expected, trainers in all programs were most likely to give more time for practice (100% in SERI and 47.8% in Tusome) or ask whether there were questions (53.3% in Lecture Pour Tous and 53.8% in NEI+). Other factors, which ranged in likelihood across programs, are presented in **Table 67.**

Table 67. **Action taken if training activities took less time than intended, as reported by trainers, by program**

What do you do if something takes less time than scheduled? (Mark all that apply)	Tusome	SERI	NEI+	Lecture Pour Tous
Nothing	4.3%	0	0	0
Move on to the next topic and/or activity	26.1%	20.0%	23.1%	26.7%
Ask whether there are questions	45.7%	50.0%	53.8%	53.3%
Talk more about the topic	26.1%	60.0%	61.5%	40%
Give more time for practice	47.8%	100.0%	38.5%	33.3%
Consult with program staff	0	60.0%	7.7%	0
Finish early	13.0%	0	15.4%	6.7%
Other	10.9%	0	0	6.7%

In terms of training approaches, at least 95% of trainers in all four programs reported that during trainings, teachers practiced new skills (both in front of large groups and in small groups or pairs). Although skill practice was the most commonly reported method used in teacher training for SERI and NEI+ trainings (50% and 46.2%, respectively, combined across small-group practice and in-school practice), practice was a relatively small component of the Tusome and Lecture Pour Tous trainings, according to trainers (**Table 68**). Instead, 63% of Tusome trainers reported that they most frequently used modeling and/or demonstration in trainings, which was also claimed to be the most commonly used training method by 40% of SERI trainers and 46.7% of Lecture Pour Tous trainers. Discussions were also frequently used by 40% of Lecture Pour Tous trainers.

Table 68.   Most commonly used training method, as reported by trainers, by program

What training method do you use most in program training sessions?	Tusome	SERI	NEI+	Lecture Pour Tous
Lecture	10.9%	0	15.4%	0%
Discussion	13.0%	10.0%	30.8%	40.0%
Modeling and/or demonstration	63.0%	40.0%	15.4%	46.7%
Small-group practice	6.5%	20.0%	30.8%	13.3%
Large-group practice	6.5%	0	0	0
In-school practice	0	30.0%	15.4%	0

Finally, we asked trainers to report on what they felt was the most useful training method for giving teachers an opportunity to learn a new instructional approach. As shown in **Figure 30,** modeling was the most commonly reported option by three of the programs (76.1% in Tusome, 46.7% in Lecture Pour Tous, and 50.0% in SERI). Small-group practice, discussion, and large-group practice were the next most common responses. Interestingly, no SERI trainers selected in-school practice, despite 30.0% of them noting that it was the most frequently used method.

Figure 30. **Most useful training method, as reported by trainers (percentage), by program**
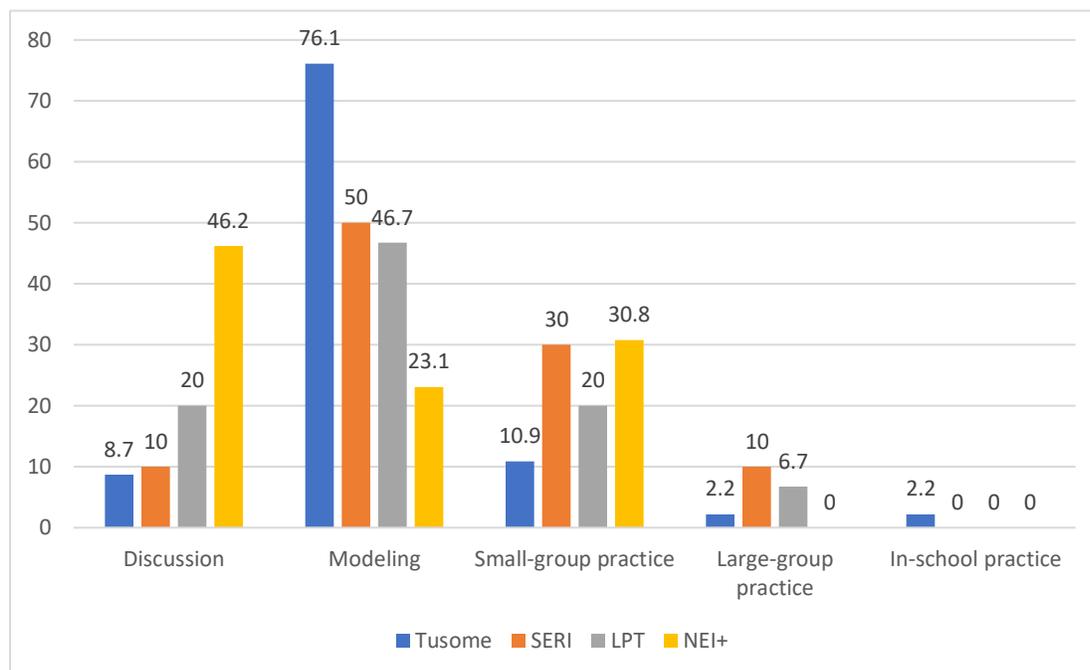
Discussion of descriptive quantitative findings

In this section, we present the discussion of our descriptive quantitative findings. These findings were derived from the classroom observations, teacher interviews, head teacher interviews, coach interviews, and teacher meeting facilitator interviews. The text below presents how some of these findings relate to learning outcomes, but this section is primarily descriptive. Our key takeaways from this section are as follows:

While we found substantial overlap in the implementation of programs, as well as the supports that teachers found most useful, there were also key differences in how class time was used. These differences tended to align with differences in program designs (as described in Section 5.3).

Some of the most interesting evidence we have relates to what teachers in these five programs were doing in classrooms. Our key takeaways are the following:

- – **Phonics.** Tusome, NEI+, Lecture Pour Tous and SERI teachers pointed to a focus on letters, sounds, and blending—the hallmarks of a phonics-based approach—as the most essential instructional element for program impact. Similarly, the teachers in all five programs saw these phonics components as a differentiator from previous programs in their countries.
- – **Teaching methods.** All five programs had large percentages of teachers cite the program's new methodology and instructional approach as responsible for improved learning outcomes. The impacts of the programs were related in part to the view of teachers that the program worked, although our analysis cannot determine the directionality of this relationship.

- **Student materials.** These programs depended heavily on student textbooks or other written materials, and these resources were seen as essential for program success. Tusome, SERI and Lecture Pour Tous used student textbooks nearly half of the time, with another one-third to one-half of instructional time spent using written materials (developed by either the teacher or students). EQUIP-T was different, with the vast majority of time spent using written (often teacher-made) materials. Interestingly, NEI+ showed significantly less time with students engaging with written materials. However, all five programs heavily emphasized the use of student materials as essential to the program's impact on student learning. Over 70% of Tusome, NEI+ and Lecture Pour Tous teachers saw the textbooks as the most useful student material, while more teachers in EQUIP-T and SERI cited supplementary readers, exercise books or other stationary as being most useful.

- **Teacher materials and teachers' improved instruction** were also seen as essential for improved student learning by teachers.

- **Whole-class instruction** was used at least 76% of the time in four of the five programs.. These findings suggest that programs can succeed if they more effectively teach the entire class, though it is important to note that teachers in all five programs created time for students to practice on their own.

- **Students' time on task.** These five programs had high percentages of class time where most, if not all, students were on task. The structure of the interventions made this possible, and it is possible that increasing this pedagogical time was a meaningful part of why the programs worked.

- **Reading enjoyment.** Teachers saw these five programs as helping students to enjoy reading more and to be more engaged.

- **Teaching strategies.** We found in our post-observation checklist that Tusome, NEI+, Lecture Pour Tous and SERI used many of the same strategies. In nearly all lessons in these programs, teachers demonstrated the skill for the student; used decoding strategies; supported students with reading accuracy, reading rate, and prosody; and used activities focused on understanding the meaning of the text. In addition, in the vast majority of classrooms, students read something on their own, and teachers circulated throughout the room and were responsive to student progress.

We asked teachers and other educators in the system how they perceived the effects of the programs and if they could determine the causal pathways through which the programs showed impact. The findings from these items are below:

- **Benefits of training.** Teacher training was seen as the most significant support provided to teachers in four of the five programs, with teachers in Lecture Pour Tous most often citing teacher materials as the most useful support. Tusome and NEI+ teachers saw teacher materials (e.g., lesson plans) as the next most important; EQUIP-T teachers saw teaching aids as the next most supportive resource; SERI teachers saw student materials as the next most supportive; and Lecture Pour Tous teachers saw training as the next most supportive resource. Clearly, teacher training, often criticized for being ineffective, can be a critical part of effective programs if done well.

- **Characteristics of program training.** All five programs were different from previous training programs in terms of a few key characteristics. Primarily, the programs incorporated significantly more discussion time and substantially more

small-group practice. In addition, participants in Tusome and SERI trainings spent more time modeling new skills and having large-group practice.

– **Most useful content.** EQUIP-T, Tusome and Lecture Pour Tous teachers saw training on reading instruction methods as most useful, while more NEI+ and SERI teachers preferred the training on conducting student assessments.

– **Logistics.** Interestingly, when asked why the programs' trainings were different from previous ones, participants in all five programs frequently cited the fact that the trainings were better organized compared to previous programs. This basic logistical competency of the programs is worth emphasizing for future programs.

– **Better teacher materials.** Teacher materials were seen by the majority of teachers as being different from previous materials because they were better organized and easier to follow. Tusome, NEI+ and SERI teachers also noted that the programs' teacher materials were different because they included step-by step instructions.

– **Student materials.** The difference related to student materials noted most frequently by teachers across the five programs was that the stories were more appropriate and enjoyable, that content is clearly presented that materials were more attractive.

– **Coaching.** Coaching frequency varied across the programs, in line with their design. The largest portion of SERI and NEI+ teachers reported that they were coached monthly, while Tusome teachers said they were coached a few times per year. All three of these programs prioritized external coaching in their implementation (more so than EQUIP-T and Lecture Pour Tous—both of which also had Learning at Scale data collection activities after COVID-related school closures). Importantly, teachers in all five programs noted that coaching impacted teaching by providing guidance on how to teach and that coaches were more supportive and friendlier than they were before.

– **Teacher meetings.** Teacher community-of-practice meetings happened more than once a month in EQUIP-T, monthly in SERI and Lecture Pour Tous, and a few times per year in Tusome and NEI+. Teachers in all five programs noted that the meetings were useful because they provided an opportunity to discuss teaching with other teachers and to receive feedback about how to handle pedagogical challenges.

– **Role of coaches.** We found that coaches in Lecture Pour Tous, NEI+, SERI and Tusome perceived their primary role as improving teaching in schools rather than implementing more administrative or inspection tasks.

– **Coaching activities.** Observations of coaching visits for Lecture Pour Tous, NEI+, and SERI revealed that the programs used similar models, with Tusome employing a somewhat different model. In all four programs, coaches and teachers typically met before the lesson, but in SERI, Lecture Pour Tous and NEI+ the teacher and coach agreed on the goals before the lesson; in contrast, Tusome coaches used a specific tablet-based tool for observation and observed the entire lesson. More Tusome coaches conducted student assessments at the end of the lesson than did other coaches.

– **Post-observation meetings.** Coaches in all programs examined met with teachers after observed lessons and discussed student engagement and student understanding.

We asked teachers to identify what they saw as the program component that was most essential for student learning.

- **The largest percentage of Tusome and SERI teachers noted that the focus on phonics was the single most effective program component for improving student learning,** while teachers in NEI+ and Lecture Pour Tous most consistently reported that **teachers' guides were the single most important aspect of the program.** EQUIP-T teachers identified **training** most frequently.

While the classroom observational and interview data we report here do not represent the full picture across all eight featured programs, we were encouraged to see some specific program design and instructional activities that appear to be related to program success.

5.4.2 Overview of instruction and student performance

Instructional practices and student performance

In this section of the report, we present our findings regarding the relationship between teachers' instructional practices (assessed using the classroom observation instrument) and students' reading ability (assessed using the EGRA). These analyses focus on treatment schools only, in an effort to better understand what instructional practices within programs are linked with higher levels of reading performance.

All analyses in this section of the report are based on linear regression models. Because effectiveness for all programs was defined in terms of improved ORF (as a part of the inclusion criteria for this study), we used ORF as our outcome variable in all of the models.

The regression result tables in this section of the report contain six columns (from left) that identify the (1) variable names with coefficient estimates for (2) Tusome,[28] (3) EQUIP-T, (4) SERI, (5) Lecture Pour Tous, and (6) NEI+. Coefficients for all regression models are bolded, and standard errors are displayed below the coefficients, in parentheses. All models controlled for class size and included clustered standard errors at the school level. We decided not to use the control schools given the complexity of the comparisons.[29] Statistically significant coefficients are signified by the inclusion of an asterisk (*) next to the coefficient (in which one asterisk is marginally significant, and two or three asterisks represent at least 95% confidence in the statistical significance).

---

[28] We collected classroom observation and student performance data in both Kiswahili and English for the Tusome program. For simplicity, the results in this section of the report focus only on the English observation data and students' results.
[29] Note that for some of the analyses, the EQUIP-T control schools outperformed the treatment schools. This comparison was not done to review the EQUIP-T impact evaluation, but this surprising result caused us to reconsider the implications of the comparisons between treatment and control in Tanzania and elsewhere.

Instructional focus

The classroom observation instrument defines "instructional focus" (or [Instruction]) as the skill that is the focus of instruction at that moment. The instructional focus options were as follows: Reading, Writing, Oral Language, PA, Grammar, Assessment, Management, or Nothing. The average time spent on each instructional area during the observed lesson is displayed in **Table 69.** Four of the five programs spent the largest proportion of time on Reading, ranging from 29.7% in NEI+ to 66.4% in Lecture Pour Tous. NEI+ teachers spent the most time on oral language (31.6%), while EQUIP-T treatment teachers spent a significant amount of time on Phonological Awareness (29.9%).

Table 69.   Percent of observed class time spent on various instructional activities, by program

Variable name	SERI	EQUIP-T	Tusome	Lecture Pour Tous	NEI+
Reading (R): % of [Instruction]	47.7%	33.8%	54.9%	66.4%	29.7%
Writing (W): % of [Instruction]	23.4%	16.5%	2.0%	8.9%	14.9%
Oral Language (OL): % of [Instruction]	14.5%	8.8%	16.4%	10.8%	31.6%
Phonological Awareness (PA): % of [Instruction]	5.3%	29.9%	8.8%	1.5%	11.6%
Grammar (G): % of [Instruction]	1.2%	0.1%	7.4%	0.3%	0.5%
Assessment (A): % of [Instruction]	4.6%	2.6%	5.1%	1.6%	7.8%
Management (M): % of [Instruction]	2.8%	7.9%	4.3%	8.4%	3.4%
Nothing (N): % of [Instruction]	0.4%	0.4%	1.2%	2.1%	0.5%

We then regressed each of these instructional focus areas separately on oral reading fluency scores (controlling for class size). Because assessors specifically observed reading lessons taking place in literacy programs—all sought to provide students with greater access to printed materials—we used the Reading group as the point of reference. In other words, the coefficients for each variable in **Table 70** are relative to the time spent on Reading. The hypothesis was that while some variation in activities is important, reducing time spent on Reading to focus on other instructional areas is unlikely to positively impact student performance on reading.

No instructional focus was significant across all five programs but Grammar was significant for three programs (i.e., increases in time spent on Grammar relative to Reading were associated with lower ORF scores). However, with less than 2% of class time in EQUIP-T and SERI spent on Grammar, this finding must be interpreted with caution. For EQUIP-T, SERI and Lecture Pour Tous, it was also shown that increased time on Assessment was associated with lower reading scores, which is interesting given the focus of these programs

on increasing checking for student understanding. It may be that increasing the duration of the Assessment time reduces its impact on learning, particularly since it comes as a trade-off to more time spent on Reading. The magnitudes of these impacts are still relatively small (with a 10 percentage point increase in time on Assessment leading to a 1.61 cwpm decrease for EQUIP-T, a 3.80 cwpm decrease for SERI and a 2.08 cwpm decrease for Lecture Pour Tous). Longer Management time was also negatively associated with reading scores in EQUIP-T, which potentially points to the importance of efficient classroom time utilization. We also found that increased time on PA in SERI was marginally associated with higher reading scores. Overall, these models account for a very small proportion of the variation in student ORF scores (as shown by the small R-squared values at the bottom of the table).

Table 70. Ordinary least squares regression of instructional practices on ORF, by program

Variable Names	Results				
	SERI	EQUIP-T	Tusome	Lecture Pour Tous	NEI+
Reference: Reading (R): % of [Instruction]	—	—	—	—	—
Writing (W): % of [Instruction]	0.123	0.057	0.245	-0.058*	-0.072
	-0.115	-0.046	-0.27	-0.033	-0.089
Oral Language (O): % of [Instruction]	-0.119	0.001	0.067	-0.014	0.098
	-0.132	-0.08	-0.081	-0.031	-0.088
Phonological Awareness (PA): % of [Instruction]	0.361*	0.035	0.144	0	-0.005
	-0.214	-0.029	-0.166	-0.047	-0.062
Grammar (G): % of [Instruction]	-1.045***	-3.824***	-0.261**	0.419*	-0.402
	-0.332	-0.654	-0.124	-0.245	-0.252
Assessment (A): % of [Instruction]	-0.380**	-0.161**	0.158	-0.208***	-0.017
	-0.159	-0.063	-0.191	-0.069	-0.108
Management (M): % of [Instruction]	0.083	-.110**	-0.421	0.003	-0.119
	-0.301	-0.055	-0.268	-0.074	-0.103
Class Size	0.141	-0.01	0.042	-0.009	0.005
	-0.172	-0.016	-0.087	-0.041	-0.039
_cons	26.545***	9.957***	17.888***	6.619***	2.508
	-5.519	-2.826	-3.868	-1.97	-4.407

Variable Names	Results				
	SERI	EQUIP-T	Tusome	Lecture Pour Tous	NEI+
Observations	885	944	920	909	776
R-squared	0.069	0.066	0.043	0.027	0.066

Notes: The "Nothing" category is suppressed from this output.
Standard errors are in parentheses.
*p < .1; **p < .05; ***p < .01.

Language component

We defined the "language component" (or [Language Part]) as the unit of language that was being taught or referenced during instruction in the classroom observation. The options were as follows: Story, Sentence, Word, Word Part, Letter, Sound, and Other. SERI, Tusome and NEI+ the most time at the Story level, while EQUIP-T and Lecture Pour Tous teachers had the largest focus on Word Part (**Table 71**). These variations may be explained in part by the time in the school year that students were assessed.

Table 71.   **Percent of observed class time spent on various language components, by program**

Variable name	SERI	EQUIP-T	Tusome	Lecture Pour Tous	NEI+
Story (ST): % of [Language Part]	28.6%	6.4%	30.5%	2.3%	24.3%
Sentence (SE): % of [Language Part]	19.8%	4.3%	13.5%	6.7%	17.5%
Word (W): % of [Language Part]	19.6%	28.5%	32.8%	15.3%	18.6%
Word Part (WP): % of [Language Part]	17.4%	32.5%	9.5%	29.5%	12.4%
Letter (L): % of [Language Part]	9.0%	9.2%	2.2%	31.1%	14.6%
Sound (SO): % of [Language Part]	5.3%	14.7%	7.2%	5.6%	11.0%
Other (O): % of [Language Part]	0.3%	4.4%	4.3%	9.6%	1.7%

Because all of the Learning at Scale programs assessed through primary quantitative data collection were using a phonics-based approach and were focused on improving reading with connected text, the regression models displayed in **Table 72** used Word as the reference group. The hypothesis was that successful programs will spend relatively less time at the word level when providing reading instruction and that the trade-off of spending more time on other language parts will be beneficial for student learning.

Interestingly, there were no significant factors in the Tusome, Lecture Pour Tous or NEI+ models. For EQUIP-T, an increased focus on Word Part (i.e., phonics, particularly at the syllable level in Kiswahili) was associated with higher ORF scores. In India, the SERI results showed that students appeared to benefit from an increased focus on time spent teaching at

the Letter level. These findings generally seem to be in line with programs where a focus on a phonics-based approach to learning was associated with higher learning outcomes.

Table 72.   Ordinary least squares regression of language components on ORF, by program

Variable Names	Results				
	SERI	EQUIP-T	Tusome	Lecture Pour Tous	NEI+
Reference: Word (W): % of [Language Part]	—	—	—	—	—
Story (ST): % of [Language Part]	**0.16**	**0.102***	**0.036**	**-0.042**	**0.028**
	-0.11	-0.054	-0.106	-0.053	-0.046
Sentence (SE): % of [Language Part]	**0.119**	**-0.065**	**0.038**	**-0.045**	**0.169**
	-0.147	-0.05	-0.13	-0.047	-0.106
Word Part (WP): % of [Language Part]	**-0.029**	**0.060****	**0.343**	**-0.022**	**-0.029**
	-0.143	-0.027	(0.219)	-0.056	-0.078
Letter (L): % of [Language Part]	**0.453***	**0.007**	**0.242**	**-0.036**	**-0.006**
	-0.169	-0.04	-0.451	-0.043	-0.045
Sound (SO): % of [Language Part]	**-0.139**	**0.019**	**0.046**	**-0.023**	**0.112**
	-0.167	-0.036	-0.214	-0.058	-0.086
Other (O): % of [Language Part]	**0.817**	**-0.221****	**-0.286**	**0.045**	**0.096**
	-1.137	-0.09	-0.188	-0.069	-0.12
Class Size	**0.007**	**0.007**	**-0.005**	**-0.019**	**0.021**
	-0.176	-0.024	-0.083	-0.042	-0.041
_cons	20.695**	7.444**	15.148*	8.222*	-1.432
	-8.935	-3.012	-7.569	-4.187	-4.476
Observations	885	944	920	909	776
R-squared	0.064	0.04	0.032	0.009	0.067

Note: Standard errors are in parentheses.
*$p < .1$; **$p < .05$; ***$p < .01$.

Materials

We defined "materials" (or [Materials]) in **Table 73** as an item that was the subject of discussion during the classroom observation (either the material that the teacher was referring to or the material that students were paying attention to). The options were as follows: Book, Written, Oral, Illustration, or Nothing. SERI (44.8%) and Tusome (46.6%) teachers each spent the largest proportion of time using books in the classroom, while books were the second most frequently used material in Lecture Pour Tous (41.7%) and NEI+ (27.9%). EQUIP-T teachers spent the most time using other written materials (76.1%), while Lecture Pour Tous saw the greatest usage of Written materials (46.7%) and NEI+ showed the largest usage of oral language (45.9%).

Table 73. **Percent of observed class time spent uses various materials, by program**

Variable name	SERI	EQUIP-T	Tusome	Lecture Pour Tous	NEI+
Book (B): % of [Materials]	44.8%	8.4%	46.6%	41.7%	27.9%
Written (W): % of [Materials]	28.0%	76.1%	32.6%	46.7%	22.2%
Oral (O): % of [Materials]	23.7%	10.4%	17.0%	7.0%	45.9%
Illustration (I): % of [Materials]	3.3%	2.7%	1.4%	0	3.2%
Nothing (N): % of [Materials]	0.3%	2.4%	2.4%	4.5%	0.8%

With such a large focus on books (textbooks and/or supplemental readers) across the programs, we used Book as the reference group for the models below. The hypothesis was that a more limited focus on books or materials being read by students will have an adverse effect on the overall performance of students.

Although there were no significant coefficients in the SERI, Lecture Pour Tous or NEI+ models, both the Tusome and EQUIP-T models showed that when teachers spent more time teaching without a specific reference material, students' scores tended to be lower (**Table 74**). Additionally, based on Tusome results, using pictures (as opposed to books) was negatively associated with ORF scores. Both factors in the Tusome models were only marginally statistically significant. The primacy of the textbook as the key Material useful in improving outcomes seems to be reinforced by these findings.

Table 74. **Ordinary least squares regression of materials on ORF, by program**

Variable Names	Results				
	SERI	EQUIP-T	Tusome	Lecture Pour Tous	NEI+
Reference: Book (B): % of [Materials]	—	—	—	—	—
Written (W): % of [Materials]	**0.012**	**0.015**	**-0.081**	**-0.037**	**0.002**
	-0.104	-0.035	-0.123	-0.029	-0.05
Oral (O): % of [Materials]	**-0.1**	**0.033**	**-0.116**	**-0.047**	**0.097**
	-0.142	-0.088	-0.115	-0.047	-0.078
Illustration (I): % of [Materials]	**0.014**	**-0.042**	**-.679***		**-0.162**
	-0.289	-0.038	-0.385		-0.154
Nothing (N): % of [Materials]	**1.607**	**-0.178****	**-.562***	**0.067**	**-0.009**
	-1.499	-0.083	-0.309	-0.112	-0.334
Class Size	**0.064**	**-0.002**	**0.006**	**-0.014**	**0.026**
	-0.206	-0.016	-0.102	-0.042	-0.038

Variable Names	Results				
	SERI	EQUIP-T	Tusome	Lecture Pour Tous	NEI+
_cons	30.888***	9.13***	26.315***	7.951***	-0.979
	-6.693	-3.055	-5.67	-1.865	-4.289
Observations	885	944	920	909	776
R-squared	0.012	0.017	0.026	0.012	0.045

Note: Standard errors are in parentheses.
*$p < .1$; **$p < .05$; ***$p < .01$.

Student response action

Our observations of how students responded in the lesson are described in detail in this subsection. The student response options were as follows (see also **Annex D)**: Read, Listen, Write, Oral Response, Physical, Manipulate, Getting, or Not Engaged. SERI, EQUIP-T and Lecture Pour Tous teachers spent the largest portion of instructional time with children Reading (**Table 75**). In Tusome, Listen accounted for the largest proportion of time (31.4%), while Oral Response was the most commonly observed response for NEI+ (43.6%).

Table 75.   **Percent of observed class time with various student response types, by program**

Variable name	SERI	EQUIP-T	Tusome	Lecture Pour Tous	NEI+
Read (R): % of [Student Response]	41.8%	41.1%	23.8%	58.8%	20.7%
Listen (L): % of [Student Response]	17.8%	5.3%	31.4%	21.4%	23.3%
Write (W): % of [Student Response]	12.6%	26.6%	9.2%	3.5%	5.3%
Oral Response (O): % of [Student Response]	17.1%	19.1%	25.7%	7.6%	43.6%
Physical (P): % of [Student Response]	7.8%	1.6%	1.8%	3.7%	5.7%
Manipulate (M): % of [Student Response]	1.3%	0.8%	0.2%	0	0.3%
Getting (G): % of [Student Response]	1.7%	4.8%	6.3%	1.5%	0.4%
Not Engaged (N): % of [Student Response]	0	0.5%	1.7%	3.4%	0.8%

Because these programs were seeking to provide more interactive lessons for students, we used Listen as the selected reference group for the models in **Table 76.** The hypothesis was that additional time spent having students only listening will be associated with lower ORF scores.

We found no significant differences based on increasing reading time relative to listening. However, in Tusome classrooms, responses in the Write and Physical categories were linked with higher student performance, such that if they spent more time doing written or physical response activities (in lieu of additional time listening), ORF was higher. This may be the result of students being more actively engaged in written and physical activities, which leads to additional learning. For Tusome, more time spent on Getting (which typically meant that most students were waiting) and having students Not Engaged (marginally significant) were linked with lower student performance. There were no significant factors for Lecture Pour Tous, NEI+, EQUIP-T, or SERI.

Table 76.   Ordinary least squares regression of student response on ORF, by program

Variable names	Results				
	SERI	**EQUIP-T**	**Tusome**	**Lecture Pour Tous**	**NEI+**
Reference: Listen (L): % of [Student Response]	—	—	—	—	—
Read (R): % of [Student Response]	**0.148**	**0.024**	**-0.029**	**-0.03**	**0.019**
	-0.183	-0.114	-0.094	-0.039	-0.099
Write (W): % of [Student Response]	**0.363**	**0.084**	**0.393 ****	**-0.072**	**0.017**
	-0.224	-0.129	-0.156	-0.055	-0.098
Oral Response (O): % of [Student Response]	**0.046**	**0.017**	**-0.037**	**-0.049**	**0.108**
	-0.17	-0.121	-0.146	-0.064	-0.116
Physical (P): % of [Student Response]	**0.213**	**0.184**	**0.956****	**-0.078**	**0.137**
	-0.259	-0.191	-0.427	-0.049	-0.195
Manipulate (M): % of [Student Response]	**-0.068**	**0.182**	**0.332**	**—**	**0.567**
	-0.225	-0.134	-0.327	—	-1.071
Getting (G): % of [Student Response]	**0.105**	**-0.031**	**-0.880*****	**-0.072**	**-0.769**
	-0.313	-0.213	-0.269	-0.181	-0.475
Not Engaged (N): % of [Student Response]	—	**-0.099**	**-0.485***	**-0.129***	**-0.067**
		-0.255	-0.267	-0.065	-0.415
Class Size	**-0.043**	**0.001**	**0.055**	**-0.004**	**-0.011**
	-0.209	-0.02	-0.089	-0.042	-0.05
_cons	18.397	5.907	20.055***	9.051***	-1.267
	-15.634	-11.896	-6.989	-3.116	-6.44
Observations	885	944	920	909	776

Variable names	Results				
	SERI	EQUIP-T	Tusome	Lecture Pour Tous	NEI+
R-squared	0.023	0.029	0.069	0.017	0.048

Note: Standard errors are in parentheses.
*p < .1; **p < .05; ***p <.01.

Level of student response

We examined where teachers directed their instructional attention. Focus is defined as where teachers were spending their time (e.g., is an individual student reading to the class? Are students working in small groups?). The response options were as follows: Whole Class, Large Group, Small Group, or Individual. The vast majority of time spent across classrooms in all programs was in Whole Class student response, ranging from 66.8% in Lecture Pour Tous to 83.6% in SERI (**Table 77**). Lecture Pour Tous had the highest proportion of time dedicated to Individual student responses (32.1%).

Table 77.   Percent of observed class time with various student response levels, by program

Variable name	SERI	EQUIP-T	Tusome	Lecture Pour Tous	NEI+
Whole Class (W): % of [Focus]	83.6%	76.9%	80.8%	66.8%	79.3%
Large Group (L): % of [Focus]	2.4%	2.6%	0.7%	0.9%	6.5%
Small Group (S): % of [Focus]	2.5%	2.8%	2.6%	0.2%	6.5%
Individual (I): % of [Focus]	11.5%	17.8%	15.9%	32.1%	7.7%

Because most observed classrooms spent most of the time focused on the Whole Class, we selected Whole Class as the reference group for the models in **Table 78.** The hypothesis was that Whole Class response may be better than Large Group, which is defined as approximately half of the class responding; but that effective uses of Small Group (i.e., pairs or small groups) might be beneficial to students.

Larger portions of classroom time in Large Groups (relative to Whole Class) were associated with lower reading scores in both EQUIP-T, Tusome, Lecture Pour Tous and NEI+, as hypothesized. Small Group responses were only significantly associated with higher ORF in EQUIP-T and Lecture Pour Tous. No factors were significant in the SERI model. These findings have implications for how instructional time is spent in classrooms in these programs, however the low proportions of large and small group focus across all programs may be driving these results.

Table 78. **Ordinary least squares regression of level of student response on ORF, by program**

Variable names	Results				
	SERI	**EQUIP-T**	**Tusome**	**Lecture Pour Tous**	**NEI+**
Reference: Whole Class (W): % of [Focus]	—	—	—	—	—
Large Group (L): % of [Focus]	**-0.156**	**-0.223****	**-0.796****	**-0.293*****	**-0.167****
	-0.205	-0.091	-0.354	-0.11	-0.083
Small Group (S): % of [Focus]	**-0.158**	**0.203*****	**-0.194**	**0.599****	**0.001**
	-0.303	-0.073	-0.318	-0.245	-0.116
Individual (I): % of [Focus]	**-0.012**	**-0.076***	**0.111**	**0.012**	**-0.099**
	-0.133	-0.039	-0.125	-0.024	-0.093
Class Size	**0.087**	**-0.008**	**-0.021**	**-0.012**	**-0.001**
	-0.207	-0.016	-0.097	-0.042	-0.047
_cons	29.73***	11.949***	19.659***	5.906***	6.005
	-4.836	-2.162	-3.449	-1.931	-4.28
Observations	885	944	920	909	776
R-squared	0.004	0.035	0.014	0.013	0.022

Note: Standard errors are in parentheses.
*p < .1; **p < .05; ***p < .01.

Post-observation checklist

We designed the post-observation checklist to measure whether classrooms included the best-practice activities that we expected might happen during an effective reading lesson. The checklist was quite simple: each item examined whether the activity happened at least once during the lesson. We used the list, which consisted of 27 items, to create classroom checklist scores across six dimensions (using exploratory and confirmatory factor analysis). The factors were as follows: Student Centered, Demonstration (teacher shows students what is expected), Simple View (of reading), Application (students practice skills), Responsiveness (teacher adapts to students' behaviors), and Efficiency (instructional time is maximized). The hypothesis was that higher scores on each of these dimensions could be linked with higher ORF scores.

Our hypothesis was met with mixed results in the models presented in **Table 79.** Higher scores for Student Centered activities were actually associated with lower results in Tusome but they were associated with higher scores in Lecture Pour Tous. The other statistically significant results were the negative relationship between Demonstration scores and student reading outcomes in EQUIP-T, SERI and Lecture Pour Tous, as well as the positive relationship between Efficiency and student reading outcomes in NEI+. This finding may be

the result of teachers spending too much time demonstrating and explaining lesson content, in lieu of allowing additional time for students to actually practice the skills.

Table 79.   Ordinary least squares regression of post-observation checklist scores on ORF, by program

Variable Names	Results				
	SERI	EQUIP-T	Tusome	Lecture Pour Tous	NEI+
Student Centered	**1.179**	**0.683**	**-3.307****	**2.426*****	**-0.627**
	-1.864	-0.786	-1.407	-0.91	-0.738
Demonstration	**-7.328****	**-1.795***	**2.46**	**-2.39*****	**-1.828**
	-3.206	-1.071	-2.382	-0.812	-1.165
Simple View	**-0.837**	**0.593**	**2.281**	**0.149**	**0.337**
	-1.775	-0.679	-1.707	-0.423	-0.73
Application	**1.957**	**1.257**	**0.673**	**0.483**	**1.668**
	-1.384	-0.82	-1.589	-0.545	-1.015
Responsiveness	**-0.048**	**0.462**	**0.104**	**0.263**	**-1.268**
	-1.713	-0.832	-1.736	-0.36	-1.099
Efficiency	**-0.768**	**0.816**	**-3.806**	**-0.058**	**3.019****
	-3.469	-1.803	-5.885	-0.996	-1.141
Class Size	**0.028**	**0.013**	**0.053**	**-0.011**	**0.017**
	-0.21	-0.019	-0.101	-0.034	-0.038
_cons	40.201***	1.708	10.707	3.037	3.871
	-10.345	-5.645	-15.754	-2.258	-5.331
Observations	885	944	920	909	792
R-squared	0.05	0.047	0.026	0.039	0.068

Note: Standard errors are in parentheses.
*$p < .1$; **$p < .05$; ***$p < .01$.

Key takeaways from the observation results

Our findings suggest that some classroom time-utilization elements were associated with improved literacy outcomes. In other words, how teachers use their time does make a difference, even within the context of effectively implemented literacy programs. One key finding is that not very many instructional time relationships were statistically significant in our regression models. Each of the models showed that the instructional practices from the classroom observation instrument accounted for a very small proportion of the variance in reading outcomes. This might be because the programs were implemented with relative fidelity so there was not substantial variation between program classrooms to tease out the relationships.

173

Some relationships between classroom observation results and learning outcomes were identified in at least two or three of the five programs. We found that additional instructional time spent on Grammar and/or Assessment was negatively associated with performance (relative to a focus on Reading), that using no instructional Materials was negatively associated with learning (relative to the use of Books), that spending additional time focused on the Large Group was negatively associated with learning (relative to the Whole Class), while additional time focused in Small Groups was positively associated with learning, and that more time with teachers doing Demonstrations was negatively associated with learning. These findings do not mean that these activities are ineffective on their own but that using these activities for longer than expected within these interventions was negatively associated with reading outcomes. It is certainly the case that other ways of implementing instruction make a difference, but recall that these data were derived from within classrooms in programs already shown to be highly effective at improving outcomes.

5.5 Qualitative Analysis (five programs)

5.5.1 Coding and analysis process

The themes discussed below are based on structured qualitative interviews conducted in seven programs: EQUIP-T in Tanzania, the PRP in Pakistan, SERI in India, Lecture Pour Tous in Senegal, NEI+ in Nigeria, Ghana Learning and Tusome in Kenya. Systems interviews with donors, program staff and ministry officials at the central and sub-national level were conducted for all seven programs. Due to constraints related to COVID-19 during the data collection period meant that qualitative instructional interviews with a small sample of teachers, head teachers, coaches, and meeting facilitators were only conducted for three programs (EQUIP-T, SERI and Tusome).

The interview data were transcribed into Microsoft Excel sheets, while the interviewers coded each question for a predefined hypothesis and causal arrows (explained below). Each interview was then transferred to the research team's qualitative software NVivo. During this step, each response in the interview data was coded for the question asked, hypothesis, causal relationship, and evidence quality.

The hypothesis and causal arrows were predefined based on the theory of change and were reviewed, as needed, during the analysis process. The hypotheses fell into five categories:

1. Effects of training on teachers,
2. Changes in instruction as a result of training,
3. Use of teacher and student materials,
4. Coaching processes and their impact, and
5. System-level processes and changes.

Causal relationships emerged between the key actors and beneficiaries of the programs, including students, teachers, school management, coaches, trainers, program, and the relevant ministry and district. Each of the responses was coded for the causal link between two or more of these actors. For instance, the hypothesis "training improved teachers' self-efficacy" was also coded for the causal relationship of "trainer to teacher" to indicate the process of change from these actors. Each response was also coded for the quality of evidence—i.e., high, medium, and low. We determined the quality of evidence based on the detail and relevance of the response. The highest quality evidence involved causal statements (e.g., "students learned quickly *because* of the new curriculum") and personal experience (e.g., during discussions of the effectiveness of teaching methods, more weight was given to the testimony of teachers versus district officials), and included concrete examples.

After coding all the interviews, we created summary tables for each program using the NVivo software to identify the causal relationship coded most frequently with high-quality evidence. Of the 39 causal relationships, 16 were identified as more frequently coded with high-quality evidence across the four programs. The responses with low-quality evidence were not considered in the analysis. For each of the 16 causal relationships, we generated summary tables for all the hypotheses that were also coded as high-quality evidence. Based on these summary tables, we found key themes in the responses across the four programs and identified quotes that would support the themes that emerged in each of the hypotheses. Although the 16 causal relationships were defined as those with high-quality evidence, we supplemented analyses of these relationships with evidence categorized as medium quality.

Once this process was complete, we wrote narratives based on interview data. We also used member checks, where interviewers reviewed the narratives, to identify any gaps or discrepancies and to address them with additional data where necessary. During this process, if data did not align with previous codes, they were recoded. This section of the report presents the hypotheses that were most heavily referenced in the interview data as high-quality evidence and were confirmed via summary reports by interviewers.

5.5.2 Narratives: Instruction-level data (three programs)

Findings from qualitative interviews on teaching and teacher support

1. **Training and follow-up: Included time to practice the methods (causal relationship trainer → teacher)**

Teachers in the two programs—SERI and Tusome—that included centralized, external-to-school training said that training was critical. The most important part of training was having trainers model instructional methods and having multiple opportunities to practice the methods.

For instance, in SERI, almost half of the training time was used to practice the techniques, as explained below by a coach who discussed some methods used in training for practice.

"When we did a demonstration, if there are 50 people, they worked on different parts. And everyone would watch and discuss. They would give the right/wrong/feedback. The teachers improved from these demonstrations and input. We did this every day [of the training]." (Coach, SERI)

These opportunities to practice appeared to be effective for coaches and teachers, compared to other trainings that primarily focused on listening. In another instance, a Tusome teacher said the following:

"You feel confident modeling and do it practically and you don't feel shy—you can teach like any other teacher. It made me feel confident. I used to teach upper [primary] before but this training encouraged me to come back to lower [primary] and teach." (Teacher, Tusome)

In particular, the Tusome program had a significant focus on practice, or microteaching, in trainings for teachers and for the CSOs, who had the dual roles of trainer and coach. Half of the teachers interviewed and the two CSOs interviewed mentioned practice as something that was either positive about the trainings or different from previous typical trainings.

"We usually do microteaching. They prepared us better. Microteaching, I really enjoyed." (Teacher, Tusome)

"Yes [the Tusome training was] different. [In previous trainings] the facilitator was just in front giving instructions, no microteaching. Tusome can ask questions [like,] what about when a learner does this…you do this way. I liked it." (Coach, Tusome)

2. **Training: Improved teachers' self-efficacy (causal relationship: trainer → teacher)**

According to trainers and other stakeholders working with teachers, the training also improved teachers' self-efficacy in all programs, which was critical to ensure high levels of implementation.

In SERI, the training helped teachers address their problems and needs and developed their self-efficacy.

"In phase 1, [teachers] were not confident. By phase 2, they were confident, and they said that it was working; kids were learning. So by phase 2, they were asking questions and getting better with the techniques. They were using them more." (Trainer, SERI)

"The confidence we (trainers) had, they [teachers] now have." (Trainer, SERI)

Two EQUIP-T teachers discussed how building confidence to implement new skills was important.

"The teacher was trying to build on the skills they had before the program. The program was not new, but it helps empower what they had before." (Director, Teacher Training, EQUIP-T)

"EQUIP has shone a light on how to have school-based meetings to learn skills. The meetings have encouraged teachers to work with one another. One will support another who doesn't know; it has brought confidence because they can solve difficult subjects. Students [have the] most difficulty in reading and writing things. The methods for how to teach to reading and writing have been helpful." (Meeting facilitator, EQUIP-T)

Practicing multiple times made teachers confident in Tusome. One teacher commented on the difference training had made:

"It is good. We have improved a lot. You can see [a] difference between Tusome and not-Tusome training." (Teacher, Tusome)

3. School-based training vs. centralized, external-to-school training (causal relationship: trainer → teacher)

EQUIP-T differed from SERI and Tusome in its reliance on school-based training to reach all teachers. The school-based training was led by teachers who had already attended a centralized, external-to-school training. Teachers spoke about the benefits of the school-based training, including the opportunity to collaborate with other teachers, gain course correction for skills they had not mastered, solve problems, learn new skills (e.g., inclusive teaching), and get motivation from the head teacher. However, the school-based meetings were not endorsed as strongly by participants in comparison to the centralized, external-to-school training. A widespread issue was that not every teacher got to participate in the centralized training. This was related to the per diems teachers expected to receive as "motivation" to attend.

4. Training. Collaboration: The interactions between trainer and teacher at the teacher training are positive and working toward a shared goal (causal relationship: trainer → teacher)

In the SERI program, the trainers said that unlike other training sessions, which were one-way, the SERI trainings focused on building a respectful relationship between teachers and trainers.

"It was very respectful relationships between a trainer and teacher. When you give respect to each other, relationships will grow. Teachers used to like this. Because they got and gave respect. We always shared and showed that it [was] a two-way relationship." (Trainer, SERI)

"We always tried. The teachers were very curious. We tried to be positive and not hurt the teacher. Room to Read told them to always use respectful language. Even if you are on the stage, you are not the god. You should listen. Be treated as you want to be treated." (Trainer, SERI)

The trainers asked teachers for their inputs and designed training based on teachers' concerns in classrooms.

"At the end of the day, we got a list of what teachers wanted to learn. If there is one component they wanted to [know] more, we will use that to plan for the next day. [The trainers] had the decision making. The State Office [Room to Read] used to give us the guidelines. Moreover, they gave us the components and the timeline." (Trainer, SERI)

"If we ask [teachers] to share their opinion and experiences, they did not get bored. They were learning, and they were motivated. Teachers used to tell the trainers this is a good program; it is based on learning and writing." (Trainer, SERI)

A teacher in Tusome emphasized how trainers interacted with teachers.

"[Trainer] adds value because they have new methods. They interact with others and try to solve some problems, exchange ideas." (Teacher, Tusome)

The above approaches motivated the teachers and kept them engaged in the training. Similarly, teachers in EQUIP-T said that the centralized trainers were very collaborative and discussed ideas with each other during group activities.

"The trainers were very collaborative. The way they engaged [teachers] throughout. They put them in groups in the same ward [subdistrict]—they give them some questions to answer. Then they give them directions on how to change ideas so they can help one another in helping students." (Teacher, EQUIP-T).

In addition, teachers in EQUIP-T said that school-based meetings were a good opportunity for collaborating and solving problems.

5. **Instruction. Students were more active in their learning (causal relationship: teacher → student)**

Teachers and head teachers also recognized the effectiveness of active student participation in their learning. Many interview participants reported that the previous techniques were teacher-centered, while the "I do, we do, you do" model helped shift the focus to involve students in their own learning.

"Because of that [technique], they are now really participating. Before the teacher used to write the word, and say this is what it is, they were not getting how to read. Now, children are eager to learn." (Head teacher, SERI)

"The program is designed according to the kids. It uses I do, we do, you do. [Students and teachers] catch this very quickly." (Coach, SERI)

"Following the I do, we do, you do, then learners acquire the skills because of the method. Teachers like the work. Previously teachers used to lecture. Learner is confident with the I do, we do, you do. Teacher shows them what to do." (Trainer, Tusome)

"In the past, the teacher only had 'I do' and there was no 'we do' or 'you do.' Now it is helping students to learn better. It was not very child-centered but now the child is fully involved. It is better." (Teacher, Tusome)

Programs were successful in increasing student participation in learning activities, particularly related to practicing new skills. In the EQUIP-T program schools, there was more student interaction because teachers used the "question-and-answer method" and had students work through problems on the board.

"It's not just me talking. Students are talking and doing things, too." (Teacher, EQUIP-T)

Teachers in EQUIP-T also used group work, with stronger students supporting weaker students. The materials, such as letter cards, were designed to allow students to practice the skills they had learned.

"Many teachers have skills on how to prepare students. Training. It helped them learn different methods to teach students. Group work, teaching through singing. Drawing. Picture code; students have been engaged in a practical work rather than just writing. For example, to compare between things and numbers." (Head Teacher, EQUIP-T)

"When you go to teachers college, you can find the interactive methods from the competency-based literacy methods. You can find some tutors who are working in this area very successfully, compared to other areas. This is the major contribution that EQUIP-T did." (Director, Teacher Training, EQUIP-T)

Teachers and trainers in SERI and Tusome also gave examples of greater involvement of students in their own learning.

"Before, it was mostly teacher-centered. But now we find that learners are involved and there was a lot of self-efficacy as when they are able to go to the blackboard and you feel 'wow'." (Teacher, Tusome)

"Students say their own thoughts. They look like they are important in the class. It shows how important you (a student) are to the class. When you are sharing your thoughts to everyone, that develops a feeling that you are important to the class. When the student sees the teacher giving the importance to everyone equally, the participation increases." (Trainer, SERI)

Teachers in the EQUIP-T program also related ideas and words to real-world objects. For example, they used materials given by the program to create pictures to accompany words taught in the classes. One teacher said that the words were also linked to actions, such as making a cow out of mud while discussing the word "cow."

6. Instruction shows students the relationship between sounds and letters (causal relationship: teacher → students)

Respondents from all three programs recognized the effectiveness of the teacher training provided through each program. One of the most valuable skills that teachers across SERI, EQUIP-T, and Tusome developed was using sounds to build students' reading skills. Letter–sound relationships had not been the focus of instruction previously. This approach helped students slowly progress from reading letters to words, sentences, and—eventually—paragraphs. The quotes below explain the effectiveness of this approach in different contexts.

"*[Students] can write the word. For example, when they want to write a word that [begins with the] letter with /m/, they can get it started. Even if they can't get all of the letters that are in the word, they can try to get some of them of the word written.*" (Teacher, SERI)

"*We only teach one letter in one day. What happens in that, the kids never forget the letter. They can remember the letter for a long time. This program includes the sound, the writing, and the reading of the word. It really works in favor of the kids.*" (Head teacher, SERI)

"*They are good at pronouncing the letters and words now. B-A-B-A.*" (Head teacher, EQUIP-T)

"*Before EQUIP, they were mixing up languages in reading and writing—English and Swahili. Before training [they said] 'A, B, C' but EQUIP has given a better method of pronouncing letters.*" (Meeting facilitator, EQUIP-T)

"*Positive impact. Teaching the children on reading, I know that the child is supposed to know the sounds of the letters first so the child can blend the sounds of the word. In 'look and say,' they used to just say 'cup,' but now the child knows the sounds.*" (Teacher, Tusome)

7. Instruction. Teachers monitor progress and adjust their instruction accordingly (causal relationship: teachers → students)

Because of greater student activity (see qualitative findings above), teachers in all three programs were more able to monitor their progress and adjust their instruction accordingly.

Teachers in Tusome and SERI said they could identify students who had disabilities or needed more time and attention.

"I am getting…more keen on the children [and] their response. If children cannot pronounce, I get to know and I call the children later. And [it] makes me understand children better." (Teacher, Tusome)

"Because of this program, [teachers] are more connected to the kids. They are checking the kids' work in writing. Also, they are interacting with more kids orally, so they know where they are. They know who is getting it. Before Room to Read there was much work to do, there wasn't much activity to do. Before that teacher[s] were a little bit lazy. It is not easy. The teachers like it. When we started some were concerned. Now it is now getting easier. And they do it because they see it works." (Head teacher, SERI)

"I can see the progress in the kids. Their reading quality has improved. They get connected to each other. I ask the kids what they have learned. If they read to me correctly (word accuracy). First, I read, then I ask the children how much they understand. If they didn't answer, I tell them to answer the question to each other. They do that usually. These interactions connect them to each other." (Teacher, SERI)

In EQUIP-T, a head teacher explained techniques that teachers use to demonstrate to students what they needed to do, which then allows teachers to gauge the students' reading skills.

"The method of engaging teachers and students. Dividing students into groups. Give them different questions. Then later on they will present the outcome of the exercise. For example, reading. You write something for them to read. If one group has managed to read then you give them something else to read. The aim is to make sure everyone participates. If there is a group which didn't manage to read, another group will help. One child will help another." (Head teacher, EQUIP-T)

8. **Instruction. Teachers are motivated because they can see how instruction leads to positive student learning outcomes (causal relationship: teachers → students)**

Teachers in SERI and Tusome described the progress of students enthusiastically and explicitly. And by implication, we received the impression that they were more motivated, although few teachers said so explicitly.

"There is nothing as such that is challenging. I like the program because I can see progress in the students. Earlier there was progress but now there is more. Most of the kids are improving." (Teacher, SERI)

"At first, people scolded me on why I was behind [in regard to the pacing of the letters]. It changed when the assistant block officer came last February. He saw the progress and praised me. Now every kid in my class can really learn. So it is better to teach." (Teacher, SERI)

The teachers were motivated to see significant improvement in students' reading skills due to these teaching techniques. In SERI and Tusome, teachers noticed that students could learn to read and decode earlier than they used to. The quotes below from interview participants provide strong accounts of how the programs had influenced teachers' overall teaching skills and the impact their teaching had on students.

"Because of this program, grade 1 students are now reading small words. Grade 2 children can now read on their own." (Head teacher, SERI)

"I see a big difference. Earlier they could read on their own in grade 3, now they can read with understanding in grade 2 and some words in grade 1. So, it is a big difference." (Head teacher, SERI)

"Because of Room to Read, what used to happen later in [grade] 4 and [grade] 5, was when [students] used to start to learn to read fluently. Because of Room to Read and the library, from [grade] 1 they can read the story, they can decode (the small story), they can read it fluently. [Grade] 1 can write now because of the instruction and the daily library." (Trainer, SERI)

"Since I started teaching, it's the best thing to happen to teaching in 28 years. The government is really in touch with what we wanted. After one year, all the pupils know how to read." (Head teacher, Tusome)

"I am able to know the children who have special needs and how to deal with the time takers—those who cannot read." (Teacher, Tusome)

As reported above, one teacher in Tusome said, *"when they are able to go to the blackboard and you feel, 'wow'."*

9. **Materials. Books and learning material make it easier for teachers to engage students (causal relationship: teachers → students)**

Another significant aspect of SERI, Tusome, and EQUIP-T's programs was the provision of engaging reading material and teaching aids. Teachers said that they enjoyed working with the material and that students were excited to see colorful books and stories.

"They also understand what kids need to learn, the 'matra' [vowel signs]. What happens in other books, letters and matras are together. In these books they are separate. They cover everything of that. This is the best. They have a lot of words for examples. In other books there are not many examples. In Room to Read the teacher does not need to find the examples." (Head teacher, SERI)

Besides students, parents were also excited that students could access storybooks that they could bring home. According to interviewees, these materials also helped address

absenteeism in schools. In Tusome, a teacher reported that the easy language and engaging content for students made them very excited.

"In fact, you have really helped. Even the teacher enjoys teaching when you read a text freely and answer questions and construct good sentences." (Tusome teacher)

Another Tusome teacher also supported this view:

"I like the books. When I'm away, they read the books and compete on reading. [They] enjoy the stories very much, the pictures are real, it brings their interest." (Tusome teacher)

In the EQUIP-T program in Tanzania, teacher-made materials were mentioned by nearly all teachers as being a critical part of the program. The key materials made were cards with letters, pictures to illustrate a word, and flipcharts. These materials allowed students to practice skills and connect words to real-world objects.

"EQUIP-T tried to identify different stories presented to teachers. Facilitators wanted teachers to create stories related to the context, since contexts are different. The program helped teachers utilize their own context to help make learning so meaningful. I went to Shinyanga Region, which is different from Dodoma and Lindi in the way they utilized the program. They created mathematics learning tools from the environment and invite experts who can deliver [the lessons]." (Director of teacher training, EQUIP-T)

10. Materials. Teachers' guides motivate teachers with lesson and content (causal relationship: program → teacher)

Teachers' guides were well received in all programs. The teachers' guides helped teachers plan their lessons and gave them easy-to-follow lesson plans and directions on how to teach. In short, they made teachers' work more effective and easier.

According to one teacher in EQUIP-T, *"Textbooks give direction on how and when pupils can learn—they show the stages of teaching."* This teacher said that manuals *"give skills and direction on how to teach."* She explained that the topic was written on the cover page, the specific skill was in boldface, and the lessons included the three steps to teaching that skill.

Comparing the teachers' guides in the Tusome program to others, a head teacher said,

"It used to make mine and teachers' work easier. Most of the teachers were comfortable and agreed that it's much easier. For the others, you have to create a scheme of work and lesson plan. The other system is time-consuming. You spend a lot of work writing." (Head teacher, Tusome)

According to coaches in SERI, the guide's clear directions helped teachers avoid confusion and teach appropriate content:

"It tells them the order of what order to teach the letters. It helps to avoid confusion. Before, some teachers would teach letters in the wrong order without the teacher guide. It would confuse the children because they [teachers] might pick difficult letters at first." (Coach, SERI)

Users came to see the benefits and praised the guides.

"At first, the teacher guide was confusing. Other programs like MGML [multigrade, multilevel], were not successful. They did not have teacher guide. At first Room to Read teacher guides were confusing but then they started using them. It is a positive effect. With the teacher guide, all the problems went away." (Coach, SERI)

11. Coaches/meetings improved/reinforced teachers' knowledge and skills, plus coaches provide support, motivation, problem-solving (causal relationship: coach → teacher)

For SERI and Tusome, school visits by an external coach were an important part of the program model. The role of the coaches was to give support, reminders, and feedback to teachers.

In SERI and Tusome, the coaches visited the schools after the training to provide support, give feedback, and help teachers to improve their teaching skills. For instance, in Tusome, the coaches observed class, noted changes to the lesson or areas for improvement, and gave feedback based on teachers' strengths and areas of improvement. The coach also demonstrated and practiced with the teachers and trained teachers on how to evaluate themselves using the Tusome model. A teacher said that the coach helped her understand how to teach a young student through the CBC. Teachers also said that they felt they could easily call the CSO and ask for help if they were stuck with a lesson and that the CSOs encouraged the teachers and made them feel confident. The following quotes from SERI and Tusome teachers illustrate the importance of coaching.

"It was very important. Whenever [coaches] used to visit, I improved. They shared the important parts of what to teach. The training was there and was good. We got all the information in the training. Sometimes we would forget [what was learned in the training]. They would remind us. We also get improvement. If you want to teach the kid, you have to be the kid. The regular visit was a big part of why the program has been successful. It was very important to have a connection between the coach and the teacher. It benefits the coach, it benefits the teacher, it benefits the kids, and it benefits the program." (Teacher, SERI)

"There is a very big lesson that tells us why Room to Read is successful. In other trainings, [teachers] are all alone. In Room to Read they all have a coach, they come weekly, stay for the full day. The teacher gets confident and… they get double support. This is the biggest

reason for success for three years. You lose your memory from a training, but with a coach you don't forget. You always get new things [with the regular visits]." (Trainer, SERI)

"[The] CSO, she comes twice a term. She comes to my class to support me. She comes and sees my lesson and any problem I have she corrects. She supports us and doesn't harass us." (Teacher, Tusome)

In SERI, teachers described the coaches as problem solvers who helped the teachers prioritize and do so in a welcoming manner.

"I want to say more about the connection between teaching and coaching. I have an example. I get [too friendly] with my kids. Every part of the lesson is important. Whenever the coach comes, I'm reminded of what I'm skipping. Also, from the coach I get answers. The coach is something of a guide. I sometimes might make mistakes. For me to improve, he won't scold. The interactions are friendly. Whenever the coach visits there is respect between the teacher and the coach. If he is going to scold, there wouldn't be improvement. There should be a respectful relationship. They cannot be bossy. The coach was not bossy." (Teacher, SERI)

In EQUIP-T, school-based training was conducted by teachers through a cascade model. Teachers reported the details of teacher meetings and peer support.

"We discuss the improvement of quality education and school management. We solve problems—for example, being late. Before EQUIP-T, teachers' meetings were very rare but now we do them weekly." (Teacher, EQUIP-T)

"In class, I learn from my partner teacher who did the EQUIP-T training. One thing I've learned is how to use those teaching materials like cards and charts. When we start the period of reading, we teach the syllable 'ma' and start showing the cards. After we show them we tell the pupils to make a word with 'ma'." (Teacher, EQUIP-T)

However, according to the Director of School Support, there were some challenges in implementing this model for post-training support.

"Training[s] at school level—these are not happening regularly. This is due to a lack of teaching staff and being busy with school. There are other challenges with WECs [Ward Education Coordinators]. Some have been moved. When they receive new ones, it takes time to build their capacity until they master how to support EQUIP-T." (Director of School Support, EQUIP-T)

"Teachers did not implement the training after they got it. At the ward level, WEOs [Ward Education Officers] did not provide a standardized training. WEOs did not follow up much." (Director of Teacher Professional Development, EQUIP-T)

185

When the Directors were asked for specifics, they reported that no teacher facilitation meetings had happened in the EQUIP-T schools since around October 2019 (i.e. approximately four months before data collection). However, council elections were taking place during this period, which impacted the ability of such meetings to occur.

12. Coaches were provided with training plus program- or government-provided support on how to coach better (causal relationship: trainers → coaches)

There was evidence of effective coaching training from the Tusome program in Kenya and from SERI in India. This training played an important role in coaches accepting the program and being motivated to carry out activities. Ongoing support was helpful and particularly welcomed because it had not been commonly provided previously.

"Other programs do [training] and don't follow up when they are done, not like RTI who still comes to support, continues to hold our hands, it seems to be a part of them. RTI will visit all or most of the stations." (Trainer, Tusome)

"RTI—I like especially [name of coach]. He was [a] CSO like me. He comes to support me. I give feedback and then I ask them to add. Can you tell me if I did well? So I can do better next time." (Trainer, Tusome)

"When the program came, I thought, another new program. This is another program. Why another one? We were very irritated. Then we saw the big books and saw how much we learned. When we took the training, we saw the program that we should follow it. Then when we started using [the program] we saw that the kids were learning and we should use it more and more." (Coach, SERI)

Conclusions

Across the three programs, a number of features were identified by respondents as being critical for effectiveness.

1. Training emphasized teacher practice. The initial training, the follow-ups, and the ongoing peer support were structured so that teachers had time to practice the methods.
2. Training improved teachers' confidence to implement new skills, and skills they had learned previous.
3. There was a positive, collaborative, and respectful relationship between trainers and teachers.
4. In Tusome and SERI, the teachers' guides included clear, step-by-step instructions and provided content in a child-friendly order with understandable lesson plans.
5. Coaches supported and motivated teachers and helped them solve problems.
6. Instruction focused on teaching students letter–sound relationships in a way that would support their ability to read words and increasingly complex text.
7. Students were encouraged to be active in their learning and to demonstrate proficiency in new skills.

8. Students' demonstrating proficiency in new skills allowed teachers to monitor their progress and adjust instruction.
9. Instruction was supported by engaging storybooks and materials that allowed students to practice new skills.

5.5.3 Narratives: System-level data

Report on qualitative interviews on government systems

Qualitative interviews were conducted with program staff and government officials at various levels of the system from the central ministry to regions, to subdistricts for all programs. The main aim of the interviews with government officials was to understand different system actors perspectives on key aspects of program implementation and the program's relationship to the different offices and actors in the education system, as well as their insights into the factors associated with program success. The findings from these interviews are summarized and highlighted in this section of the report.

1.   How the government was persuaded to take on the program

For SERI, Tusome, and PRP, the main factors that influenced uptake were evidence of the need for a reading program and of the effectiveness of the program, genuine collaboration between program and government, and alignment between the program and the government's curriculum goals. For EQUIP-T and PRP, a significant factor was that the program met government existing needs.

EQUIP-T
EQUIP-T faced two turning points during its implementation (by Cambridge Education) that increased government buy-in and take-up of the intervention. The first was the transition to the reading, writing, and arithmetic (3Rs) curriculum in 2014 and 2015. This transition was brokered by the Big Results Now government program, which began in 2013. However, the main challenge for the 3Rs curriculum was the lack of teacher training or materials for teachers to implement these new approaches. Therefore, EQUIP-T's ability to provide training and materials to teachers tasked with implementing the new 3Rs curriculum not only made EQUIP-T popular with teachers, but also ensured that it was seen as a program that responded to demand. The second turning point was a decision made by the program's funder, DFID, to provide budget support directly to the Tanzanian Government. In the early stages of EQUIP-T, the program had difficulty managing relationships with the Ministry of Education, Science and Technology and the President's Office for Regional Administration and Local Government. However, changing DFID's approach to supply substantial EQUIP-T funding directly to the government substantially improved EQUIP-T's access and the level of buy-in from national and subnational government levels. This change made it considerably easier to implement EQUIP-T's instructional improvement activities.

SERI

For SERI, representatives in three of the four Indian states with whom we spoke mentioned that they were convinced to adopt this USAID-funded program because of their experience with the existing Room to Read literacy program implemented in a number of states in India, combined with data showing the program was effective in improving reading skills. After the state government accepted the program, the next step involved convincing individual districts to be involved. The first district in Chhattisgarh State (Baloda Bazar) was chosen because it was near the capital and had low literacy levels. The district had also had a positive experience with Room to Read's library program.

The program point person in Baloda Bazar District Education Office explained, "*Early on, trust and faith were missing. Many NGOs would come and go as soon as their work was done. This organization cared about doing the work properly.*"

Baloda Bazar's Chief Administrator, the Mission Director, was a friend of Chhattisgarh's Head of Education, which helped the state convince the district about SERI's importance. Baloda Bazar's district administrator knew that literacy was a problem and wanted to focus on lower grades.

Tusome
In Kenya, Tusome was influenced by the success of the PRIMR Initiative, which showed significant improvements in math and reading outcomes for students in pilot schools. Because PRIMR included a strict research design (i.e., an RCT) and showed significant improvements in student's learning, the results were widely disseminated to government officials. According to a USAID official, the agency spent a lot of time with government officials explaining PRIMR's results and taking officials (including the Minister of Education and Principal Secretary) to visit schools to see the program in action. The MOE was convinced by the program's success and wanted to scale up PRIMR nationally, which eventually led to the design of Tusome. Many respondents also mentioned that Tusome's launch by the President of Kenya was an indication to government officials nationwide that the program was to be taken seriously.

PRP
In Pakistan, the initial lever for ensuring government buy-in for the PRP was providing concrete evidence of the need for the program. This point was confirmed by a high-level Ministry of Federal Education and Professional Training (hereafter, referred to as MOE) counterpart in Pakistan, who admitted that that he was tough to convince initially but got on board once there was evidence of effectiveness. From that point on, he really appreciated what PRP did and even went so far as to adopt similar things in other areas of his province.

"*We got many questions from government about 'why reading?'. There are so many other subjects, why would we just focus on the one? We had [EGRA] baseline and other evidence to show the poor performance.*" (USAID)

Furthermore, PRP leaders noted that buy-in comes from genuinely working with the government, not just claiming that you will plan things jointly. This was also confirmed by high-level MOE officials. For example, one official noted,

"Those who came up with this project were very smart. The planning was excellent. They carried out a very clever needs assessment of the society and desires of the people and government."

A second high-level MOE counterpart further noted,

"the relationship between PRP and the government was based on needs from the government. [The government] had certain work they wanted to accomplish and reached out to PRP for technical support. This was important because it wasn't the program forcing things."

Ghana Learning

Learning faced some initial implementation challenges due to the original design of the activity. Two years into its implementation, the program was redesigned to better align with MOE and GES plans and expectations. These shifts were important for ensuring government buy-in and support.

Additionally, Learning engaged MOE and GES officials and conducted a pilot project prior to the start of implementation. This work was essential for showing how the program was intending to address an essential issue (i.e., low reading scores nationally). One high-level GES official shared:

"They showed us videos. Children were reading. They were pointing to letters and words. We had the meeting at the GES conference room. We all marveled. It was a success story for us. We decided that we should scale it up to reach as many children as possible."

Another official noted:

"We [at GES] felt that when Learning comes in, it would go a long way to fix a major problem which we were facing (i.e. just 2% of children were able to read and write). It was a felt need."

2.   Expectations for system counterparts are specified, plus the system communicates expectations for districts, schools, teachers, or students

Across all programs, the authority of the implementing country's central government is critical in communicating expectations. Government endorsements for some of the programs' expectations were done by issuing circulars and seeking formal approval of program approaches (e.g., materials). In addition, good, continuous communication between the programs and both central and regional governments was critical. The

education systems in India and Nigeria—and to some extent in Tanzania—are decentralized and communication was required at multiple levels. At times, communication with the government was especially challenging, but particularly critical, in the initial stages of the programs.

SERI

The SERI program in Chhattisgarh, India, began with a memorandum of understanding signed by the state government and Room to Read. As a USAID/India Senior Education Specialist said, *"Convincing the government is crucial—in Indian schools, a lot of things are top down. Once you are successful, they are with you."*

Although Chhattisgarh State already had an hour devoted to literacy each day, which provided a good platform for the program, the state government played a critical role in setting up additional, initial conditions to implement SERI.

Some of the communication about SERI was delivered to teachers directly from the state government. The Room to Read State Manager estimated that during the course of the program, the government issued between 10 and 15 circulars that gave teachers instructions. Some of these circulars—e.g., those asking teachers to attend training—were issued at the request of Room to Read.

The other key line of communication was between the State Pedagogy Coordinator and the SERI program focal person in Baloda Bazar District. The district also set up a District Resource Group to coordinate with the program and to review materials. In fact, one District Project Coordinator (a government position) in Baloda Bazar commented that their contribution to the success of SERI was his work coordinating between the district and state levels.

Read India

In Karnataka, India, counterparts were informed about the program. Evidence suggests that both program and State officials discussed the project through official agreements, meetings and, presentations. These discussions involved exchange of feedback and planning of new changes in the program. A Project Official said:

"There was a lot of conversation about which classes to choose. They decided on classes 4-5. Because ASER showed lots of gaps in 4-5. There was also discussion on the number of days. 60 or 90 days? We decided on 60 days. We had an MoU with the government"

For government officials, system communicated expectations and information about the program through official circulars that originated at the State level and travelled to the lower bureaucratic order. In addition, State officials admitted to providing information through meetings and orientation, however a district official mentioned receiving information through a single official order. A District Official explains the procedure:

"There's an order from the state office, therefore you implement"

PRP
Similarly, in Pakistan, the government set the expectations for counterparts working on PRP. These expectations were guided by a technical advisory committee and steering committee (composed of executives, including planning, finance, and human resources) and led by Pakistan's MOE with PRP representation. A key high-level MOE official from Pakistan explained the process of priority setting and communication:

"The political leadership sets the priorities. In the education sector, we have a sort of forum to establish and determine priorities. They are all reflected in five-year plans [Education Sector Plans]. These plans must be endorsed and approved by the provincial cabinet. Therefore, this support was included in the plan at the high level, and consultative sessions were held with [the] high level, districts, and teachers. We have our own dissemination strategy regarding policies and priorities from the top down. This program [PRP] is part of a five-year plan, so it was disseminated to all districts/schools."

Another high-level ministry official in Pakistan further noted that the relationship between PRP and the government was very strong and that the program and government worked side by side. He shared that people used to joke during their government monthly education meetings by asking: "Is this a PRP meeting or an education meeting?" Although the meetings were typically government only, PRP always attended and they planned together.

Additionally, USAID played a large role in the relationship-building and expectation setting for ministry counterparts. PRP's Agreement Officer's Representative noted that

"USAID signed [an] MOU [memorandum of understanding] with all seven ministries for the program—thus, the work could take place even though [implementer International Rescue Committee] was not registered. Additionally, whenever [PRP] has issues, we take up the issues with the host government separately. For example, PRP had concerns about the teacher transfer and we raised these concerns directly with the government without PRP present."

EQUIP-T
A collaborative approach was also important in Tanzania, where a senior official explained that the curriculum body that she leads approved the materials officially. This layer of approval reinforced EQUIP-T's importance among teachers. The senior official conducted training for several regions in Tanzania and commented that because key leaders in authoritative positions within the system encouraged teachers to use EQUIP-T materials, the teachers were more receptive. The program succeeded as a result of this national-level engagement and the subsequent communication of the program's importance to teachers and the rest of the education system.

Tusome

Similarly, in Kenya, communication was important from the outset.

"We did a lot of sensitization together with RTI. We used media, emails, and worked alongside RTI. We were communicating that there is this program coming up and this is the role. Saying that this is a government program but we are working with RTI. People are sensitive sometimes and don't want to see this as coming from outside, from 'muzungus' [foreigners]. So, we also align it with the Constitution, with [the] Basic Education Act, and with the county's own plans. We had to do a lot of alignment. Sensitization was also based on [a] pilot, based on evidence." (Kenya National Coordinator of Tusome)

For Tusome, the communication flowed in a hierarchical way from the central ministry to the county and subcounty, then down to schools (i.e., head teachers and teachers). The program's communication also followed the hierarchical flow of information supported by other communication strategies—for example using WhatsApp—that allowed for direct communication with ministry staff at the county or subcounty levels. Formal communications usually involved circulars or memos, while less-formal communications were disseminated through e-mail, phone, and WhatsApp. For example, in January 2020, the MOE sent a circular to head teachers of all primary schools (via the regional directors of education, county directors of education, and subcounty directors of education) that provided guidelines on how the schools and teachers were to use the Tusome books within the context of the new curriculum. Based on interviews with officers at the subcounty level, they did receive the circular. Conversely, WhatsApp communications were used for communications reminding CSOs of trainings, how to use books, etc. Tusome also attended joint meetings with county officials for the development of education-related work plans. Because of these planning activities, Tusome activities are embedded into the official work plans of county officials.

"The TSC County Director does not communicate directly to teachers. They do this through the existing structures for the subcounty directors and the CSOs, who reach the teachers. The communication was mainly on the program activities of the day and one of the expectations is a 100% attendance rate, evidenced by the list of attendance… Once the memo is received, it flows from TSC County Directors and then it is customized and sent to CSOs, then to teachers. Sending this through the CSO ensures any information gets to the teachers easily." (County education official, Kenya)

Another avenue for conveying and discussing expectations was through the various steering committees that were set up to guide Tusome.

Finally, constant communication was a theme that ran through many of the responses and was seen as a success factor by interviewees. When asked about what aspects of the program contributed to success, one county director in Kenya made it clear, *"Communi-*

cation. This is because with[out] communication from the right people, work cannot run properly." (County education official, Kenya)

Other interviews referenced "constant dialogue with the program," "practical ways of communication," and "constant interaction with Tusome."

Ghana Learning

In Ghana, the Learning team relied on traditional Ministry systems for communicating expectations to decentralized staff. The Director General (DG) would send letters and directives that would trickle down to regions, districts, and schools. GES also held orientation sessions and workshops in collaboration with the Learning program for district and regional leadership. Once these individuals understood the information, it was easier for them to disseminate this information to decentralized staff. Within regions and at the district level, there were also coordinators who worked directly with the program. For example, a District Director of education shared the following on how she received communication from the DG:

"Communication came from DG of GES. It told us about new activities of Learning and what was involved (including roles/responsibilities for district directors). I was happy because this was coming at a time when schools were having difficulty in the number of children who were struggling to read. This was essential for building a strong foundation. This was a lofty and fantastic idea. When communication came from GES, I was expecting that it would bring out change in education in a positive way. This was followed up by a workshop organized by USAID, giving us the details and modalities on how it was expected to go. Helping us to obtain buy-in."

Learning also worked directly with regional- and district-level coordinators, as well as staff from the National Teaching Council, National Inspectorate Board, and National Council for Curriculum and Assessment. Ultimately, as one high-level MOE official put it:

"I cannot see that there is anyone in the ministry who did not know about the project."

NEI+

In Nigeria, the State was seen as creating an enabling environment that led to different agencies coming together to support the program. Officials of the LGEAs were very active, and the government encouraged the SSOs to take part in training and oversee the implementation of the program in schools. The SSO's role also shifted from traditional M&E to supporting teachers in following the training design. In an example of the State Ministry enabling the quality assurance team, a high-level State Ministry official shared,

"Children have joy in doing; and the QA officers were given necessary support and they were doing their work. Have moved from QA so [that] teachers will not be afraid when they see the SSOs, [but rather] see them as partners. This has really led to success. When you

look at the baseline [about] how the performance was low and, in the midline, you see the improvement, [it is] better than Sokoto—we are always ahead. Even [in] teacher attendance and student attendance. These are some of the good things that have happened."

However, not much was reported by stakeholders as to how the system communicated expectations from the state to the LGEAs, likely due to its decentralized approach to education. The acting education secretary for Ninji did share that the State Ministry told them that NEI+ is working in the communities, and they are expected to provide their maximum cooperation. Some of these expectations included providing the program assistance, calling head teachers' attention to program activities, and taking staff to remote schools.

Lecture Pour Tous

The DEE was the official leader of the Ministry's Lecture Pour Tous program, with technical support from the Chemonics team. The inspector of DEE noted that the program set the expectation of partnership, collaboration, and "accompaignement". This expectation grew out of the way the technical assistance team for Lecture Pour Tous initially established their relationship with, and subsequently worked with, the DEE.

To clarify and emphasize these expectations, the DEE created an internal committee for all matters concerning Lecture Pour Tous implementation. The DEE oversaw coordination with the other arms of the Ministry concerned by the work of the program, notably the training directorate (DFC), the institute for education research and assessment (INEADE), the planning directorate (DPRE), and the human resources directorate (DRH). However, a management committee across these entities was not set up. Regional consultations were held to inform IA and IEF staff. Staff at the regional level shared that they received communication from DEE through memos and administrative letters. The district education officer explained that after they receive information from DEE, they first communicate to the IEF who shares it with the school director, and then the school director communicates to teachers. At the district level, communications are also shared via WhatsApp groups, which the district staff feels is an easy and faster way to share information.

"…Main communication vector, however, was through trainings (cascade model)—this is how IA and IEF (and even school directors) came to understand the new expectations for their roles. [The] Main difference was a new expectation related to serving as coaches for teachers. [This] Wasn't an official change in the job description, just a change in how they approached their job and job responsibilities. [The] Idea of "encadrement" through observation and feedback was introduced in this way, as opposed to control of teachers/teaching." (Inspector in the DEE)

Although the DEE was the point of contact for regions, the DFC was responsible for communicating expectations to pre-service teacher training centers and did so by working through regional education offices (IA) and directly with CRFPEs. A note from the minister, with a related note from the head of the DFC, were officially transmitted regarding the introduction of the national language modules.

The Director of INEADE, who is relatively new in the position, felt that Lecture Pour Tous was communicating in parallel to INEADE, i.e., that Lecture Pour Tous would inform them and include them, but not fully develop the ideas and approaches with INEADE staff. He said he would have preferred a more consensus-driven approach in which the program and INEADE determined together the focus and implementation of activities. The Director suggested that in his opinion, communication and engagement of all actors at all levels would have been better if program documents were shared.

3. System actors persuade or convince teachers that the program is valuable for their job

For the majority of programs, communication of expectations by the central government was considered by officials to be sufficient for implementation to take place, and there was little discussion of persuading and convincing teachers that the program was valuable. For SERI, advocacy with teachers at the subdistrict level was also deemed necessary to ensure their participation.

SERI

For SERI, district officials faced challenges resulting from teacher skepticism of external initiatives. Teachers were also not used to scripted lesson plans, which were proposed by the program.

"Across the board, I find people saying that there were NGOs saying do this or do that, this package or that package, and we were apprehensive when Room to Read came along." (Senior Education Specialist, USAID/India)

To ensure teacher buy-in, SERI's district focal person went around Baloda Bazar, block by block, convincing teachers to participate. He worked hardest on one block where teachers are more politically active. Because the focal person was a well-known and respected person in the area, he used his influence to motivate teachers to *"take up the profession with sincerity to gain respect in return."* He added, *"I told the master trainers, 'it's more like social work than a profit-making business. You have to love your work.'"*

The district focal person used evidence to convince teachers if they were unwilling to participate. Because he had first-hand experience with the *"systematic approach and methodology"* of Room to Read, he used this experience to convince teachers of its usefulness. He said he took a friendly approach and suggested that they could drop the

program if they did not see any improvements in six months. Using these strategies, he was able to convince all teachers to participate.

It is worth noting that convincing the teachers to participate did not stop once the training started.

"At the training session, the priority is to clear the doubts of the teacher. The team also works in a friendly way. They have a WhatsApp group to share challenges. They try to deal with the challenges presented by teachers. We convinced them it is a program of the government with NGO support." (District Institute of Education and Training Senior Lecturer, India)

Read India

Initially, the program faced some resistance from teachers, according to one education official. He noted that some teachers argued that they already had an activity-based program that tested children's learning levels (Nanhi Kali, an afterschool program). Officials say they convinced teachers largely with the use of class level monitoring data. As one official noted, *"[Teachers] were surprised to know the low level of learning."*

Monitoring visits by Cluster Resource Persons (CRPs) participating in the program also helped convince teachers in another way, by providing better support to teachers.

"I know that someone is coming to see my class, so I take it seriously. CRPs used to be postmasters. They used to just deliver messages to the teachers. Now they go to the classroom, observe the classrooms, share ideas with the teachers and in cluster meetings." (Program Staff, describing the changing the attitudes of teachers)

PRP

In Pakistan, system actors felt compelled to ensure that teachers adopted the new program once they received a directive to do so from their provincial secretary. Nearly all district-level officials noted that expectations were set by senior management and that once an official letter was signed, it was simply their job to ensure that teachers attended trainings and used program materials in schools. One district official further noted that, *"viewing implementation first-hand allowed me to see just how effective the program was. This convinced me to work hard to ensure implementation could go forth as intended, despite difficulties with religious groups in my district."*

4.   System monitors performance relative to stated expectations

Most programs cited monitoring as key to successful implementation. All programs involved district-level staff in program monitoring and some involved direct monitoring by the central government. There was variation among the programs as to whether the monitoring process was led by the government or by the program.

SERI

The State Ministry in Chhattisgarh, India, played several roles in monitoring schools. The first role was to reinforce the district monitoring system. The State Pedagogy Coordinator in Chhattisgarh explained his philosophy: *"Once you monitor, people will implement. If you stop monitoring, people will stop."* He also explained the way the state team interacted with the district-level monitoring team:

"The [cluster resource coordinators] collect information like are all schools making use of the library? Is there a library? Are teachers trained and using the Room to Read methodology? Are they following the schedule at the right pace? There are 38 [cluster resource coordinators]. My team has a call center, and my team would call one person in the [coordinator] group to receive information." (State Pedagogy Coordinator in Chhattisgarh)

As the State Pedagogy Coordinator alludes to above, members of the state team were directly involved in analyzing the data produced by the district. An additional role for the state was to monitor the implementation of the program in schools directly. The State Pedagogy Coordinator set up an online process for the state office to monitor the availability of books in a sample of schools. He explained his process:

"A few schools said they had zero books. We shared this info with the district. We sent a memo that all books should be available in the school. We will punish people if books are dumped in one place. Books travel: District – block – cluster – school. Somewhere along the chain it gets dumped."

At the local level, cluster coordinators were responsible for monitoring 15 to 20 schools. The coordinators visited schools regularly with a checklist to complete about library use and use of Room to Read's teaching methodology. They gave feedback directly to the teacher in addition to compiling reports from their visits. Cluster coordinators met at the block level twice a month, with around 30 coordinators per block. One meeting each month was to discuss the last school visit, while the other discussed pedagogy in general. The cluster coordinators' data from monitoring visits were compiled and analyzed by Room to Read. The data show that in February 2019, toward the end of the program, 215 out of 500 schools were visited by a cluster coordinator.

The process for improving the performance of schools that were not implementing the program well began with a request from Room to Read, based on its data analysis. District officials would take action based on this request. One district coordinator in Baloda Bazar explained that he had a formula for reprimanding teachers but had found that an observation visit from him was sufficient to motivate the teacher. He said, *"If the boss is there, the work gets done. I am the boss. If the boss is not active, then the program won't work."*

Read India

The State level government officials in Karnataka, India, emphasized the centrality of monitoring performance in two ways. First, the State officials supervised the role of lower bureaucratic structures such as district/block resource person and teachers in monitoring the performance. Data on school performance was collected either through online dashboard or visits by district officials. A High-Level State official discussed their role in monitoring:

"If you do not monitor the program, it will not be a success. There should be hand-holding of teachers and CRPs [cluster resource persons]. The dashboard has helped a lot. We would monitor, to see who has entered data."

Second, monitoring was presumed to be a success factor (see the excerpt above) and helped understand the program's impact. The monitoring data was used to improve implementation of the program. The use of data involved collaborations across departments within the government structures such as districts and blocks, as well as collaborations with program partners, such as Pratham. A High-Level explains the way collaborations involved in acting on monitoring data:

"Cluster level meetings were held for 12-15 schools. At the meetings they reviewed the status of implementation. By sharing data, Pratham coordinators helped us. We held two meetings with Pratham to get the analysis. I personally counseled the DIET [district institute for education and training??] principal that schools were not performing. Bella discussed the project with one teacher in one underperforming school in the district. He interacted with parents too. Parents said the children were very active. After getting data we disseminate data to districts and block"

District officials were involved in the operational aspect of monitoring. The intermediary role of district officials between the State and the schools involved receiving guidelines from the State and communicating it to the schools. District officials also planned school visits, supervised the assessment, made sure data was uploaded on the online dashboard, and prepared report on the progress of the schools. A district official explains their role in monitoring:

"We go for the visits after schools. We want to know how are the schools doing? How are the children doing? And then give the report to the BRC [block resource person]. We gave statistics and progress cluster and block wise. If they find any loopholes. They took action"

PRP

For PRP in Pakistan, respondents noted that even with government monitoring, it was essential to have program staff visit schools and support teachers. A district education manager for PRP shared their monitoring process:

"There are about 1,200 schools in Dera Ismail Khan (800 all-boy schools); 300 schools were part of PRP. We received information/data from mentors on a regular basis (on their visits, the learning outcomes, the teacher processes, about materials, everything). This was special type of information, focused on these schools. It was accurate and precise and focused more on quality than our normal data."

For PRP, a high-level minister in Pakistan's Khyber Pakhtunkhwa province also explained how government's monitoring the program at different levels encouraged engagement:

"We had a steering committee. We looked at project implementation and checked how they are meeting their targets or not. Overseeing the entire process from teachers up to provincial level. This project was monitored from two levels: (1) project internal; (2) District Education Officers who were involved. We used to receive attendance sheets of each and every training by the end of the day so we could see if trainings were occurring as expected."

NEI +
High-level officials reported that the EMIS team, which also reviews the assessment reports and identifies schools/districts that need more support, monitors the performance of the program in terms of student assessments:

"[They were] absolutely able to achieve the goal. We made a comparison between the last NEI project and NEI+ when we had a joint meeting with Sokoto. We realized that 95% of the pupils in Bauchi initially were unable to read in the local language fluently. But after the midterm, then we realized that substantial numbers [of students] are now able to read. But also, the EGRA showed that improvement and that has impacted the performance."
(Director Planning at the Ministry of Budget and Economic Planning)

The program also supported monitoring of the resources allocated to basic education:

"[The] State education account was introduced by NEI and continued by NEI+—it is a tool that tells you at a glance what [the] government is investing in education. We see where stuff is going and see where there is a need to shift more resources. Especially higher education was getting more than required. Basic education, which is the foundation level, had not been funded very well. Before we didn't know what was spent, but now we are able to identify gaps and have been able to increase budget to basic education. Basic Education was later given more priority than higher. Initially it was very low." (Permanent Secretary of State Ministry of Education)

Lecture Pour Tous
Discussions with interviewees in Senegal largely focused on monitoring at the regional level. This reporting kept the DEE at the central level informed of program progress. Multiple respondents at the district and regional level elaborated on this multi-layered reporting

system. As one IEF explained, *"[Monitoring is done at the school] using coaching and training tools. After this, we write a report to a higher level."*

Two regional officials then elaborated,

"Each IEF makes a report. Then we combine reports (at the IA level) and submit to MoE. Reports are submitted every 3-6 months." – Inspector, Kaolack

"I send reports to the national-level Ministry: financial reports, technical reports, how project was run and activities submitted (circular report). The project sends someone to make sure everything is monitored properly and communicates to the MOE. The report is submitted to the DEE" – Regional education official, Kaolack

Training activities are also heavily monitored, with respondents noting that DEE send supervisors to evaluate the trainings, and that training reports are shared at the local and national level:

"For every training, a report is generated and sent to the IA. The IA then collects reports from all IEFs and compiles in a final report, that is sent to the DEE" – IEF, Fatick

Additionally, in regards to pre-service work, the coaching and communication director at DFC discussed conducting observations of training modules, and was able to share the responses of teachers regarding their reactions to the modules on national language teaching.

EQUIP-T

In Tanzania, the districts incorporated quality assurance using a government system that was adapted to the EQUIP-T model. A district Director of School Support in Bahi, Tanzania, explained how the system works. He said that when school quality assurance officers visit schools, they have to undergo several areas of review. First, they look at the student learning outcomes, including the 3Rs. Second, they look at how the teachers use the teaching and learning methods. Third, they observe the curriculum and curriculum modules. Fourth, they look at the leadership; water, sanitation, and hygiene facilities; and community activities. Having the core activities of EQUIP-T be a part of the government's quality assurance review allowed the intervention to be more meaningfully incorporated into the daily activities of the quality assurance officers.

Based on the student learning outcome focus in Tanzania, a District Education Officer in Dodoma explained the process that he and his colleagues took to address issues in student learning:

"When I'm informed that there is a problem among some pupils, I go to that particular school. I discuss with teachers to know exactly what the problem is. We try and set the solution. One solution is to see that all classes have enough TLMs [teaching and learning materials] so that the class can be conducive and…attractive to the pupils. I have instructed

all schools in their capitation [to] make sure they allocate 5,000 TSh to buy manila cards [colored paper on heavy stock] so that they can make good and attractive teaching aids."

The integration of EQUIP-T activities into the government's quality assurance system meant that EQUIP-T-supported districts were more likely to focus on instructional quality issues. However, we were not able to find evidence that these monitoring visits occurred very frequently.

Tusome

In Tusome, monitoring against expectations was an important aspect of the program. At the central level, MOE staff periodically visited schools to observe the program in action. While they were at the school, they visited classrooms, talked to teachers and head teachers, and observed students reading.

"There is a monitoring exercise going on right now and many staff are in the field. Yes, they are going all around observing those children who are going into grade 4. They will compile a report on what they see, including challenges, and they will share it when they come back. I took part in a number of them and saw children in grade 3 reading in Swahili and English. In my 40 years as a professional, I had not seen anything like it." (MOE official, Kenya)

Although these high-level visits by the central ministry are important for government ownership and buy-in, the biggest responsibility for monitoring lies with the CSOs at the subcounty level. Tusome gave these officers tablets preloaded with apps to use for classroom observation. CSOs visited classrooms, observed teachers, and input information into the tablet using the apps. The information was then uploaded to a server and became visible to education officials throughout the Kenyan education system who had access to an internet connection.

Many respondents stated that access to implementation data through the dashboard was one way that Tusome differentiated itself from other programs and contributed to the program's success. It was also a way to hold CSOs accountable for doing their job. Their supervisors could check the dashboard to see whether they had been doing their observations and because of the embedded global positioning system (GPS) codes, the supervisors also could tell whether the CSOs were actually at a school or not when the observation was uploaded.

"…the dashboard is something that they used. Cabinet Secretary is using Tusome as an example to recognize counties that are doing well, and call out those that are not." (Program staff, Tusome)

"…also, the issue of tracking of lesson observation was a big plus since it was digital and there was GPS. In Kenya, [CSOs] will tell you that they have gone to observe, but 'from

their homes.' The issue of GPS to track is something that should be sustained." (County education staff, Nyamira County)

Ghana Learning

Implementation oversight for Ghana Learning was primarily a program responsibility, though system actors regularly monitored schools. For example, at the district level, the circuit supervisors who are part of the district education directorate monitored schools and supported teachers. Learning gave these supervisors tablets with a monitoring tool, which they used during school visits to record information.

Learning also developed a dashboard that allowed MOE and GES staff to review information on schools and districts. Several respondents highlighted that this was very useful for them when it came to monitoring performance and making decisions. One official noted that if poor performance was found in a school or district, they would communicate with the district to determine what was causing such performance:

"When we reviewed schools, we would review low-performing and high-performing schools in order to understand where some struggle and what others are doing well."

Additionally, they noted:

"I had a tablet, my inspectors had tablets. And we could review the dashboard to understand what was happening in the schools (using real-time data). We visited schools, even from the top levels. It was quite exciting to see the children reading in their mother tongue."

Lastly, district teams monitored schools. For example, when Learning was interested in evaluating whether the program's materials and training were being used correctly, it reached out to district teams, which supported the program in monitoring implementation and extending assistance to the districts and schools that required it.

NEI+

NEI+ monitored the implementation of the program at the school level by gathering data through School Support Officers (SSOs), who are based in the LGEAs. The SSOs provide constant supervision to teachers to ensure that they are using the pedagogical training in classrooms. According to some interviews, SSOs reported the data from classroom observations directly to NEI+ and LGEA teams do not have direct access to the data, which is managed and analyzed by the EMIS team.

5.   System institutionalizes changes in policy, procedures, or practices as a result of the program

We found multiple examples of policy and practice changes across programs cited by respondents, including changes in curricula, language policy, monitoring procedures, budgeting and planning, and teacher professional development. While this list is not

comprehensive, interviewees noted multiple components of the programs that were institutionalized. For example,

- EQUIP-T was successful in instituting school-based teacher Communities of Learning as a national policy in Tanzania.

- PRP was successful in revising the national curriculum for reading. Ghana Learning had a lasting impact on approaches to school monitoring and coaching.

- NEI+ saw major success in the incorporation of more basic education funding in State budgets.

- Lecture Pour Tous successfully informed changes to the design of a new bilingual education model for the country, as well as plans for the national reform to convert public primary education to this bilingual model and in conjunction launch a national reading program incorporating nearly all core components introduced by Lecture Pour Tous .

- Tusome saw institutionalization of all major aspects of their programming: book production, procurement, and training.

- Changes to expectations for coaches in Senegal, specifically how coaches observe lessons and provided feedback to teachers, was made explicit for the new inspectors at IEF and is being incorporated into the cahier de charge des IEF, currently under review by the inspector general for primary education.

- SERI and PRP influenced assessment and curriculum in areas that were not included in the original program.

- Ghana Learning influenced the coaching and monitoring practices of GES, including the continued use of Learning's tablets and monitoring dashboard.

SERI

For the SERI program in Chhattisgarh, institutionalization took place at the level of the program in Baloda Bazar District and at the level of state education policy. The first level was when the Chhattisgarh State education system carried out elements of SERI in Baloda Bazar with only minimal support from Room to Read. SERI was designed to transfer responsibility to cluster coordinators from Room to Read's Literacy Facilitators. Several respondents said that the role of the cluster coordinators in monitoring schools was the critical step in transferring ownership of the program to the government.

The second level of institutionalization involved SERI influencing some state policies. These policies were affected by the success of SERI but were not adopted directly from SERI. According to the State Pedagogy Coordinator, the program was unlikely to be expanded in its current form, in part because the state did not want to be seen to be favoring the

approach of one organization (i.e., Room to Read) and because additional technical capacity would be needed for the state to adopt Room to Read's approach to training literacy instruction. However, SERI influenced Chhattisgarh State to use the fluency and comprehension categories in its state-level assessments and to assess every student in the state. Based on this assessment and ASER results, Chhattisgarh education officials found out that 50% of students were not reading at grade level. As a solution, the State Pedagogy Coordinator advised new measures to improve students' skills.

"In April, the whole month will be dedicated for a reading campaign. Five mothers will do an assessment and can approve if students are able to read or not. If not, they are taught by teachers in April. There will be a two-day workshop to design a reading campaign. Room to Read is supporting them technically in these and other areas going forward." (State Pedagogy Coordinator, India)

Furthermore, during our interviews, state education officials were making plans to use Room to Read's *Tarang* student workbook across the entire state of Chhattisgarh, not just the four districts supported by the program. However, this goal was subsequently dropped.

Read India
While one MOE official emphasized the degree to which the State education department internalized the program by centering support delivery around cluster resource persons (who bought in because they were engaged fully in the program), other interviewees noted challenges to institutionalization. One interviewee emphasized that teachers internalized the methods, which was one core goal of the program, but could not speak to the second goal, involvement of the state in the program. Another interviewee noted that there was not a clear plan for the continuation or the program, nor guidelines to integrate it into regular teaching.

PRP
Among the main focuses of PRP were to impact policy change and to ensure institutionalization of program practices. Continuous professional development was one of the practices that provincial governments most consistently incorporated as a result of their engagement with PRP. Most MOE ministers in Pakistan commented that the CPD model was the most significant contribution of PRP. According to a key high-level ministry counterpart, the PRP CPD framework garnered significant support from the state ministry, receiving an allocation of 1 billion PKR per year for its implementation in provinces. Before PRP, there was no such allocation for training and professional development. He further added, *"It was beautifully integrated into the entire teachers' program in the province—truly sustainable."*

On a provincial level, the KP and Azad Jammu and Kashmir governments also incorporated the PRP CPD model and were working to expand its use comprehensively throughout each district in their respective provinces. To this end, MOE officials in KP expanded CPD into districts that were not part of the PRP.

"There are four elements in our CPD model (classroom observations—every head teacher must observe one teacher per week; self-reflection—one page daily per teacher; quarterly meetings among teachers; professional development day—eight days per academic year, covering English, Urdu, math, and science). We identify weak areas based on assessment results. Self-reflection and classroom observations, and quarterly review meetings were all part of the PRP CPD model. The main difference was the professional development day, which was a bi-weekly meeting under [the] PRP model." (Deputy Director of Provincial Institute for Teacher Education, Pakistan)

Furthermore, a key high-level ministry official in Pakistan noted that as a result of PRP, the system now approaches issues differently:

"Due to PRP, [the] government now uses needs assessments to inform training (diagnostic assessment). They also helped [with] ethnic and gender inclusion in government materials. [The] government now uses steering committees that include people from all levels of the system (whereas they used to just be for senior officials)."

Additionally, according to Pakistan's MOE officials, system officials in Sindh province changed the pre-service training as a result of the program. Previously the pre-service training was only nine months; by late 2020, it was a four-year program. Further, practicum was minimal in the pre-service training, but it was increased to be a major part of teachers' four-year pre-service training. Also, in Sindh, the curriculum was revised with PRP support to focus entirely on reading (a factor that was missing prior to PRP).

Lastly, the KP Secretary of Education noted several important changes enacted in his province as a result of PRP's successful implementation. First, all schools in the province changed to a phonics-based approach to reading and provided annotated scripted lessons to teachers as a guide. Furthermore, the KP government revised textbooks up to grade 5 to incorporate reading, writing, listening, and speaking standards, with one KP ministry official noting, *"I have taken as a policy, to the textbook board, the need to incorporate more PRP supplementary materials into all schools."*

EQUIP-T
In Tanzania, according to district officials, one of the significant changes at the policy level after EQUIP-T was integrating the program's focus on the lesson into the quality assurance system:

"Now everything that EQUIP-T was doing is part of their normal activities…. They have a normal quality assurance team within the school. It incorporates the lesson activities within the system." (Director of School Support, Bahi, Tanzania)

A district-level director of teacher professional development in Tanzania's Dodoma region also said that another essential change was to the teacher training system, which happened as a result of the program. Proper follow-ups were also more likely to be carried out to

ensure that teachers followed the instructional procedures they learned in training. The director was candid and said that this approach was not fully successful at first. Teachers were initially reluctant and felt confused and overwhelmed about how to relate EQUIP-T methods to the common materials in Tanzania. However, gradually, the teachers recognized EQUIP-T as part of the broader education program. In these instances, Academic Officers and Ward Education Officers helped teachers relate the typical Tanzania materials to the EQUIP-T materials and worked with the head teacher to delegate decision-making power on what materials to use to the school and the individual teacher, so that teachers did not feel overwhelmed or overworked. In some cases, this meant a reduction in emphasis on the key elements of EQUIP-T, particularly materials utilization.

Tusome

Tusome influenced several changes in policies, procedures, and practices within the Kenyan education system. The most frequently mentioned example of institutionalization of the Tusome approach was in the government's development of the competency-based curriculum. According to government officials, the CBC borrowed not only from Tusome's pedagogical approach, but also from its approach to book development. Respondents also mentioned that the government took up Tusome's approach to book procurement and distribution and that the government began using its own budget to procure Tusome books. Although Tusome's focus was on in-service teacher training, one government official pointed out that the pre-service teacher training institutions have integrated the Tusome approach into pre-service programs based on the light-touch Tusome pre-service intervention. Other less frequently mentioned ways in which Tusome changed government policies and procedures included changing the designation and job description of Teachers' Advisory Centre Tutors to CSOs, who focus much more on supporting teachers; adopting Tusome approaches to teacher training in implementing PRIEDE; and influencing changes to the teacher appraisal process.

"*We are moving to embed Tusome approaches into the pre-service curriculum… approaches have been infused into the teacher training program so as teachers graduate, they are already conversant with the Tusome approach…. We will continue to align the competency-based curriculum with Tusome approaches from grades 1–4. [We] have developed materials, which also takes [the] same format as Tusome books. We have a learner book and teachers' guide."* (High-level MOE officer, Kenya)

"*The teaching methodology has been incorporated into the normal teaching at the school level. Making joint work plans has also been improved. Policies are made at the national level and the work at the county level is implementation. I see a lot of Tusome in the new CBC, meaning the methodologies were adapted and incorporated in CBC."* (County Director of Education, Kenya)

Ghana Learning

Although some respondents were concerned about the sustainability of the program, other suggested that the program did influence the coaching and monitoring practices of the MOE and GES. For example, the GES and MOE are expanding the provision and use of tablets and dashboard to 10,000 schools under a World Bank-funded project (GALOP). In day-to-day monitoring processes, the district director of education shared, "The program has become part of our municipal/district action plan. It has been integrated into our routine monitoring. Therefore, each CS has the program as part of his/her monitoring itinerary. When you go to a school, all the things that you monitor, this program is part of that. That is taken care of as normal, routine duties." (District Director of Education, Ghana)

NEI+

Ministry officials spoke about the inclusion of the NEI+ program in the budget. They considered it a huge achievement to institutionalize the program components in basic education. Officials shared that there is a strategic plan in place that includes efforts to ensure the sustainability of the program. The Director of Planning at the Ministry of Budget and Economic Planning, in particular, was very optimistic for the continuity of the program, stating,

"[The] Ministry had a paradigm shift in terms of how best to share the utilization of the budgeted resources to improve learning outcomes. Making the SBMC very functional to carry out some oversight work to support the reform NEI+ project... And so many factors—does not only relate to physical structures but also the quality of teachers and the support that they provide to the pupils as they come to the school. The MOE was able to respond positively…".

Speaking about the impact that the program has shown in terms of students reading skills, the Permanent Secretary of State shared that there is a change in policy now to provide more opportunities for pupils to read. The training provided to teachers, head teachers, and SSOs will also prove to be useful in ensuring continuity of the program on ground. However, the Permanent Secretary of State also expressed some concerns about the lack of follow-up mechanisms to ensure that the sustainability plan is implemented.

Lecture Pour Tous

The National Reading Program (PNLSE) codified nearly all of Lecture Pour Tous' most important elements in its approach to reading education. Additionally, the director of INEADE spoke at length about the Ministry's initiative on bilingual education (MOHEBS)-seeing this as a direct policy shift that resulted from Lecture Pour Tous. The MOE internalized (or intends to internalize) Lecture Pour Tous's approach through the MOHEBS. According to the director, Senegal has a long history of trying to address learning in national languages. Lecture Pour Tous moved the issue forward by analyzing the situation more concretely through language mapping and by demonstrating the impact of teaching in the first language. The design of MOHEBS is built in part on Lecture Pour Tous's experience and

research findings and provides a plan to scale bilingual teaching. In addition to this, Lecture Pour Tous helped develop a framework and standards defining categories of reading proficiency for early grades in local languages that are aligned to the Global Proficiency Framework.

6. <u>System capacity in key technical areas was reinforced or developed by the program</u>

Every program involved significant capacity building of government staff. Respondents reported that staff capacity was built through the development of TLMs, teacher professional development, coaching, reading pedagogy, evaluations, assessments, and use of data. Each program differed in its approach to building system capacity. EQUIP-T and SERI were both intentional in the way that responsibility and ownership were transferred gradually to the government. In both cases, the process ended with the government being able to take on some—but not all—of the program components. Similarly, PRP and NEI+ were successful in supporting provincial or State governments to implement parts of the program. Both Ghana Learning and Lecture Pour Tous were successful in improving processes and inputs for materials development at the central level, and Lecture Pour Tous was successful in building the capacity of inspectors (IEFs) and other teacher support actors to take a more supportive approach to coaching. Through Tusome support, the Kenyan Government was able to run the large components of the program independently, although continuing to rely on donor finance.

SERI

The main focus of capacity building in SERI was the government cluster coordinators. Room to Read's Senior Program Officer in Chhattisgarh explained that cluster coordinators were involved from SERI's beginning. In the first phase of the program, the "I do" phase, the coordinators observed the Room to Read Literacy Facilitators but did not make decisions. In the second phase, "We do," the coordinators monitored teachers alongside the Literacy Facilitators. SERI gradually reduced the number of supporting Literacy Facilitators from 50 in year 1, to 25 in Year 2, and eventually to zero by the "you do" phase. In the "we do" phase, data collected by the cluster coordinators were analyzed by the Room to Read program officer, who reported results to the MOE and district offices. The "we do" phase was successful in transferring responsibility to the government system, although this transfer of responsibility was limited to the cluster coordinators. The intention was to follow the "we do" phase with a "you do" phase in different districts, where the role of the Room to Read program officer would be removed. However, district officials said that the initial training program could not be run independently of Room to Read. Therefore, full institutionalization of the program was not achieved.

PRP

In Pakistan, many federal and provincial ministers acknowledged the technical guidance that PRP provided in reforming the curriculum and teachers' CPD. PRP had an instrumental part in developing the capacity of two institutions: (1) the provincial institute of teacher

education (teacher professional development framework on the ground) and (2) the curriculum wing (curriculum and development of books and teacher development). A Federal Minister of Curriculum explained how the MOE built the capacity of the institutions:

"PRP did not create a curriculum but instead provided technical guidance to the MOE for curriculum design. Alignment with PRP methods was key (including PRP materials being included in the curriculum and textbooks and item banks being provided to teachers for reading assessment)."

PRP assisted the curriculum wing in policy areas, such as revising the scheme of studies to include reading time and dedicated reading spaces in schools. The program also supported development standards for Urdu and Pashto. The Federal Minister of Curriculum shared the following about textbook revision:

"During our textbook production, ethnic sensitivities and gender sensitivities were introduced because of the influence of PRP. All books used to have only Muslim characters (and mostly males). These social issues were contentious at first but it's important. I told the textbook people to go look into the societies and see what is there—the materials should represent the society."

PRP also supported provincial governments in developing their own curriculum, textbooks, and teacher training that aligned with the curriculum.

"The changes that they worked with were the ones that were identified by the government as well. PRP saw how the government was thinking, and they were looking into the issues the same way but brought new skills to see how the goals could be met. We all saw the same target, but they showed us how to hit the target directly. They made the targets and approach clear to us. We had the funds for implementation, but we were unable to carry out the needs assessment. Once they implemented, we were able to replicate it very easily." (Former Provincial Secretary of Education, Pakistan)

"Training to supervisory staff (on sensitization of approach and materials) was important. This will help us carry out the work moving forward—as a supervisory/support role in addition to the training from mentors/teachers. I, myself, took the training and there were many things that were new to me. It helped me as a mother as well because it helped me to teach my son." (Director of Training and Coordination, Federal Directorate of Education, Pakistan)

EQUIP-T
In Tanzania, the EQUIP-T program helped the government build its capacity to collect data at the school level. This raising of skill levels impacted how the system worked in general, not just the pedagogical changes relevant to EQUIP-T's work.

"Previously they were only able to collect data at the district level, but now you can dig to the school level. Dropout rate was now also at the school level, not just the council level. Pupil–teacher ratio was 50 on average, but some schools were much larger—this disaggregation shows where intervention was needed. Can now do teacher redeployment. Example, in first data collection, found some 1:300, this was in the Tabora region; Kariwa council had an alarming [pupil–teacher ratio] overall. There was great variation; when they started deployment, they received a lot of money for redeployment." (Director of School Support Services, Makuru, Tanzania)

Another official pointed out that the program built the capacity of the system to continue teachers' professional development.

"Teachers capacity building made an impact on success. Program strengthened the system to make more awareness on the continuous professional development of teachers. We have that in the GoT [Government of Tanzania] system, [but] mostly it wasn't done continuously." (High-level official, Tanzania)

According to a senior education official, a weakness of EQUIP-T was that some central cadres of education officers were not trained by the program. Similarly, the program's inability to reach all levels of the system programmatically was seen as a weakness (although this was a purposeful program decision, to focus on regions, districts, and schools).

"We had these trainings but not specifically on particular officers. That's a weakness. They didn't help to build the capacity at the ministry level. LGA level yes, especially for planning and in-service training, but at Ministry/PO-RALG level, no." (Senior education official, Tanzania)

Tusome
Kenyan officials acknowledged that Tusome developed the capacity of actors within the education system in several areas, such as development of TLMs, evaluations, assessments, reading pedagogy, coaching, and information and communication technology (e.g., how to use tablets for coaching and how to use the dashboard). The main recipients of training were teachers and CSOs, which led to some groups of officials, such as the Quality Assurance and Standards Officers (QASOs), to feel neglected. The training and capacity building of teachers and CSOs were seen by many as one of the biggest contributors to the success of the program. This capacity building included training, following up regularly, and ensuring that they had the necessary resources and tools to do their work. In response to the question about the biggest contributors to Tusome's success, education officials from the central level to the subcounty level mentioned the training and capacity building:

"Training of teachers can be ranked number one…then also the support of CSOs. Teachers discovered the trainings were friendly and helped the learners. Also, during the training, they got reimbursed very quickly, and that was motivation." (Subcounty officer, Kenya)

"Teacher training was the most important aspect." (County TSC officer, Nyamira County, Kenya)

Ghana Learning
The Learning activity focused on building the capacity of all stakeholders from the Ministry of Education (MOE) and Ghana Education Service (GES), including partners in the National Council for Curriculum and Assessment, National Teaching Council, and National Inspectorate Board. It also provided national core trainers, circuit supervisors, and district teacher support teams with intensive capacity building in reading instruction to support their roles as trainers, monitors, and school-support providers. Interview respondents overwhelmingly acknowledged that capacity strengthening was a cornerstone of the Learning program. The Ghana Learning program held workshops regularly to train staff across the spectrum, including school coordinators/coaches, master trainers, and teachers. One specific example that some officials provided was that by engaging stakeholders at different levels for developing and validating the learning materials, the system's capacity for writing and evaluating materials was enhanced.

"At every level, from the highest down to the schools, we had capacity building for all staff."
(Coordinator of Private Schools [GES])

"If we have a very good literacy curriculum, we owe it to the Learning project. If we have good training teams; if we have good writing panels; if we have very authentic book validation process…we owe it to the Learning project." (Former Executive Secretary of National Council for Curriculum and Assesment (NaCCA)

NEI+
In addition to building the capacity of teachers, head teachers, and SSOs, government officials shared that the program provided them training in different areas. For instance, district officials were trained as master trainers so that they could train and supervise teachers. The director of the State Universal Basic Education Board (SUBEB) and some MOE directors were also involved in the pedagogical trainings. Many of them were grateful for the training provided by NEI.

"For myself, I've been very proud. It has really saved my neck. For the first time, I was there as an admin. I was just picked, and it was tough. They were there and gave us a series of training. Give us much encouragement. That guided me and my colleagues to

know the working materials. We were all there, even the accountant. EMIS [officers] were highly trained also." (Director Planning, Research and Statistics)

Ministry officials also provided some examples where the program built the capacity of the system. NEI+ supported SUBEB by setting up budget and planning tools. This improved the capacity of the budget and planning team within SUBEB. The EMIS Director also commended the data collection system built by NEI+, which helped them conduct school census using tablets and create reports using Power BI.

Lecture Pour Tous

Most interviewees recognized that Lecture Pour Tous significantly developed the capacity of schoolteachers, school directors, and coaches. According to a district official in charge of school supervision, *"The first change I noticed was the increased quality of training for school directors and teachers. I (directly) trained school directors and teachers. [Lecture Pour Tous] has a very original approach. There is a lot of training but also the close supervision is a new element"*. Another district official (IEF) served in both a coaching role and as trainer of trainers. They explained that the training is usually done in 2 to 5 separate sessions, which includes two sessions of initial training and 2 to 3 follow-up sessions and further capacity building. They also shared the following:

"The trainers manual (the greatest contributor to the program's success) clearly establishes the structure of the training. The trainer just has to apply/execute it. We ask teachers to answer questions based on their own experiences. The guide has clearly indicated: activities of the trainer/activities of the trainees/expectations and outcomes. We learned a lot from [Lecture Pour Tous] about planning for a training. They have a day for going through how to plan the training, with distribution of roles and tasks. This gives a good idea of/visibility into what we are going to do. Because of this, time is used effectively, and people are motivated/the environment is good."

This official added that the training and follow-up mechanisms introduced by Lecture Pour Tous are well received by teachers. There is a shift from inspection-oriented monitoring to a softer and kinder approach towards teachers. The teachers are more relaxed, and they are happy to invite them to their classes- *"[When] Teachers see me they are so happy and laugh and say 'hi'. Teachers call me into their class themselves."*

In addition to training, some staff also shared that their capacity was improved by working on module development, on the training of CRFPE (teacher training) instructors, and by conducting classroom observations. The Director of INEADE shared that Lecture Pour Tous built their capacity by improving approaches and methods for evaluating materials and providing better pedagogical specifications for procuring materials. While interview responses were largely positive towards Lecture Pour Tous' training and capacity building,

one interviewee from DFC stated that from his understanding, Lecture Pour Tous did not provide any training for the DFC staff.

7.   Counterparts play substantive roles in the implementation

Senior government officials in all programs said that the buy-in and meaningful involvement of government officials was critical for success. However, total government responsibility was restricted to selected areas of a program. For example, the government took over responsibility for program monitoring in SERI, but not program design or teacher training. In PRP and Lecture Pour Tous, some of the most substantial roles in the program were played by government teacher trainers. In EQUIP-T, NEI+ and Tusome there was a substantial role for government counterparts not only in implementation but also in planning at regional and district (Tanzania). State (Nigeria) or county (Kenya) levels.

SERI

In India, the government played a substantial role in implementation. A USAID Senior Education Specialist said that her most important role with SERI was convincing the government to give orders to the schools because it was necessary for this communication to come from the government, rather than Room to Read. The Education Specialist further advised that she helped set up a State Resource Group composed of identified government experts who operated like a think tank and advisory body. This advisory group was actively involved in giving feedback to SERI. She further explained how her involvement led to genuine collaboration between USAID and the government

"I was always there for [Room to Read]—whatever they needed. This award was made under a mechanism that required co-design. I tried to build with them all these steps—a scaffolded approach, [the government had] as much flexibility as possible in state selection. I left it to them, flexibility in the budget, too. When it came to scale, I helped them think through the strategy." (Senior Education Specialist, USAID/India)

Room to Read staff spoke of the importance of the government's role in the success of SERI.

"Ownership was key. The point person (in the Ministry) didn't change. He has taken full ownership. The district focal person was there for the full time and was motivated. Because there was a need—most schools were underprivileged. The State Pedagogy Coordinator was present from the very beginning, even before the demonstration results came in. He knew it worked. He had been to seminars at [the] country level." (Room to Read's Senior Program Officer in Literacy, Chhattisgarh, India)

"Programs may be good, but if the enabling factors are not there, it won't be implemented. We got the enabling factors right—the clarity of agreement from state to district to schools. They are in sync with what is expected. The cluster coordinators and support staff became critical. If they faced challenges, they could get help to address them. The state already has

a focus on improving learning outcomes. They saw that the work in Raipur was successful. They saw this in the field and the monitoring data. The State Resource Group was formed which review[ed] content and look[ed] at the approach." (Room to Read's State Manager, Chhattisgarh, India)

At the district level, a District Project Coordinator in India said that their role was to coordinate with the state and Room to Read. Several interviewees said that the significant involvement of staff at the district level was also critical for program success.

"District officials were very enthusiastic. They selected teachers to be resource persons, which involved a whole day of screening. They got training from Room to Read to become master trainers. The mentoring and coaching were provided by Room to Read. Program should be accepted by district/state. It should not be forced from the top. From the beginning, block and ground-level people should be involved, and their resources should be involved. If only external experts are involved, everything might collapse. Some local officials should be there so we can build their capacity." (State Pedagogy Coordinator, India)

PRP

In Pakistan, ministry counterparts regularly spoke about the strong working relationship and joint activities conducted between PRP and the ministry. The majority of this work, however, consisted of PRP providing technical expertise and support to the ministry under the program's system development component. For example, although the reading curriculum is a ministry document, its development was a joint effort that was strengthened by the ministry and PRP working together.

"The entire curriculum development process included experts from schools and districts (to ensure buy-in from the start). Once the curriculum was approved, orientation workshops were held for head teachers and language teachers. For the first time, 100% of Urdu teachers in primary schools received training on use of the new curriculum. PRP facilitated these workshops/trainings and provided materials, but technical leads were all from MOE." (Former Director of Curriculum, Federal Directorate of Education, Pakistan)

Furthermore, Pakistan ministry officials said that the education departments were involved at multiple stages in the curriculum revision, needs assessments, teacher training, and textbook revisions. There was also close communication between MOE officials and PRP.

"It was also clear that during the planning of the project, the program was aware of the entire government, and therefore, not a single place was left unattended. PRP worked with every department and level within the government. In a way, this project was a replica of the entire department, with something new to say." (KP Ministry official, Pakistan)

EQUIP-T

The EQUIP-T program was criticized by some respondents for having limited the engagement of the curriculum body in developing and reviewing materials initially. The

materials were ultimately approved by the curriculum body, and due to that increased engagement, EQUIP-T was noted to be more effective at involving the government in materials development compared to another donor-funded program, Tusome Pamoja:

"We would want to see more engagement of TIE staff in developing TIE materials. I do not think they are involving us to a large extent. At a certain point they are asking us to review them. They should ask TIE to be part of writing of the zero draft of the material." (Director of Materials, Tanzania)

Education officials in Tanzania described how they collaborated with EQUIP-T and how their collaboration strengthened their relationship to be more effective. The MOE officials acknowledged that the relationship between the government and EQUIP-T was clearer than the relationship they had with Tusome Pamoja. The closer relationship with EQUIP-T occurred in part due to EQUIP-T's provision of budgetary support to the government. This budgetary support mechanism increased over the life of the program, and these mechanisms increased the government–program collaboration and allowed government counterparts to play a large role in the program planning.

"EQUIP-T funding goes through [the] government financial system. If I was asked how much EQUIP-T had given to LGAs, they know. This was transparency. They are open on the level of funding and the activities. Can check exchequer system to see how much it is. Tusome Pamoja—we don't know how much is set for the project, while EQUIP-T has. GoT don't know how much is going to training and LGAs for Tusome Pamoja, but we know for EQUIP-T. We don't know how much is TLMs; LGAs, we don't know where they print. Even the sustainability of the project—EQUIP-T can be more sustainable than Tusome Pamoja because of engagement and openness." (Assistant Director of Secondary Education, Oganga, Tanzania)

The ministry and the program held monthly and quarterly meetings to discuss and agree on EQUIP-T activities, which led to the program being perceived as more willing to adapt its interventions to match government priorities than other donor-funded interventions. EQUIP-T even added a construction activity to its program design in response to a request from the GoT.

"EQUIP-T didn't have a construction activity. GoT saw it as a challenge and a high need to have the construction component within the program. EQUIP-T wasn't very ready to do the construction. Please let's sit together and let's agree, GoT will do a presentation and show the challenges. The program wants to strengthen the 3Rs while students are sitting under trees. Build classrooms and strengthen the 3Rs. Improve the environment as well as improve teacher training, build the planning capacity of districts, provide TLMs. To us, the GoT, the construction was key to improving outcomes. It was a negotiation that we did with EQUIP-T and later we agreed to work together in construction." (High-level ministry official, Tanzania)

Tusome

Ministry officials in Kenya played a substantive role in implementing Tusome; in fact, it was implemented mainly by ministry officials. Many education officers at the county and subcounty levels saw Tusome activities as part of MOE activities.

"Tusome is our program…no way to differentiate between Tusome and government. Its activities are part of our core mandate…[We] monitor the teachers, support the teachers in induction and training, [and] oversee the distribution of TLMs to ensure that every child has a textbook." (County education officer, Elgeyo-Marakwet County, Kenya)

Government officials were involved in every aspect of Tusome. There were two high-level decision-making bodies: (1) a National Steering Committee (NSC), chaired by the Minister of Education with members from USAID, RTI, and senior management from partner agencies; and (2) a technical committee chaired by the Principal Secretary with representation from various education agencies, including the TSC, KICD, and Kenya Institute of Education (KIE). These committees met regularly at the central level to make decisions about Tusome's direction.

"As an implementing partner for USAID, RTI had to work with the government. RTI respects [the] government and recognizes that ownership lies with the government. Relationship was largely cordial. Each organization had its own policies and respected it, but government policy is the law of the land. Sometimes ran parallel with the government, but still lots of intersection with Ministry….We had several different committees: National Steering Committee, which is chaired by minister or cabinet secretary and included membership from USAID, senior management from partner agencies, and maybe some [chief executive officers]. Also, the National Technical Team, which was chaired by the Principal Secretary and had technical staff from KICD, TSC, KIE, etc., met regularly and were involved with decision making and planning. NSC was responsible for policy making." (Central MOE official, Nairobi, Kenya)

"When Tusome was starting, we had a week-long design project with USAID and DFID to come up with a project design document. We came up with design document jointly with them and described what Tusome meant. We came up with a structure; steering committee chaired by Minister, technical committee of midlevel staff headed by national coordinator, and also county-level structure. This formed a very good structure where the Minister of Education chairs a meeting every three months to get to know how it's being implemented, and seek help for struggles. Government appointed a technical coordinating committee, derived from many departments of education. This was a joint work plan in every activity." (District education manager, Kenya)

According to Tusome's National Coordinator from 2016 to 2017, her biggest contribution in terms of her role was to create harmony among various people working with Tusome, which was helpful to solve challenges and address gaps. "We realized early on that we had left out

the primary teacher colleges, and this was a gap. We were able to bring them on, which was important for sustainability. I think KNEC [the Kenya National Examinations Council] was also brought on board early as well."

Ghana Learning

By design, the majority of Learning's activities were intended to be implemented by the MOE and GES, with technical direction from the Learning team. The main exception to this approach was the program's use of Ghanaian vendors from the private sector to oversee the distribution of materials to schools.

In order ensure consistent MOE and GES involvement, the program created a learning advisory committee that included representatives from all levels of the MOE and GES, as well as university representatives. This committee was responsible for formally approving books and materials. It also played a significant role in validating the materials being developed in local languages.

Although the government was involved, much of the implementation was still seen as being led by Learning staff. As one high-level GES official noted:

"I would say that about 20% of the work was done by GES, 20% by MOE, and 60% by staff of Learning. This is so because they had implementation roadmaps. Any time they wanted to embark on these roadmaps, they would invite GES to come in and validate and determine whether or not their plan would benefit schools. After that validation, the program would get GES/MOE to get teachers and other stakeholders to get on board (e.g., training)." (Coordinator of Private Schools, GES)

NEI+

State and LGEA officials that were interviewed shared that they were involved in the program on multiple levels. Many State officials were directly involved in training the teachers and supervising program implementation at the LGA and school level. In fact, many officials discussed their involvement in monitoring and quality assurance of the program. At the local level, SSOs also oversaw implementation in schools. Speaking about the effects of training on teachers, one official shared the following:

"When the teachers went [to their classrooms] they portray and replicate what we taught them. The methodology, the child-centered approach, and other things, and the Let's Read, and Mukaranta. We went to supervise them and that's how we found out that it made a difference."

As noted in previous sections, high-level officials spoke about their involvement in strategic planning and ensuring sustainability of the program.

Lecture Pour Tous

Two interviewees from DEE noted the large role they and other DEE staff played in Lecture Pour Tous activities related to teacher training and coaching. DEE staff served as master trainers (of other trainers) and also provided supervision during all teacher trainings. To support coaching, DEE staff trained all IAs and IEFs on how to use the Lecture Pour Tous coaching guide.

One respondent from INEADE confirmed that DEE handled overall coordination of Lecture Pour Tous activities, noting that INEADE supported technical coordination related to materials, assessment, and research. INEADE staff carried out language mapping survey work with technical support from Lecture Pour Tous and with participation from the MOE's directorate for national languages. INEADE also oversaw materials development and worked with private publishers to produce the LPT books. INEADE supported research activities, including an impact evaluation conducted at three timepoints (baseline, midline, and endline).

The Department for Communication and Training (DFC) was primarily involved in pre-service training matters, the development process of work plans for pre-serving training, development of modules, and supervision of the CRFPE instructor training. According to the Director of Training and Communication at DFC, the Technical Assistance Team to Lecture Pour Tous relied on the DFC to work with CRFPE to determine how best to "fit" the modules into those existing programs. DFC accomplished this by working through the CRFPE coordinator who oversaw the work of the concerned CRFPE (note that the coordinator is one of the directors selected from among the participating centers).

The DFC also worked with the technical assistants for Lecture Pour Tous to train staff at IA, IEF, and CRFPE to develop strategic communication plans at their levels and trained them on communications activities, including working with local radio stations for messages to parents promoting national language literacy. Interviews with district education managers (IA) also indicated their key role in supervision of activities at the regional level.

"[I am the] institutional focal point for the IA (regional body). I coordinate activities between the IEFs. [Including] reports, supervision and technical committee meetings. [I am] In charge of the technical committee mission to implement what the Instruction Academy is giving them, following an action plan." (District Education Manager, Senegal)

8.   Challenges faced in implementation

Although systems actors commented on the success of every program, they also highlighted some of the challenges in implementing within the contexts they worked in. Although there were several challenges, there was no clear theme(s) that emerged across the programs.

EQUIP-T

In Tanzania, the failure of teachers to consistently implement the school-level teacher Communities of Learning meetings that were an essential part of EQUIP-T's intervention

was a consistently cited problem. Due to some gaps between the centralized, external-to-school trainings and school-based meetings, and delays due to Local Government Authority planning for the trainings, the overall effectiveness was diminished. The delays stemmed in part from the late release of program funds to these Local Government Authority structures by the Ministry of Finance; the transfers were out of the control of EQUIP-T. In addition, it was difficult for midlevel civil servants to provide sufficient high-quality supervision to schools, which also reduced the effectiveness of the instructional improvements that EQUIP-T focused on. The following quote came from a district with very large class sizes and limited teaching staff in some schools:

"*Training at school level—these are not happening regularly. This is due to lack of teaching staff and being busy with school. There are other challenges with WECs. Some have been moved. When they receive new ones it takes time to build their capacity until they master how to support EQUIP-T.*" (Director of school support, Chamwino District, Tanzania)

In addition, the use of funds was highlighted as a significant challenge by a ministry official, who further explained that funds were tied to approved plans and activities. He elaborated that after an agreement was reached on activities at the national level, the activities were then rolled out to the regions. However, one of the challenges was minimal flexibility in making changes to adapt to the school realities in the EQUIP-T-approved activities. The ministry official also commented on the lack of funding for salaries, refreshments, and training expenses, which typically are not covered by donor-funded programs like EQUIP-T.

One respondent noted that EQUIP-T initially had goals that did not align with those of the government, as many programs are accused of. EQUIP-T's ministry national program counterpart also raised some challenges that he experienced with the program implementation design and EQUIP-T's lack of flexibility, another complaint typical of programs that are tasked with following donor guidelines and terms of reference. However, he also noted an important way in which EQUIP-T changed over time—i.e., adapting to government structures—to more effectively support the government.

"*One of the weaknesses of EQUIP-T [was that] they were trying to establish a separate project structure. The first two years they didn't do much until they decided to adapt the government structures. This brings ownership and understanding. This most importantly builds capacity of existing staff. Once they decided to adapt the structure, it fast-tracked the implementation.*" (Ministry national program counterpart)

The Ministry national program counterpart's experience above demonstrates one of the key changes that allowed EQUIP-T to be implemented effectively. The program recognized the need to work within existing government structures after two years of difficult implementation because of the initial mismatch in priorities between the program and the GoT.

SERI

In SERI, two main challenges were identified: (1) the lack of teaching staff and (2) staff attrition.

"Poverty and migration are issues in hard-to-reach areas. The number of teachers is less in these areas. We have placed some guest teachers in schools where there are no teacher[s]. We intensify program[s] when children are there [and not migrating]—extend the amount of time on the program each day. And they also work during summer vacation." (District Project Coordinator, Barwani District, Madhya Pradesh State, India)

"There is a challenge in proper monitoring of the program because there was an LF [Literacy Facilitator] and a BRC [block resource coordinator] and CRC [cluster resource coordinator]— now the BRC and CRC are no longer there [i.e., the position no longer exists]. They are trying to address this challenge. Deputy block-level people are the only ones supporting the project. But others may come later." (Senior District Institute of Education and Training lecturer, Champawat District, Uttarakhand State, India)

PRP

Several high-level ministry counterparts in Pakistan also noted challenges working with PRP in the initial stages. One of the biggest challenges was obtaining buy-in at the beginning of the program. Like EQUIP-T, this challenge was partially due to the initially misaligned priorities between PRP and the government (i.e., the program tried to introduce a new way of working and was focused on teaching reading while reading time did not even exist in the government's scheme of studies). Further, ministry officials stated that there were coordination issues from PRP's side at the beginning of the program. These issues were rectified through a combination of efforts, including more consistent working meetings with government officials, strong support from USAID, focused alignment of PRP activities with government priorities, and changes in PRP provincial-level leaders.

Tusome

For Tusome, Kenyan education officials reported some challenges with the program— especially in the early years and, more recently, as the program was entering closeout. As Tusome prepared to close, CSOs were no longer getting reimbursed for school visits. As a result, many spoke about their challenges with sustainability and donor dependence.

"…challenges in the past and that we continue to face is that of resources. When support from USAID ends, what happens? As a government, our resource base will not allow for it. We might reach a level whereafter the program might not continue. Especially in the monitoring and evaluation, Tusome has been very effective. Not sure whether we will be able to continue. Shall we be able to manage?" (MOE official, Nairobi, Kenya)

"Also, when fare reimbursement stopped, pace of support has slowed down." (County officer, Nyamira County, Kenya)

"First challenge is transport, now there's nothing. But even before, it was not enough—400 shillings, but sometimes it cost 800 shillings, and we have to make up for it." (Subcounty officer, Elgeyo-Marakwet County, Kenya)

Because Tusome brought many benefits to education officials who participated, such as training opportunities, training allowances, transport allowances for school visits, and tablets, those who were not part of the program felt left out.

"QASOs were doing the same work as CSOs, but we weren't facilitated. This brought some conflict. It's the same way that PRIEDE has treated QASOs—as orphans. We are expected to oversee and report, but not trained. No effective linkage between QASOs and CSOs, even with CBC. Not trained, but willing and able. If [monitoring and evaluation] had been domiciled around QASO, [it] would have been better." (County officer, Siaya County, Kenya)

"Quality assurance and middle-level management officers were initially left out and the project worked directly with the CSOs. Later these were involved. Initial sharing and communication of an activity was not done well. Little financial facilitation of the officers to support program activities." (County officer, Nyeri County, Kenya)

Tusome's approach to developing and procuring textbooks threatened the stronghold that publishers had in this sector, which also brought a legal challenge. Publishers, some of whom had allies within the Kenyan government, sued to overturn what they saw as the Government of Kenya allowing a monopoly by a United States firm to produce books.

"There were people who fought the program…For example, publishers felt they were losing because government was giving away free books. Felt that there was a monopoly." (MOE official, Nairobi, Kenya)

However, it was not just publishers that felt their position and authority within the Kenyan education system was threatened. Government agencies, such as the teachers' unions and KICD, also opposed Tusome's methods.

"Others were worried that the program prevented teachers from being creative because of the teacher guides. Or that fast learners did not benefit and were being held back. There were also concerns raised about creating a dependency by those worried what happens when the program ends? Other donor partners were concerned that the government was giving all this attention to Tusome and perhaps not to their own programs. Even within the ministry, KICD was not too happy. But these challenges were not insurmountable, eventually everyone had to come around." (MOE official, Nairobi, Kenya)

By the start of 2020, Tusome's fate was becoming increasingly unclear. The program was in its final year and, accordingly, some resources were being withdrawn with the expectation that the MOE would be taking these expenses on. At the same time, the MOE was

continuing to roll out its new competency-based curriculum to all schools, and there was some uncertainty and confusion around Tusome's place within the new curriculum. County and subcounty officials compared the CBC to Tusome, and many thought that Tusome was better. Although the MOE had distributed a circular that attempted to explain how the Tusome books were to be used within the new curriculum, field officers still expressed misgivings.

"When CBC came, the books were not compatible with what MOE gave. Head teachers were saying there are lots of inadequacies with the new books, and teachers were confused." (County officer, Nyamira County, Kenya)

"Teachers' attitudes were changed; they were excited. I have attended some of the trainings and teachers took to it. When it comes to CBC, people did not understand. They found it confusing." (County officer, Nyamira County, Kenya)

"It is only now that [we have] clear guidance on when to use Tusome books and CBC books. The MOE sent a circular that has now cleared things up." (County officer, Nyamira County, Kenya)

Although Tusome staff tried to engage with the MOE in the development of the new curriculum, they did not get everything for which they advocated. For example, time for Kiswahili was reduced, and the program had to adjust to make up for that.

"The new curriculum substantially changed things… Changed the number of lessons for Kiswahili from five lessons a week to three a week, which gave less time. Therefore, curriculum body asked us to reduce the content. So, we had to reduce some significant areas including in the books…some key areas were removed. The third thing around the curriculum reform that has impacted [the program was that] they introduced a new subject called Literacy, which brought a lot of confusion. From the design we were engaging with the curriculum body, but they wouldn't listen. But now they are seeing it was a mistake. The teachers got confused. It's just a big mess." (Tusome program staff, Nairobi, Kenya)

Ghana Learning
Several respondents shared that there were challenges related to the language policy. The MOE was interested in promoting reading in local languages, and the program helped the Ministry implement it in 100 districts. However, there was tension in terms of public buy-in. Although there was reluctance to teach in local languages instead of English from some stakeholders, others were concerned that only a few languages were promoted in the program. The former NIB Executive Secretary explained, "here was a perception that some people were being left out or that it was hegemonic. 'There are 47 languages, why are you people in education supporting 11?' was pushback received from some people." The Director of Early Childhood Education also shared that while this program was running, she was also

responsible for coordinating the "Jolly Phonics" program that promoted reading in English for the remaining districts that were not included in the Learning program.

Other challenges that were highlighted by some respondents included delays in timelines and some minor delays in the provision of materials. The former NTC Director also spoke about lack of clarity in the beginning about expectations of working collaboratively with the USAID-funded program. They explained that the GES/MOE had previously worked with DFID-funded programs for which DFID provided funding and the GES/MOE implemented the activities. In the Learning project, they wished that the GES/MOE provided clarity early on about grants, partnership, and technical assistance.

NEI+

One of the challenges that some respondents shared was the program's limited engagement with certain stakeholders. According to some officials, NEI+ did not involve the administrators of the LGEA directly and, therefore, was not able to garner support throughout the state. The Secretary, Bauchi Agency for Nomadic Education shared that NEI+ was not able to engage them as a full autonomous agency within the SUBEB, sharing, "They don't involve the agency in some of their activities. So even in the closeout they did not involve us, and when they came to distribute computers and other things, they did not consider us, and we are the ones that are supposed to be working and we handle marginalized groups."

In another interview, the Director, Planning at the Ministry of Budget and Economic Planning shared that the program was unable to collaborate with UNICEF. According to the director, UNICEF had been working for several years in basic education and had a great influence on how things were done. The director pointed out that effective collaboration with UNICEF was an area that NEI+ needed to work on. The director also highlighted that NEI+ needed to take a more holistic view of different sectors that are linked with education: "If you improve education, how does that affect the water sector; effective linkage between education and other sectors like WASH, primary health etc."

In terms of implementation, a major contextual challenge raised was the shortage of teaching personnel at schools. For instance, in hard-to-reach areas, there may be only one teacher who is also acting as headmaster. Respondents noted that if this teacher was called away for program activities no one was able to care for the pupils. Additionally, several schools have teachers managing large classes and multiple grades. The secretary of the Bauchi State Agency for Nomadic Education suggested that unless the Ministry addresses this shortage of teachers, program implementation will be difficult.

The EMIS Director shared additional challenges with respect to teacher training. One issue was that the expectation for the master trainers to cascade the training was not implemented as expected. Another challenge noted was that the teacher management

information system application designed by NEI+ was not ready in time to be implemented alongside the training.

Lecture Pour Tous

Some officials shared challenges in processes and logistics. For instance, the materials that were initially delivered to schools had a mismatch with the actual number of students. Learning from these challenges, the program sought IEF's help to allocate books in the proper quantity to individual schools, as IEF they would have more accurate enrollment figures for each school in comparison to the statistics and projections maintained at the central ministry. From where he sits, the Head of Communications at DFC felt that cumbersome administrative procedures were significant constraints ("lourdure administrative"). For example, he said that activities that were developed or planned in October were not fully approved until January or February. From his perspective, these kinds of delays meant that at times a window for an activity to achieve greater relevance was missed.

In terms of teacher training, an official shared one challenge specific to multi-lingual communities, where the first language of teachers and students varied. This made it difficult to identify a single language for instruction to train teachers on.   One teacher trainer also expressed concerns about scaling trainings after technical support to Lecture Pour Tous ended, saying they will need support and training to sustain the program efforts.

When asked about challenges the program faced with implementation, an inspector in the DEE shared his belief that the program handled many things directly and then informed local counterparts (IA and IEF) rather than developing and planning more activities together. The director of INEADE also highlighted that Lecture Pour Tous worked at the decentralized levels (IA and IEF), but felt that it would have been even better if the program put more emphasis on building autonomy at the decentralized (and school) levels. The director of INEADE shared, "We need to break the existing culture of 'waiting for orders from the MOE (or from the project).'"

The Head of Communications at DFC also felt that, in his opinion, Lecture Pour Tous communication activities did not sufficient make use of national level mass communications techniques to build broader support among the population, and importantly, among education sector leadership, for national language literacy.

9.   Program- (not Ministry)-led communication with districts and schools

A key hypothesis of our research was that learning at scale would be more effective, and institutionalized to a greater extent, when governments rather than programs led communication about program activities with districts and schools. As a way to explore this hypothesis, we sought evidence to the contrary—that program's rather than governments led this communication effort. There was moderate evidence to support this alternative

hypothesis from EQUIP-T, Lecture Pour Tous, and SERI. The overall conclusion is that communications are given authority if they are issued by the government, rather than by the program. However, the program may play a role in initiating or drafting communications and, in some cases, is responsible for issuing the communications.

For instance, in SERI, Room to Read officers in Chhattisgarh emphasized the government communications that were initiated by the state government, whereas the State Pedagogy Coordinator said, *"Whenever support from Room to Read is requested, a communication is issued by the state."*

. In Senegal, staff in the DFC (that was not leading program implementation), said that although the program collaborated with the DFC, at times they felt that they were working in isolation and taking the lead on activities. They suggested that instead of serving in a supportive role, DFC should have been in the lead role.

District-level officers in Pakistan also stated that although they were responsible for ensuring that teachers attended PRP trainings, their role with the program was limited and PRP primarily acted independently at the district level.

Similarly, under PRP, a district education manager said that they did not monitor subdistricts and schools. This point was confirmed by a Minister of Teacher Training in Pakistan who explained, "[There was] *No real monitoring of performance. We are in charge of certification and quality assurance, so we had a mandate to monitor trainings, but this didn't happen formally. PRP did all real monitoring and we trusted their data."*

10.   Necessary inputs or resources are reliably made available

The availability of key inputs and resources was one of the biggest contributing factors to Tusome's success in Kenya and was also brought up in discussions about Lecture Pour Tous, NEI+ and Ghana Learning.

Tusome

Respondents in Kenya noted that among the inputs mentioned, the provision of high-quality books for every student, tablets for CSOs, travel reimbursements, technical guides, and timely payment of allowances made the most impact. The inputs and resources were especially appreciated at the county and subcounty level.

"[Tusome] achieved close to 100% of mandates. But it is the distribution of textbooks. If there's anything that they have contributed, it is distribution of text[s] at a 1:1 ratio. That is the most important. Also building the capacity of the teacher… over the holidays all the teachers come together at their zone level for training. That alone assures almost 100% attendance because they are close to where teachers are. Some other programs, teachers have to travel far, and they don't come. Also, the small token that the teachers and trainers get is appreciated. The support, small token, [and] facilitation fees given to trainers and

trainees. PRIEDE is saying why don't you refer to the Tusome model?" (County officer, Elgeyo-Marakwet County, Kenya)

"Also having the tablets helped. We can use it to collect other kinds of data. You can take photos or even Google. Signed a contract with Tusome to say after two years, you will take it [the tablet]." (CSO, Elgeyo-Marakwet County, Kenya)

"Tusome has really helped. In each month up to last September, they gave fare reimbursement. Helped to fuel our movement. RTI also gave us tablets and that has been very effective in providing feedback. Gave us training notes, which was important for us to understand better. Since the reimbursement stopped, though, it has been difficult to get to schools; we are mostly using our own resources. This has slowed our movement. Ministry does not provide transport to schools." (CSO, Nyamira County, Kenya)

"Making [each] teacher a skilled teacher was most important, then the books—getting to 1:1. Facilitation of movement for CSOs so we can go to support school[s]. Some people just train and leave you and you never see them again. There was constant interaction with Tusome. You'll give feedback that books are no longer there, and they will respond. There was constant interaction." (CSO, Siaya County, Kenya)

Officials at the MOE also recognized the importance of the inputs made by Tusome.

"Hard to pick [the biggest contributor to success] because they are all important and are the building blocks. Training is key, but also so is the provision of materials at 1:1 ratio. First time that has ever happened." (High-level MOE official, Nairobi, Kenya)

"Other than the pedagogy, availability of teaching and learning materials, where and when they were required." (Key MOE counterpart, Nairobi, Kenya)

Ghana Learning

In addition to capacity building efforts (including workshops and trainings for MOE and GES staff, as well as teachers and coaches), the Learning program was praised for their provision of books and learning material for students, as well as paying for experts to consult on material development. In addition, the program provided scripted lessons to the teachers. The Director of Early Childhood Education commended the efficiency of the program in delivering books (which was done through Ghanaian vendors in the private sector):

"I was overwhelmed by the books/materials. The way they distributed the materials to the schools. They had these vans and they delivered materials right to the doorstep of the students. The materials were very attractive, and students loved them. We print out books. They send them to the depots, and they stay there forever and forever and never get to schools." (Director for Early Childhood Education)

The GES and MOE provided human resource support and vehicles for school visits.

NEI+

In Nigeria, respondents noted that teacher training was provided by the program. NEI+ also provided tablets for the SSOs. Some of the LGEA officials highlighted that the state did not provide many resources—either material or personnel. However, other officials noted that the state contributed toward the development of materials and has financed book printing.

Lecture Pour Tous

In terms of necessary inputs and resources, the Ministry and district officials shared that there were enough materials provided by the program. District officers also appreciated the materials in the local language. Lecture Pour Tous provided vehicles for monitoring and the government has requested for those to be donated to them. The training of teachers, coaches, and staff were considered a significant contribution from the program.

11. Biggest contributors to success

All respondents were asked about the biggest contributors to the success of each program. The most common responses across programs were good collaboration—and alignment of goals—between the program and the government; effective pedagogical approaches, including phonics; and training that made teachers' task simpler.

For the SERI program in India, the following five factors emerged:

- The design of the program was well articulated, with structured lesson plans. The level of structure was particularly important for the large-scale implementation.

A scientific approach to reading, based on neuroscience, distinguished SERI from other approaches and was the foundation for its effectiveness.

Teacher training was practical, with a focus on how to implement methods in the classroom.

The demonstration phase of the program was important to convince government officials and teachers that the program was effective.

The government insisted that SERI should not be seen as a Room to Read program; i.e., government ownership was key.

For Tusome in Kenya, the following aspects were cited by more than one respondent as the biggest success factors:

- Availability of books in a 1:1 ratio.

Teacher training, including using a pedagogical approach; making teachers' life easy; motivating teachers by the success of students; paying allowances on time; and holding the trainings close to where the teachers live.

Follow-on support and coaching.

Collaboration and communication with government.

For PRP in Pakistan, the biggest success factors were noted as follows:

Monthly teacher inquiry group meetings, which were cluster-based teacher meetings to share experiences with the program, including difficulties and solutions.[30]

PRP's coordination and strong working relationship with the MOE, including building on government needs and plans.

PRP's strong leadership and consistency at the top levels of management.

Strong teacher and student materials, teacher training, and the phonics-based approach to teaching reading.

For EQUIP-T in Tanzania, the key success factors were as follows:

- Timing. EQUIP-T filled the demand for training and materials in the 3Rs that other programs and the government were not able or prepared to do at that time.

School-based communities-of-learning approach for in-service training. Although many local leaders were concerned about the quality of this approach, the national leaders were more likely to support it because of the reduced cost.

Changing the EQUIP-T budget management system to allow government involvement in financial decision making in program funds. This is a program decision that none of the other interventions included and produced substantial goodwill and buy-in from the government.

Working closely with the LGAs allowed EQUIP-T to be perceived as working within government structures.

Pedagogical methods applied by teachers. For some teachers, the methods included phonics and participatory teaching.

For Ghana Learning in Ghana, key success factors included the following:

- Strong collaboration between Learning staff and officials from the MOE and GES: Including government counterparts throughout the development and planning processes was essential for buy-in and continued support. As such, many stakeholders were invested in the program as it aligned with their goals.

A strong training model, combined with hands-on support (though coaching and monitoring), for ensuring that teachers were comfortable and confident in the new phonics-based methodology.

The development and distribution of engaging teaching and learning materials: Appropriately leveled readers and easy-to-follow teachers' guides were important, especially paired with Learning's efficient distribution process, which ensured that all materials were available in schools, on time

The development and use of a data dashboard that provided stakeholders with access to the information they needed to monitor program performance and inform decisions on necessary inputs and program shifts.

---

[30] Although participants found teacher inquiry groups to be highly effectives, data from PRP's RCT on the impact of professional development activities indicated that teacher inquiry groups were the least cost-effective activity (with coaching being most cost-effective) and that teachers who did not receive an assignment to a teacher inquiry group were able to compensate for their absence through alternative means (i.e., WhatsApp groups).

For NEI+ in Nigeria, key success factors included the following:

- Focus on student learning in trainings. Most of the respondents spoke about the positive impact that teacher training and quality assurance officers had on student learning specifically.

Supportive monitoring. One of the factors that contributed to the program's success was the shift in the monitoring approach. It allowed SSOs to be viewed by teachers as partners.

Collaborative decision-making between NEI+ and the Ministry. This included putting in place sustainability plans.

Changes in the enabling environment. A few respondents noted positive changes to the state's budget to provide more resources for basic education.

For Lecture Pour Tous in Senegal, key success factors included the following:

The practical approach to teacher training. Several officials appreciated that the training included opportunities for teachers to practice using the new methods and materials.

New coaching techniques. Respondents also said they found coaching techniques innovative and useful for reinforcing the training with teachers. Several respondents noted that the approach to coaching shifted from inspection to support of teachers.

Teachers' ownership of their students' learning. While the DFC had very little involvement in in-service teacher training, the head of communications at DFC expressed that the capacity building of teachers was such that they took ownership of their students' learning.

Provision of materials and the use of local language in these materials, coupled with community mobilization to spark enthusiasm for local language reading instruction among parents. The district education inspector of Kaolack shared, "what strikes me, that I like the most, is the use of local language. When they're new in G1, students can't automatically speak French. Local Language provides a little step up to help them. As the student begins learning French, you slowly pull away the LL."

Conclusions

We conclude that a number of factors are important for an education system to produce learning outcomes when it is adopting a new program focusing on literacy instruction. Our conclusions are based on findings that were consistent across the three countries studied:

1. **Programs as a priority**. It is essential that the government see the program as a priority. Programs were able to convince the government to prioritize their activities through evidence—both of the scale of the problem and of the effectiveness of the proposed solution—and by aligning the program goals with existing government priorities.
2. **Systems communication**. The priority given to the program should be communicated throughout the system. In all the countries we studied, there was a strong hierarchy in government structure; thus, communication was most effective when emanating from the MOE. However, several programs reinforced the communications using informal methods, such as WhatsApp.
3. **District-level buy-in**. District officials played a critical role in communicating expectations to teachers and convincing them to adopt the program.

4. **Support for subnational officials**. A critical part of the systems is the district and subdistrict staff who interact regularly with schools. For this interaction to be effective, these staff must be provided with capacity building, resources, and support.

5. **Monitoring**. A key component of the interaction between district officials and schools was the monitoring of both implementation (e.g., teachers attending training) and student outcomes. The process of monitoring may have been more significant to communicate the importance of the program to teachers than to provide data to inform course corrections.

6. **Transfer of responsibility**. Several programs were effective in transferring responsibility to the government system through a deliberate phased approach.

7. **Capacity Building**. Capacity building was effective through direct engagement of government counterparts in essential areas of program implementation—for example, by conducting joint activities between the government and the program.

8. **Institutionalization**. Through the above methods, programs were institutionalized in a piecemeal fashion. All programs were successful in institutionalizing some activities; none of them succeeded in achieving complete institutionalization. There is no evidence that programs were more successful on the impact on learning when institutionalized in the government system, but such institutionalization is likely to contribute to the sustainability of activities beyond the life of the program.

5.6 Cost Data

The Learning at Scale study team is still working with partners from each of the eight programs (and their lead organizations) to determine how cost analyses can be included in the final study report. At this stage, we simply introduce basic information on donors, overall contract amounts and cost implications.

Program donors and contract ceilings

As noted in **Table 80**, most (*n* = 6) of the programs included in this study were funded by USAID, one (*n* = 1) was funded by DFID, and one (*n* = 1) had multiple donors. There is a vast spread in the contract ceiling amounts provided by these donors. The program with the least amount of funding was SERI at approximately $5,800,000. The largest amount of funding was PRP at $181,000,000.

Table 80.   Program donors, contract amounts, and years

Program	Country	Imple-menter	Donor	Contract amount[1]	Date contract signed	Contract end date	COVID-19 impact?
EQUIP-T	Tanzania	Cambridge Education	DFID	$154,409,992	2013	2020	N
Ghana Learning	Ghana	FHI 360	USAID	$76,583,859	2015	2019	Y

Program	Country	Imple-menter	Donor	Contract amount[1]	Date contract signed	Contract end date	COVID-19 impact?
Lecture Pour Tous	Senegal	Chemonics International	USAID	$75,733,663	2016	2021	Y
NEI+	Nigeria	Creative Associates	USAID	$82,201,162	2015	2021	Y
PRP	Pakistan	IRC	USAID	$181,078,305	2013	2020	N
Read India	India	Pratham	Multiple		2016	2020	N
SERI	India	Room to Read	USAID	$5,824,688	2015	2020	Y
Tusome	Kenya	RTI Inter-national	USAID	$59,325,525	2015	2021	Y

*Development of the program model was funded by several donors prior to current USAID-funded program.
[1]Each contract amount was converted into real 2019 US dollars.

Cost analysis considerations

In a cost analysis, stakeholders often anticipate that the documented expenditures will align, or closely align, with the contract's ceiling amount. However, there are many reasons why the contract ceiling should not be compared to the results of a cost analysis. For example, a costing exercise focuses on the costs incurred to generate a single impact measure, but many of the programs included in this analysis had multiple outcomes, and some had other programs running concurrently with the program selected for this analysis. Our selected outcome measure—grade 2 fluency—may not reflect the entirety of a program's impact or the entirety of a program's costs (Dhaliwal et al., 2013). Also, a costing exercise may not include costs related to a program's management, general operations, reporting, and development (e.g., sunk costs into developing teacher training and teaching and learning materials). Instead, the costs analyzed represent a very specific type of cost for a single, specific outcome measure.

Additionally, there are contextual, country-level, and contractual influences on costs that restrict the ability to directly compare the cost-effectiveness of these programs to each other. As shown in **Table 81,** each program may start with a different impact baseline, each may be embedded in a different system, and the operational cost for each program may vary by country (Levin et al., 2018). For example, the wage level of skilled labor in one country may be higher than the wage level of the same type of skilled labor in another country. Country-level cost influences include a government's parameters on teacher per diems required for activities such as teacher trainings and its directives to publish teaching and learning materials in-country. For example, teaching and learning materials may be less expensive to publish out of country, even when the extra cost of shipping is added into the

total. However, to comply with a country's directive to publish in-country, organizations may have to pay the higher in-country prices and potentially provide capacity-building sessions to in-country printers. Organizations implementing these programs often have little influence on these national governments' decisions, although the decisions are likely to have substantial impacts on program costs.

However, the contractual influences on costs may be the most important. The structure of donor contracts can limit a program's ability to be as cost-conscious as possible. For example, for USAID-sponsored programs, funding is not necessarily related to the size of the country's education system. Instead, USAID contract amounts for flagship education programs are typically influenced by congressional allocations for education by country rather than by the number of students in that country or involved in that particular program. These funding decisions are well beyond the scope of the implementing organizations but often have substantial impacts on per-student costs. Finally, donor changes to contracts can force organizations to react quickly and may restrict their ability to make better financial choices if more time were available (e.g., paying less to ship teaching and learning materials by sea, which takes longer, than by air).

These contextual, country-level, and contractual influences can affect an organization's procurement and per-student allocations. Given these influences, the purpose of this costing exercise is not to determine which program was the most cost-effective. Instead, this research seeks to describe the programs' expenditures and cost drivers as a means of learning from each other and affecting how future programs are designed and implemented.

Table 81.   Limitations to and complexities of cross-program comparisons

Cost influences	Examples
Contextual	Different impact baselines Embedded in different systems Operational and personnel costs vary by context
Country-level (government)	Set per diems for training and support Directives such as guidance to publish books in-country
Contractual	Contract amounts are based on external factors Contractual changes in program reach or scope

Source: Levin et al., 2018.

Finally, COVID-19 has exerted an additional influence on this cost analysis that limits the ability to compare results across programs. All but one of the programs included contract years 2020 and 2021. Because of the pandemic, some of the programs' final impact estimations (i.e., endline findings) remain unavailable as of mid-2021. In these cases, the Learning at Scale team expects to use the proximal "success" (i.e., intermediate midline findings) for the cost analysis (Dhaliwal et al., 2013).

Program life cycle

One of the most interesting aspects of cost analyses is the variation in the programs' life cycles. A program life cycle provides a visualization of the general phases of a contract and helps frame the expenditures needed for a cost analysis.

Looking at the variation in start-up time among the programs is also interesting. Start-up time includes more than just setting up an office and hiring staff. It can also include the development of teacher training and support models and TLMs, and at times includes major program redesigns.

As previously mentioned, six of the eight programs were impacted by COVID-19; the other two programs ended before COVID-19 began to affect implementation. For some organizations, the pandemic disrupted implementation and endline data analysis of their programs ($n = 3$). For others ($n = 3$), program closeout was affected.

Similar to the number of contract months, the dosage within these contracts also varied. Some of this variation was dictated by government literacy regulations. That is, the governments of the countries where these programs operate have different regulations regarding the minutes of literacy instruction per day. For example, Kenya requires 30 minutes of literacy instruction per subject, while Ghana requires 90 minutes. Because the programs must work within these government systems, differences in the number of minutes of literacy per day can impact the pedagogical structures of different programs.

Scale

One of the qualifications for inclusion in the Learning at Scale research was that programs had to be working at scale or nearly at scale. Because of the population and geographic size differences among the countries included in this research, "working at scale" could look very different for different programs. Tusome is the only program analyzed here that was working across all public primary schools in a country. SERI and Read India were both working within the same country (India) but working nearly at scale looks very different for these two programs. SERI is working within four states (i.e., most of Uttar Pradesh and portions of Uttarakhand, Madhya Pradesh, and Chhattisgarh), although similar Room to Read programs are being implemented elsewhere in India. The Read India program included in Learning at Scale is operating in particular districts in one state (Karnataka), although the program is also being implemented in multiple states in India in collaboration with either state or district administrations.

This variation in scale does not infer that one approach (i.e., covering an entire country versus targeting specific districts) is better in terms of strategy or efficiency. Indeed, targeting specific districts as Pratham has done may have the benefit of directing resources where they are most needed, quickly. However, the differences in scale can have implications for costs. For programs covering tighter geographic areas, coordinating activities such as teacher trainings might be logistically easier and less expensive than for programs that are spread out over a wider geographic area. Additionally, distributing the necessary teaching and learning materials is certainly more complex at a national scale than at the district or regional level.

All of these considerations will be taken into account throughout discussions on how to most appropriately conduct cost analyses for these programs.

6 DISCUSSION AND RECOMMENDATIONS

Having reviewed program features and elements for eight of the world's most successful, large-scale early grade reading programs, pored over interview responses for hundreds of participants, analyzed data from hundreds of classroom observations and thousands of student assessments, and conducted in-depth cost-effectiveness analyses, we ultimately came to two key conclusions: First, these highly effective programs have compelling similarities and, at the same time, substantial differences. There appear to be some consistent features and implementation approaches that may be important in many of the most successful programs that can and should influence the sector's view of how large-scale implementation can happen. On the other hand, the substantial differences in programs across a variety of metrics make clear that there is no one-size-fits-all option or silver-bullet approach to improving outcomes at large scale. Ultimately, while some promising approaches are common across many of these interventions, there are multiple pathways to improving outcomes at scale.

Our second conclusion is also quite simple: **It is possible to substantially improve learning outcomes at scale** and, in many cases, within government systems. The sector has spent much of the past 15 years with an increasingly vehement concern about the dismal state of learning in many LMICs. We share these concerns and view the push to fundamentally change instruction in LMICs to improve learning as the key problem for our sector to solve. The Learning at Scale programs, and our findings, reveal that it is possible to make substantial changes to instruction and at substantial scale. This is particularly salient in the COVID-19 era, because concerns about learning loss are paramount. Our analyses of the scale of learning loss suggest that children may have started the 2021 academic year having lost most of the entire 2020 academic year's worth of learning (Angrist et al., 2021). Adding this COVID-19 learning loss to the limited learning in a typical year, our view is that there is nothing more essential right now than taking what we know about how to improve learning outcomes at scale and investing heavily in programs that work.

Based on the findings of this report, we present the following recommendations in three main sections: Effective Program Design, Program and Instruction Considerations, and Systems Issues.

6.1 Effective Program Design

- **Invest in Learning at Scale-type programs in response to COVID-19 learning loss.** COVID-19 has caused substantial learning loss in systems that were already producing poor learning outcomes. The Learning at Scale programs are examples of the sort of large-scale, highly effective programs that should be the primary focus of policy makers, donors, and implementers. We recommend that designing,

implementing, and monitoring effective, large-scale programs to improve learning be the primary—if not exclusive—focus of the sector in 2022 and beyond.

- **Implement large-scale programs with rigorous evaluations.** One of the main difficulties in undertaking the Learning at Scale study was the lack of evidence about effectiveness for some existing large-scale programs and the lack of scale of many rigorously evaluated programs. To improve outcomes at the scale needed, more programs must be implemented at scale and use designs that will allow us to measure impact and rethink implementation when and if programs struggle.

- **Fund what works.** Our survey of the education sector revealed that for some multilateral and biliteral donors, we were unable to identify any large-scale, highly effective interventions, despite substantial funding investments. Donors should consider using the findings from Learning at Scale to design, implement, and evaluate future programs. Without evidence, we cannot be sure programs work. Without scale, we are not supporting the vast majority of students in LMICs who need better teaching to overcome upsettingly low learning outcomes.

- **Consider equity and regional differentiation.** We were unable to identify effective, large-scale interventions in lusophone (Portuguese-speaking) contexts or Arabic-speaking countries, and we found only one such intervention in a francophone country. While several donors find working in anglophone contexts easier, the sector should not ignore countries or portions of countries that are not English-speaking. In fact, it may be that needs are larger in non-English-speaking LMICs, and we cannot be certain that what works in anglophone contexts will work elsewhere.

- **Consider structured pedagogy and teaching-at-the-right-level programs.** Seven of the eight Learning at Scale programs can be characterized as either structured pedagogy programs (six) or teaching at the right level (one) in design. These programs, while relatively new in their large-scale implementation in LMICs, are showing substantial evidence of impact.

6.2 Program and Instruction Considerations

- **Utilize key program elements.** Our findings from across the eight programs show that ten program elements were included in several Learning at Scale interventions and identified as being key to program success. These elements are listed in detail on page 3, but we name the top five here:

- Teacher training focused on modeling and practicing new skills
- Structured teachers' guides
- Coaches' use of structured tools to support teachers
- Face-to-face methods for initial trainings
- Explicit and systematic teaching of skills using direct instructional pedagogical methods.

- **To improve reading, use phonics and spend time actually reading.** Our classroom instructional findings are based on five programs. We found that that these programs focused the majority of instructional time on reading activities, used reading materials in classrooms as the primary resource, and primarily taught using an explicit and systematic phonics-based approach.

- **Incorporate practice, modeling, and discussion for effective teacher training.** Train teachers differently, using more practice, modeling, and discussion. We observed that teachers in the Learning at Scale programs reported having more time for small-group practice and discussion. These teachers saw modeling, practice, and discussions of instructional methods as the most useful training methods. This type of focused skills-based teacher training contrasts starkly with the generalized teacher training pervasive in many LMICs.

- **Coach differently to impact outcomes.** The consensus in the sector is that coaching works, if implemented appropriately. Our findings revealed that the type of coaching observed to work was more supportive, more friendly, and focused on pedagogical improvement rather than inspection. Engaging coaches should be coupled with coaches receiving training and structured tools to support teachers.

- **Avoid some classroom instruction "don'ts."** We identified some instructional practices that were negatively related to learning outcomes within the broader interventions. Reducing time spent on reading in order to spend more time on Grammar or Assessment in the classroom was negatively associated with learning outcomes, as was using no instructional Materials compared with using Books. Additionally, while demonstration in the classroom is good, too much demonstration may limit the amount of time available for student practice and therefore negatively impact learning outcomes.

- **Use learning materials that work.** Give students books, supply teachers with teachers' guides, and provide supplementary readers. We collected qualitative

evidence that teachers and other educators saw the books, teachers' guides, and, in some cases, supplementary readers as contributing to program impact. With respect to student materials, teachers identified engaging stories and materials that allowed students to practice and connect to real-world objects as essential. Regarding teacher materials, they were better organized, easier to follow and teachers' guides with lesson plans were deemed important.

- **Monitor student progress throughout lessons.** We found that teachers across all programs showed very high rates of 'responsiveness' to student needs in the classroom. This was demonstrated by teachers consistently checking for student understanding, circulating to monitor student progress, and rephrasing explanations when students didn't understand the content being taught.

6.3 Systems Issues

- **Align with government priorities.** Work closely and strategically with governments so that the program is a government priority. Our qualitative results spoke to government leaders as being essential to program success, particularly when the program and its requirements were seen as priorities for the officials themselves. This strategy works most effectively when the program is linked directly to existing ministry priorities or initiatives. The successful programs analyzed here found ways to do that, although those ways varied.

- **Communicate through the system to increase awareness and expectations of the program.** We recommend working with government not only to gain buy-in to the program but also to communicate about program activities regularly through normal ministry channels and informal ones. Such communication can be essential in convincing teachers of the value of the new methods and materials.

- **Enlist ministry counterparts in delivering and managing inputs needed to effect classroom change.** Successful programs build clarity about how the program relates to the daily activities of midlevel civil servants in the government system. Work with the system to identify the roles of key actors in the system and set clear expectations for implementation for actors throughout the system. Monitoring can communicate the importance of expectations and provides a basis for accountability.

- **Implement specific capacity building related to school support and monitoring.** Many studies recommend capacity building, but our findings go beyond

that. Our evidence suggests that equipping low-level education officers with tools and knowledge about exactly how to support teachers implementing the program was essential and that this was most effective when government officers knew how to effectively monitor the program at the school, district, and subnational levels. Having ready access to supporting data and clear lines of sight between individual and system performance regarding implementation quality are also important.

- **Change the organizational structure to support pedagogical improvement.** We recommend that efforts to focus on teaching, coaching, support, and monitoring go beyond piecemeal trainings. Instead, programs should work closely with government to reorient the entire system so that all actors understand their roles, implement ongoing support, and continually communicate to reinforce the new instructional behaviors needed to maintain program impact.

REFERENCES

Altet, M., Sall, N., Seck, M., & Ouedraogo, I. (2017). *Teacher training based on the results of research on their actual practices in the context of sub-Saharan Africa: The OPERA project in Burkina Faso, issues and challenges.* AFD Research Paper No. 2017-41. Paris: Agence Française de Développement. https://www.afd.fr/en/teacher-training-based-results-research-their-actual-practices-context-sub-saharan-africa-opera-project-burkina-faso-issues-and-challenges

Angrist, N., de Barros A., Bhula, R., Chakera, S., Cummiskey, C., DeStefano, J., Floretta, J., Kaffenberger, M. Piper, B., & Stern, J. (2021). Building back better to avert a learning catastrophe: Estimating learning loss from COVID-19 school shutdowns in Africa and facilitating short-term and long-term learning recovery. *International Journal of Educational Development*, *84*(July), special issue with UNICEF, 102397. https://doi.org/10.1016/j.ijedudev.2021.102397

Araujo, M. C., Carneiro, P. M., Cruz Aguayo, Y., & Schady, N. R. (2016). *Teacher quality and learning outcomes in kindergarten.* IDB Working Paper Series, No. IDB-WP-665. Washington, DC: Inter-American Development Bank. https://publications.iadb.org/en/publication/12400/teacher-quality-and-learning-outcomes-kindergarten

Banerji, R., & Chavan, M. (2016). Improving literacy and math instruction at scale in India's primary schools: The case of Pratham's Read India program. *Journal of Educational Change, 17*(4), 453–475. https://doi.org/10.1007/s10833-016-9285-5

Banerjee, A. V., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., & Walton, M. (2016). Mainstreaming an effective intervention: Evidence from randomized evaluations of "Teaching at the Right Level" in India. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.2846971

Bottoms, G., & Schmidt-Davis, J. (2010). *The three essentials: Improving schools requires district vision, district and state support, and principal leadership.* Atlanta, GA: Southern Regional Education Board. https://www.sreb.org/publication/three-essentials

Brown, J. L., Jones, S. M., LaRusso, M. D., & Aber, J. L. (2010). Improving classroom quality: Teacher influences and experimental impacts of the 4Rs program. *Journal of Educational Psychology, 102*(1), 153–167. https://doi.org/10.1037/a0018160

Bruns, B., Costa, L., & Cunha, N. (2017). *Through the looking glass: Can classroom observation and coaching improve teacher performance in Brazil?* Policy Research Working Paper No. 8156. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/27962

Brussow, J. A., Gaumer Erickson, A. S., Noonan, P., & Jenson, R. (2013). *Coaching observation checklist.* Lawrence, KS: University of Kansas, Center for Research on Learning. http://www.researchcollaboration.org/page/coaching-observation-checklist

Colbert, V., & Arboleda, J. (2016). Bringing a student-centered participatory pedagogy to scale in Colombia. *Journal of Educational Change, 17*(4), 385–410. https://doi.org/10.1007/s10833-016-9283-7

Creative Associates International. (2018). *Northern Education Initiative Plus annual report, October 1, 2017–September 30, 2018.* Prepared for USAID/Nigeria under Contract No. AID-620-C-15-00002. Washington, DC: Creative Associates International. https://pdf.usaid.gov/pdf_docs/PA00WRR3.pdf

Crouch, L., & DeStefano, J. (2017). *Doing reform differently: Combining rigor and practicality in implementation and evaluation of system reforms.* International Development Working Paper No. 2017–01. Research Triangle Park, NC: RTI International. https://www.rti.org/publication/doing-reform-differently

Dhaliwal, I., Duflo, E., Glennerster, R., & Tulloch, C. (2013). Comparative cost-effectiveness analysis to inform policy in developing countries: A general framework with applications for education. Chapter 8 in P. Glewwe (Ed.), *Education policy in developing countries* (pages 285–338)*.* Chicago: University of Chicago Press.

Education Development Center, Inc. (EDC). (2018). *USAID Philippines Basa Pilipinas Program: Final project report, January 1, 2013–July 31, 2018.* Prepared under Contract No. AID-492-C-13-00004. Washington, DC: United States Agency for International Development (USAID). https://pdf.usaid.gov/pdf_docs/PBAAJ709.pdf

Early Language and Literacy Classroom Observation (ELLCO) [Slide deck]. (2006). Presented at California Preschool Instructional Network (CPIN) meeting. https://cpin.us/sites/default/files/fcab_resources/fcab_res_langlit/fcab_ll_bg/ELLCO_10.17.pdf

Fairfield, T., & Charman, A. E. (2017). Explicit Bayesian analysis for process tracing: Guidelines, opportunities, and caveats. *Political Analysis*, *25*(3), 363–380. https://doi.org/10.1017/pan.2017.14

Fairfield, T., & Charman, A. (2019). A dialogue with the data: The Bayesian foundations of iterative research in qualitative social science. *Perspectives on Politics*, *17*(1), 154–165. https://doi.org/10.1017/S1537592718002177

Fleisch, B. (2016). System-wide improvement at the instructional core: Changing reading teaching in South Africa. *Journal of Educational Change, 17*(4), 437–451. https://doi.org/10.1007/s10833-016-9282-8

Freudenberger, E., & Davis, J. (2017, December [revised]). *Tusome external evaluation—Midline report.* Prepared for the Ministry of Education of Kenya, USAID/Kenya, and DFID under USAID Contract No. AID-615-TO-16-00012. Washington, DC: Management Systems International (MSI), a Tetra Tech company. https://pdf.usaid.gov/pdf_docs/PA00MS6J.pdf

Fullan, M., & Quinn, J. (2016). *Coherence: The right drivers in action for schools, districts, and systems.* Thousand Oaks, CA: Corwin/SAGE.

Gallagher, M. J., Malloy, J., & Ryerson, R. (2016). Achieving excellence: Bringing effective literacy pedagogy to scale in Ontario's publicly funded education system. *Journal of Educational Change, 17*(4), 477–504. https://doi.org/10.1007/s10833-016-9284-6

Gaumer Erickson, A. S., Noonan, P. M., Ault, M., Monroe, K., & Brussow, J. (2020). *Observation checklist for high-quality professional development [Version 3].* Lawrence, KS: Research Collaboration Lab, Center for Research on Learning, University of Kansas. http://www.researchcollaboration.org/page/high-quality-professional-development-checklist

Global Education Evidence Advisory Panel. (2020, October 27). *Cost-effective approaches to improve global learning: What does recent evidence tell us are "smart buys" for improving learning in low and middle income countries?* Retrieved from http://documents1.worldbank.org/curated/en/719211603835247448/pdf/Cost-Effective-Approaches-to-Improve-Global-Learning-What-Does-Recent-Evidence-Tell-Us-Are-Smart-Buys-for-Improving-Learning-in-Low-and-Middle-Income-Countries.pdf

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching, 8*(3), 381–391. https://doi.org/10.1080/135406002100000512

Hamm-Rodríguez, M., Richardson, E., & Guajardo, J. (2018). Teacher motivation and behavior change: Results of the Teacher Motivation Diagnostic Tool in northern India. In S. R. Pouezevara (Ed.), *Cultivating dynamic educators: Case studies in teacher behavior change in Africa and Asia* (pp. 249–288)*.* Research Triangle Park, NC: RTI Press. https://doi.org/10.3768/rtipress.2018.bk.0022.1809.8

Handford, V., & Leithwood, K. (2019). School district contributions to students' math and language achievement. *International Journal of Education Policy & Leadership, 14*(9), 1–21. https://doi.org/10.22230/ijepl.2019v14n9a863

Harris, A. (2012). Leading system-wide improvement. *International Journal of Leadership in Education, 15*(3), 395–401. https://doi.org/10.1080/13603124.2012.661879

Joddar, P. (2018). *Impact evaluation of the literacy program-partnership approach under Scaling-up Early Reading Intervention (SERI) funded by USAID: 2018 endline evaluation report for Chhattisgarh and Uttarakhand*. Delhi: Room to Read.

Joddar, P. (2019). *Impact evaluation of Room to Read's literacy program under partnership approach in Madhya Pradesh and Uttar Pradesh, India: 2019 endline evaluation report.* Delhi: Room to Read.

Jukes, M. C. H., Turner, E. L., Dubeck, M. M., Halliday, K. E., Inyega, H. N., Wolf, S., Zuilkowski, S. S., & Brooker, S. J. (2017). Improving literacy instruction in Kenya through teacher professional development and text messages support: A cluster randomized trial. *Journal of Research on Educational Effectiveness, 10*(3), 449–481. https://doi.org/10.1080/19345747.2016.1221487

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains—Summary.* Policy and Practice Summary. Seattle: The Bill and Melinda Gates Foundation. http://k12education.gatesfoundation.org/resource/gathering-feedback-on-teaching-combining-high-quality-observations-with-student-surveys-and-achievement-gains-3/

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying effective classroom practices using student achievement data. *Journal of Human Resources, 46*(3), 587–613. https://doi.org/10.3368/jhr.46.3.587

Kelly, C. J. (2016). *A comparative case study of the role of the school district in influencing school improvement: Supporting and turning around low-performing schools* [Unpublished doctoral dissertation]. School of Education, College of William and Mary. https://doi.org/10.21220/w4bc7w

Knudson, J., Shambaugh, L., & O'Day, J. (2011). *Beyond the school: Exploring a systemic approach to school turnaround.* Policy and Practice Brief. California Collaborative on District Reform. https://cacollaborative.org/sites/default/files/CA_Collaborative_School_Turnaround_0.pdf

Leithwood, K. (2010). Characteristics of school districts that are exceptionally effective in closing the achievement gap. *Leadership and Policy in Schools, 9*(3), 245–291. https://doi.org/10.1080/15700761003731500

Leithwood, K., & Azah, V. N. (2017). Characteristics of high-performing school districts. *Leadership and Policy in Schools, 16*(1), 27–53. https://doi.org/10.1080/15700763.2016.1197282

Levin, H., McEwan, P., Belfield, C., Bowden, A., & Shand, R. (2018). *Economic evaluation in education cost-effectiveness and benefit-cost analysis.* Thousand Oaks, CA: Sage.

Main, K., & Pendergast, D. (2017). Evaluating the effectiveness of a large-scale professional development programme. *Professional Development in Education, 43*(5), 749–769. https://doi.org/10.1080/19415257.2016.1241817

McAleavy, T., Elwick, A., & Hall-Chen, A. (2018). *Sustaining success: High performing government schools in London.* Reading, Berkshire, UK: Education Development Trust. https://www.educationdevelopmenttrust.com/our-research-and-insights/research/sustaining-success-high-performing-government-scho

Meyers, C. V., Redding, S., Hitt, D. H., McCauley, C., Dunn, L., Chapman, K., Ambroso, E., & Chen-Gaddini, M. (2017). *Four domains for rapid school improvement: A systems framework.* Four Domains Series, The Center on School Turnaround at WestEd. San Francisco, CA: WestEd. https://www.wested.org/resources/four-domains/

Molina, E., Fatima, S. F., Ho, A., Melo Hurtado, C., Wilichowski, T., & Pushparatnam, A. (2018). *Measuring teaching practices at scale: Results from the development and validation of the* Teach *classroom observation tool.* Education Global Practice Policy Research Working Paper 8653. Washington, DC: World Bank. http://documents.worldbank.org/curated/en/464361543244734516/pdf/WPS8653.pdf

Mourshed, M., Chijioke, C., & Barber, M. (2010). *How the world's most improved school systems keep getting better.* New York: McKinsey & Company. https://www.mckinsey.com/~/media/mckinsey/industries/public%20and%20social%20sector/our%20insights/how%20the%20worlds%20most%20improved%20school%20systems%20keep%20getting%20better/how_the_worlds_most_improved_school_systems_keep_getting_better.pdf

Murphy, J., & Hallinger, P. (1988). Characteristics of instructionally effective school districts. *Journal of Educational Research, 81*(3), 175–181. https://doi.org/10.1080/00220671.1988.10885819

National Institute of Child Health and Human Development (NICHD). (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Reports of the subgroups.* US National Institutes of Health Publication No. 00-4769. Washington, DC: US Government Printing Office. https://www.nichd.nih.gov/publications/pubs/nrp/documents/report.pdf

Osmond-Johnson, P., & Campbell, C. (2018). Transforming an education system through professional learning: Developing educational change at scale in Ontario. *Educational Research for Policy and Practice, 17*(3), 241–256. https://doi.org/10.1007/s10671-018-9233-7

Piper, B. L., Mejia, J. M., Spratt, J. E., Betts, K. L., Sitabkhan, Y. A., Sowa, P. A., & Ralaingita, W. D. (2019). *Training teachers or robots? Unexpected findings of a 7-country teacher professional development study*. Presented at Comparative and International Education Society conference, San Francisco, CA.

Piper, B., Ralaingita, W., Akach, L., & King, S. (2016). Improving procedural and conceptual mathematics outcomes: Evidence from a randomised controlled trial in Kenya. *Journal of Development Effectiveness, 8*(3), 404–422*.* https://doi.org/10.1080/19439342.2016.1149502

Piper, B., & Zuilkowski, S. S. (2015). Teacher coaching in Kenya: Examining instructional support in public and nonformal schools. *Teaching and Teacher Education, 47,* 173–183. https://doi.org/10.1016/j.tate.2015.01.001

Piper, B., Zuilkowski, S. S., Dubeck, M. M., Jepkemei, E., & King, S. (2018). Identifying the essential ingredients to literacy and numeracy improvement: Teacher professional development and coaching, student textbooks, and structured teachers' guides. *World Development, 106,* 324–336. https://doi.org/10.1016/j.worlddev.2018.01.018

Piper, B., Zuilkowski, S. S., Kwayumba, D., & Oyanga, A. (2018). Examining the secondary effects of mother-tongue literacy instruction in Kenya: Impacts on student learning in English, Kiswahili, and mathematics. *International Journal of Educational Development, 59,* 110–127. https://doi.org/10.1016/j.ijedudev.2017.10.002

Piper, B., Zuilkowski, S. S., Kwayumba, D., & Strigel, C. (2016). Does technology improve reading outcomes? Comparing the effectiveness and cost-effectiveness of ICT interventions for early grade reading in Kenya. *International Journal of Educational Development, 49,* 204–214. https://doi.org/10.1016/j.ijedudev.2016.03.006

Popova, A., Evans, D. K., & Arancibia, V. (2016). *Inside in-service teacher training: What works and how do we measure it?* RISE Programme Working Paper. https://www.riseprogramme.org/sites/www.riseprogramme.org/files/inline-files/Evans_Inside_In_Service_Teacher_Training_CLEAN_v2016-06-22.pdf

Pratham. (2020). *Government of Karnataka partnership program 2019-20: Learning levels and program update.* Delhi: Pratham.

Pritchett, L. (2015). *Creating education systems coherent for learning outcomes: Making the transition from schooling to learning.* RISE Working Paper No. RISE-WP-15/005. https://riseprogramme.org/sites/default/files/inline-files/RISE_WP-005_Pritchett_1.pdf

Rawle, G., Binci, M., Pettersson Gelander, G., Harb, J., Jasper, P., Khan, S., Medardi, D., Romarri, A., Rorich, M., & Ruddle, N. (2019). *EQUIP-Tanzania impact evaluation. Endline quantitative technical report, volume 1, results and discussion.* Prepared for DFID. Oxford, England: Oxford Policy Management Ltd. https://www.opml.co.uk/files/Publications/8383-assessing-equip-t/opm-ie-el-quant-report-vol-i-final-sent-rev.pdf?noredirect=1

Rincón-Gallardo, S. (2016). Large scale pedagogical transformation as widespread cultural change in Mexican public schools. *Journal of Educational Change, 17*(4), 411–436. https://doi.org/10.1007/s10833-016-9286-4

RTI International. (n.d.). *Tusome Early Grade Reading Activity* [Impact story]. Research Triangle Park, NC: RTI. https://www.rti.org/impact/tusome-early-grade-reading-activity

Seidman, E., Kim, S., Raza, M., Ishihara, M., & Halpin, P. F. (2018). Assessment of pedagogical practices and processes in low and middle income countries: Findings from secondary school classrooms in Uganda. *Teaching and Teacher Education, 71,* 283–296. https://doi.org/10.1016/j.tate.2017.12.017

Shannon, G. S., & Bylsma, P. (2007). *Nine characteristics of high performing schools: A research-based resource for school leadership teams to assist with the school improvement process* (2nd ed.) Olympia, WA: Office of Superintendent of Public Instruction. Retrieved from https://files.eric.ed.gov/fulltext/ED499819.pdf

Social Impact (2018). *Ghana early grade reading program impact evaluation 2018 midline report*. Prepared for USAID/Ghana under the Evaluating Systems program, Contract No. GS-10F-0294V. Washington, DC: USAID.

Social Impact (2019). *Ghana early grade reading program impact evaluation endline report.* Prepared for USAID/Ghana under the Evaluating Systems program, Contract No. GS-10F-0294V. Washington, DC: USAID.

Soine, K. M., & Lumpe, A. (2014). Measuring characteristics of teacher professional development. *Teacher Development, 18*(3), 303–333. https://doi.org/10.1080/13664530.2014.911775

Togneri, Wendy & Anderson, Stephen E. (2003). *Beyond islands of excellence: What districts can do to improve instruction and achievement in all schools.* Washington, DC: The Learning First Alliance and the Association for Supervision and Curriculum Development. https://learningfirst.org/sites/learningfirst/files/assets/biefullreport.pdf

United States Agency for International Development (USAID). (2020). *Kenya: Tusome early grade reading* [Fact sheet]. https://www.usaid.gov/documents/1860/tusome-early-grade-reading-activity

Walls, E., Tulloch, C., & Harris-Van Keuren, C. (2020). *Cost analysis guidance for USAID-funded education activities.* Prepared for Building Evidence in Education (BE2). Washington, DC: USAID. https://www.edu-links.org/sites/default/files/media/file/USAID%20Cost%20Analysis%20Guidance_Final%20Feb20_0.pdf

Wolf, S., Turner, E. L., Jukes, M. C. H., & Dubeck, M. M. (2018). Changing literacy instruction in Kenyan classrooms: Assessing pathways of influence to improved early literacy outcomes in the HALI intervention [Health and Literacy Intervention]. *International Journal of Educational Development*, *62*, 27–34. https://doi.org/10.1016/j.ijedudev.2018.02.004

ANNEXES

Annex A. Organizations Contacted for Learning at Scale Consideration

Africa Development Bank
Abdul Latif Jameel Poverty Action Lab (J-PAL)
Aga Khan Development Network
Agence Française de Développement (French Development Agency [AFD])
Basic Concepts Unlimited
Basic Education Coalition (BEC)
Bill and Melinda Gates Foundation
Boston Consulting Group (BCG)
BRAC
Bridge International Academies
Brookings Institution
Build Africa, Uganda
Cambridge Education
Campaign for Female Education (CAMFED)
Center for Global Development (CGD)
Chemonics International
Creative Associates International
Delivery Associates
Department for International Development (DFID), United Kingdom (merged into new Foreign, Commonwealth, and Development Office [FCDO] in 2020)
DevTech
Department of Foreign Affairs and Trade (DFAT), Australia
DNA Economics
Dubai Cares
EducAid
Educate a Child
Education Development Center, Inc. (EDC)
Education Development Trust
FHI 360
Ghana Education Service
Global Partnership for Education (GPE)
Imagine Worldwide
Interaction

International Rescue Committee (IRC)
JET Education Services
Juarez and Associates
JumpStart, South Africa
Learning and Educational Achievements in Pakistan Schools (LEAPS)
Luminos Fund
Mitchell Group
Pratham
Qatar Foundation
Rising Academy Network
Room to Read
RTI International
Save the Children
School to School International
South Africa Department of Basic Education (DBE)
STiR Education
The Citizens Foundation
Tikichuela (Peru)
United Nations Children's Fund (UNICEF)
United Nations Educational, Scientific, and Cultural Organization (UNESCO)
United States Agency for International Development (USAID)
University of Notre Dame
University of Witwatersrand
World Bank
World Education
World Learning
World Reader
World Vision
Young 1ove

Annex B. Learning at Scale Program Interview Questions

1. Please describe your intervention.
   a. What level of schooling? Primary? Secondary?
   b. Does the program involve a literacy component?
   c. What is the timeline for the program? When will it end?
2. What learning outcome evaluation results exist?
   a. What level(s) are the outcome data available for?
   b. How many schools are the outcome results externally valid to?
   c. What is the impact of the program (as defined by the program)?
   d. What is the impact of the program (as defined by effect sizes in standard deviations)?
   e. When was the impact evaluation done?
   f. What is the design of impact study?
   g. How does the impact study try to show causality?
3. How many schools are you implementing in currently? How many schools will you be implementing in by the end of the program?
4. How many entire districts or clusters is the program implementing in?
5. Is the program integrated into the system? Evidence for that? If your support stopped, would the system maintain the program? How do you know?
6. How is this different from programs that have not been integrated into the system in the same country?
7. What is the instructional strategy in the classroom that makes the difference?
8. If selected, can we have access to the raw impact data?
9. If selected, can we have access to cost data?
10. Can you identify high-performing and low-performing districts/locations?
    a. Could we collect data in these high-performing districts/locations?
    b. Can you help get us access to those locations?
11. What else do we need to know to understand your program and its eligibility for Learning at Scale?

Annex C. Full Data on Program Elements and Key Elements for Program Implementation

Legend for cell shading: L = light green (Component of Program Implementation), D = dark green (Key to Program Implementation). Blank = no shading.

Category	Items	Kenya Tusome — Component of Program Implementation	Kenya Tusome — Key to Program Implementation	Pakistan Reading Program — Component	Pakistan Reading Program — Key	Read India — Component	Read India — Key	EQUIP-T, Tanzania — Component	EQUIP-T, Tanzania — Key	SERI, India — Component	SERI, India — Key	Ghana Learning — Component	Ghana Learning — Key	Lecture Pour Tous, Senegal — Component	Lecture Pour Tous, Senegal — Key	NEI+, Nigeria — Component	NEI+, Nigeria — Key
Materials	Supplementary readers	L		L		L		L		L		L				L	
Materials	Program materials aligned to government curriculum	L	D	L				L	D	L	D	L		L		L	
Materials	Structured teachers' guides (scripted lessons)	L	D	L	D					L	D	L	D	L	D	L	D
Materials	Student books for all students (1:1)	L	D			L				L		L		L		L	
Materials	Materials developed with government	L		L	D			L	D			L	D	L		L	
Materials	Teaching aids (big books, letter cards, pocket chart, etc.)	L		L		L	D	L	D			L		L			
Materials	Word walls, letter charts, etc.			L		L	D					L		L			
Materials	Local-language materials			L										L	D	L	D
Materials	Student books (textbooks)	L								L		L		L		L	
Materials	Consumable student books (workbooks)			L	D							L		L		L	
Materials	Lesson plans (little to no scripting or structure)	L				L	D							L			

Category	Items	Kenya Tusome		Pakistan Reading Program		Read India		EQUIP-T, Tanzania		SERI, India		Ghana Learning		Lecture Pour Tous, Senegal		NEI+, Nigeria	
		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	Component of Program Implementation	Key to Program Implementation	Component of Program Implementation	Key to Program Implementation
Materials	Online or soft-copy materials for teachers			■		■								■			
Materials	Textbook taken home			■										■		■	■
Pedagogy	Phonics-based instruction	■	■	■	■	■							■		■	■	
Pedagogy	Direct instruction (explicit and systematic)	■	■	■						■	■	■		■			■
Pedagogy	Pair work																
Pedagogy	Continuous assessment			■	■	■										■	
Pedagogy	Gradual-release model, "I do, We do, You do"										■	■		■			
Pedagogy	Increased instruc-tional time in lessons	■	■	■	■						■						■
Pedagogy	Mother-tongue program			■													■
Pedagogy	Group work					■	■										
Pedagogy	Bilingual program	■														■	
Pedagogy	Implemented with numeracy program					■	■										
Support	Coaches have structured tools	■	■	■	■	■				■	■	■		■		■	■
Support	Coaches are provided with program/ teacher materials	■		■		■	■				■			■		■	
Support	External-to-school coaching	■	■	■	■	■								■		■	

Category	Items	Kenya Tusome		Pakistan Reading Program		Read India		EQUIP-T, Tanzania		SERI, India		Ghana Learning		Lecture Pour Tous, Senegal		NEI+, Nigeria	
		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	Component of Program Implementation	Key to Program Implementation	Component of Program Implementation	Key to Program Implementation
Support	Coaches are government staff	░		░	█					░	█	░		░		░	
Support	School-based communities of practice meetings			░		░	█	░	█			░	█				
Support	External-to-school communities of practice meetings	░						░						░		░	
Support	Program support coaches in school	░								░			█				
Support	Coaches use tablets or other devices (smart phones, etc.)	░	█											░			█
Support	Internal-to-school coaching							░	█						█		
Support	Communities of practice use structured tools							░									
Support	Coaches reimbursed by program (transport support)	░		░		░											
Support	Coaches meet in groups/with supervisors					░			█								
Support	Virtual communities of practice (WhatsApp, SMS; not face to face)			░										░		░	█
Support	Virtual coaching (WhatsApp, SMS; not face to face)													░		░	

252

Category	Items	Kenya Tusome		Pakistan Reading Program		Read India		EQUIP-T, Tanzania		SERI, India		Ghana Learning		Lecture Pour Tous, Senegal		NEI+, Nigeria	
		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	Component of Program Implementation	Key to Program Implementation	Component of Program Implementation	Key to Program Implementation
Support	Coaches are hired by program																
System Support	Program has regional staff																
System Support	Program uses monitoring data to make decisions about implementation																
System Support	Program invested in capacity building at a decentralized level																
System Support	Program shares achievement data with government decision makers																
System Support	Program designed to align with existing education plans																
System Support	Program responsible for distribution of materials																
System Support	Government uses monitoring data to make decisions about implementation																
System Support	Program monitors frequency of coach visits																
System Support	Government staff responsible for conducting monitoring																

Category	Items	Kenya Tusome		Pakistan Reading Program		Read India		EQUIP-T, Tanzania		SERI, India		Ghana Learning		Lecture Pour Tous, Senegal		NEI+, Nigeria	
		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	Component of Program Implementation	Key to Program Implementation	Component of Program Implementation	Key to Program Implementation
System Support	Program invested in capacity building at a central level	●		●	●●			●				●		●		●	
System Support	Program supports government beyond literacy instruction	●		●				●	●●			●		●		●	
System Support	Program staff embedded in government offices			●				●				●	●●	●	●●	●	
System Support	Program develops and uses dashboard for result/data sharing	●	●●	●		●						●	●●	●		●	●●
System Support	Program mobilizes additional local resources to support schools	●		●				●		●	●●	●		●		●	
System Support	Government responsible for monitoring frequency of coach visits							●		●	●●	●		●			
System Support	Government responsible for distribution of materials	●						●						●			
System Support	Program implements community materials/supports (e.g., mobile library)					●								●			
System Support	Program gives community/local grants	●		●										●			

● = Component of Program Implementation (light shading) ●● = Key to Program Implementation (dark shading)

254

Category	Items	Kenya Tusome		Pakistan Reading Program		Read India		EQUIP-T, Tanzania		SERI, India		Ghana Learning		Lecture Pour Tous, Senegal		NEI+, Nigeria	
		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	Component of Program Implementation	Key to Program Implementation	Component of Program Implementation	Key to Program Implementation
System Support	Program mapping/scoping exercise prior to intervention																
System Support	Program sends funds to government																
Training	Initial face-to-face training																
Training	Refresher face-to-face training																
Training	Teacher training emphasizes modeling/practice																
Training	Nonresidential teacher training																
Training	Teacher training (lowest level of cascade) done by government officers																
Training	Structured training manuals																
Training	Training of trainers done by program staff																
Training	Training for head teachers																
Training	Training of trainers done by government staff																
Training	School-based training																

Category	Items	Kenya Tusome		Pakistan Reading Program		Read India		EQUIP-T, Tanzania		SERI, India		Ghana Learning		Lecture Pour Tous, Senegal		NEI+, Nigeria	
		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	Component of Program Implementation	Key to Program Implementation	Component of Program Implementation	Key to Program Implementation
Training	Pre-service training component	▨		▨										▨		▨	
Training	Residential teacher training	▨				▨		▨									
Training	Virtual teacher training (WhatsApp, SMS; not face to face)													▨		▨	

Annex D. Classroom Observation Instruments Developed for Learning at Scale Research

Classroom Observation Instruments
Timed Lesson Observation; Post-Observation Checklist; Material Inventory
Learning at Scale
Data Collection- 2020

Before the Observation

Greet the teacher. Confirm that the teacher teaches the class you are seeking to observe. During the observation, find a place to sit where you can observe the lesson with as little disruption to the classroom as possible. Keep observation tools in front of you.

Introduction and Consent

"Hello! My name is _____, I'm a researcher from RTI International. I'm here to learn about the _____ program that is working with your Ministry to improve learning outcomes of students. Thank you for taking the time to talk with me today. This research is part of a multi-country study called "Learning at Scale" that is looking at programs that have shown improvement in student learning outcomes to understand what factors contributed to that success. We'd like to observe your reading lesson today as part of this study.

I would like you to feel comfortable teaching this lesson as you do on any day. I will be taking notes while you are teaching so that I can accurately capture the lesson. These notes will remain confidential. Data will be anonymized, and neither your name nor any other identifying information will be stored or reported alongside any classroom observation findings.

This is completely voluntary. You can refuse participation in this observation or end the observation at any time. If you choose to refuse or end this observation at any point, there are no negative repercussions.

Do you agree to the observation? YES/NO

Observation Information

Observer Name: _____

Date of Observation: _____

Country: _____

Program of Study: _____

Observation Start Time: _____

Observation End Time: _____

Location of Observation: _____

Observation ID#:   _____

School: _____

Grade Observed: _____

District/County/Province: _____

Note: This observation will be conducted electronically, on tablets with Tangerine[TM] software, and ID#s will be randomly generated.

Timed Lesson Observation

Date: |___|___|/|___|___|/|___|___| Start: |___|___|:|___|___| End: |___|___|:|___|___|
        date   month   year   hour   minute   hour   minute

Directions: Using the timing device, at each 2-minute interval, the assessor captures a snapshot of what is happening at each level. The assessor notes what is happening at the exact 2-minute interval, not what happened previously. Only one per level can be chosen.

Lesson Minute	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46
Instruction																							
Reading (R)																							
Writing (W)																							
Oral Language (OL)																							
Phonological Awareness (PA)																							
Grammar (G)																							
Assessment (A)																							
Management (M)																							
Nothing (N)																							
Language Part																							
Story (ST)																							
Sentence (SE)																							
Word (W)																							
Word Part (WP)																							
Letter (L)																							
Sound (SO)																							
Other (O)																							
Materials																							
Book (B)																							
Written (W)																							
Oral (O)																							
Illustration (I)																							
Nothing (N)																							
Student Response																							
Reading (R)																							
Listening (L)																							
Writing (W)																							
Orally (O)																							
Physical (P)																							
Manipulate (M)																							
Getting (G)																							
Not Engaged (N)																							
Focus																							
Whole Class (W)																							
Large Group (L)																							
Small Group (S)																							
Individual (I)																							
On Task																							
Most (M)																							
Half (H)																							
Few (F)																							
Climate																							
Positive (P)																							
Negative (N)																							
Ordinary (O)																							

Instruction: This level is to code the overall skill that is the focus of instruction.

Reading (R): This includes skills such as print concepts (e.g., directionality, spacing, word length), letter knowledge (e.g., name, shape, sound), decoding, comprehension, and blending.

Writing (W): This includes instruction in handwriting, copying, spelling, student-generated writing, journals.

Phonological awareness (PA): Identify, compare, generate, manipulate units of sound (e.g., beginning sounds, clapping syllables, identify rhyme)

Oral Language (O): This includes skills such as, songs, rhymes, vocabulary, and word meaning.

Grammar (G): This include knowledge of capitalization, plurals, and punctuation.

Assessment (A): New content is not the focus; the purpose is to measure progress of previous learning.

Management (M): Any activity for classroom management. For example, the teacher is taking attendance, asking the students to get out books. No academic demands are placed on the students.

Nothing (N): Nothing is being taught at this time point.

Language Part: This level is to code the unit of language that is being taught or referenced. Yield to Word Part if instruction is showing how to use blend or segment word. Yield to Sentences, if the task is to connect words to create a sentence.)

Story (ST): Refers to a story, a rhyme, a song or a passage of connected text. It is anything longer than one sentence, where the sentences connect.

Sentence (SE): A single sentence.

Word (W): An isolated word or words.

Word Part (WP): A word part or syllable are being blending or segmented. (e.g., d - o - g; Mon – day; I-me-ka-ti-ka).

Letter (L): Any letter.

Sound (SO): This could be the Sound a letter makes (e.g., letter "Ay" makes the sound "ah") or a combination that make one sound (e.g. ch, th).

Other (O): The focus is on other content besides language (e.g., counting objects; reference the days when there is no reference to print).

Materials: This level is to code the types of materials used during instruction. Code the material that is the subject of discussion – the thing the teacher is referring to or that students are paying attention to. It could be visible– such as writing on the board or in a book, or it could be something spoken about such as a memorized poem

Book (B): The range of books that students may use. For example, pupil book, a supplemental reader, textbook, an electronic book.

Written (W): Any written text that isn't a book for students that is being read or created by the teacher or the students (e.g. letter cards, pocket charts, chalkboard)

Oral (O): Any language that is presented orally. This can be initiated by the teacher or the students.

Illustration (I): An illustration, picture, realia or any object.

Nothing (N): Nothing is being used or referenced.

Student Response: This level is to code what the students are doing during an observation interval. What type of response do they have to the classroom activity? This provides a measure of student activity throughout the class.

Reading (R): The students read something silently or aloud.

Listening (L): Students are listening or waiting to answer a question. This includes raising hands.

Writing (W): Students respond by writing.

Orally (O): Students respond orally (e.g., singing a song, answering a question).

Physical (P): Students have a physical response, (e.g., raising a number of fingers on a hand; writing a word in the air)

Manipulate (M): Students respond by manipulating something with their hands. (e.g., moving blocks to represent sounds)

Getting (G): Students are getting something or waiting for the teacher (e.g., a book, school bag; lined up waiting to have their work checked).

Not Engaged (N): Students are not engaged (e.g., looking out of the window).

Focus: This level is to code the response type used in the activity expected by the teacher. Is an individual reading to the class? Are collaborating in small groups?

Whole (W): Response from whole class.

Large (L): Response from a large group of students (roughly 50%).

Small (S): Response from a small group of students (pairs or small groups).

Individual (I): Response from an individual student.

On Task: This level is to code what the rest of the class is doing as compared to those noted in the Focus level.

Most (M): Approximately 3 out of 4 students are engaged in the current activity or are doing something sanctioned by the teacher.

Half (H): Approximately 2 out of 4 students are engaged in the current activity or are doing something sanctioned by the teacher.

Few (F): Few students are engaged in the current activity or are doing something sanctioned by the teacher.

Climate: This level is to code ways the tone of the classroom climate.

Positive (P): Use of student names; encouragement; including student voices or ideas

Negative (N): Use of negative language with the students; gender or other stereotypes; possible physical punishment

Ordinary (O): The tone of the classroom is neither noticeably positive nor negative.

POST-OBSERVATION BINARY CHECKLIST

Directions: Complete immediately after observing the lesson. Tick yes if you observed the activity during the observation; tick no if you did not observe it. Advanced observers may tick items as they are seen.

	Demonstration: The teacher showed the students what she expected them to do.		
1.	The teacher said that she wants students to do the activity that she does, "First I'll do it, then you'll do it."	yes	no
2.	The teacher explained what she was demonstrating, "Notice how when I read this text my voice changes for the character who is sad."	yes	no
3.	While demonstrating, the teacher "thought aloud" to show her thinking about decoding or meaning.	yes	no
4.	The teacher demonstrated how to use a resource. (word wall, alphabet strip, page number)	yes	no
	Research Based/Simple View: Instructional activities that advance code and meaning skills.		
5.	Playing with language. (alliteration, clapping syllables, onomatopoeia, rhyme, reduplication)	yes	no
6.	Print awareness activities. (discussing book title/cover, tracking text, discussing letter shapes, counting letters)	yes	no
7.	Instruction of letter sounds and their symbol.	yes	no
8.	Word strategies for decoding. (blending, chunking, look at letter, write for sound)	yes	no
9.	Attention to accuracy, rate or expression while the teacher or students read connected text.	yes	no
10.	Activities to learn the meaning of words.	yes	no
11.	Activities or discussions about the meaning of connected text. (e.g., teacher read alouds or student read text)	yes	no
	Application: Students practice skills.		
12.	Students practiced skills without the teacher that were demonstrated earlier by the teacher.	yes	no
13.	Students read letters, individual words or connected text on their own or with another student.	yes	no
14.	Students discussed the text read. (give an opinion, evaluate a character)	yes	no
15.	Students wrote or drew (letters, words, or sentences) without copying.	yes	no
	Responsiveness: The teacher adapts to student behaviours.		
16.	The teacher used a visual method to check understanding. (thumbs up, writing answers on slate, point in book)	yes	no
17.	The teacher asked questions to check for individual understanding.	yes	no
18.	The teacher helped students to achieve the correct answer for an initial incorrect response. (hints, scaffolds)	yes	no
19.	The teacher circulated, leaned in, redirected, to monitor student progress.	yes	no
20.	The teacher rephrased explanations when students did not understand.	yes	no
21.	The teacher included students who did not volunteer to answer.	yes	no
	Resources: Materials that are needed for implementation are used.		
22.	Students read from something that is dedicated for their individual use. (book, exercise book, slate, tablet)	yes	no
23.	Instruction used a fixed item for all to reference. (alphabet chart, chalkboard, graphic organizer, posters)	yes	no
24.	Instruction included an item designed for interaction between teacher and students or between students. (read aloud, pocket chart or similar, phonics wheels, storyboards)	yes	no
	Efficiency: Instructional time is maximized.		
25.	The teacher had materials ready at the start of an activity or the class.	yes	no
26.	The pace of the instruction is lively.	yes	no
	Student-Directed		
27.	The content was student directed.	yes	no

MATERIAL INVENTORY

Directions: Complete immediately after filling in the binary checklist. Tick yes if teacher and student materials are available IN THE CLASSROOM; tick no if materials are not available in the classroom.

	Teacher Materials			
	Teacher has [program] teacher materials IN THE CLASSROOM			
1a	Teachers' guide	yes	no	n/a
1b	Teaching aids (e.g. pocket chart, letter cards)	yes	no	n/a
	Student Materials			
	Student [program] materials are available IN THE CLASSROOM			
2a	Textbooks	yes	no	n/a
2b	Supplementary readers	yes	no	n/a
2c	Student workbooks / exercise books	yes	No	n/a
	Student Textbook Ratio			
	Student textbook ratio IN THE CLASSROOM			
3a	How many students are in the classroom? [Count students in the class]			
3b	How many students have textbooks in the classroom? [Count]			
	Library Materials			
	Library materials in the school and classroom			
4a	Is there a school library?	yes	no	n/a
4b	Is there a classroom library?	yes	no	n/a
4c	How many students in the class have any books from the school/classroom library in their desk/bag? [Ask students to show the books]			

The following table provides explanations to guide training and to support increased consistency across uses.

	Demonstration: The teacher showed the students what she expected them to do.
1.	The teacher briefly said that she wants them to do the activity that she does, Training: The teacher might say something like, First I'll do it, then you'll do it. She could also gesture what she wants them to do.
2.	The teacher explained what she was demonstrating, Training: The teacher might say, "Notice how when I read this text my voice changes for the character who is sad."
3.	While demonstrating, the teacher "thought aloud" to show her thinking about decoding or meaning. Training: The teacher explicitly shares her thought process or strategy used. Examples include: • "That word did not make sense, so I will reread it and make sure I look at each letter." • "I don't know the meaning of this word, but I know that since it ends in ly it is probably a describing word."
4.	The teacher explicitly demonstrated or explained how to use a resource. Training: Resources might include word wall, alphabet strip, page number
	Research Based/Simple View: Instructional activities that advance code and meaning skills.
5.	Manipulating language. Training: Features of the language structure and sound are highlighted, compared, manipulated. Meaning is not the focus. Print may or may not be present. It could be during a song, during a read aloud or verbal interaction. Examples: • Alliteration: beginning sounds (boy, bench) • Rhyme: paka – taka; plane – train • Clapping for syllables • Reduplication: word part or alternate form repeated (baya baya; ping pong) • Onomatopoeia: words that sound like they look (boom, click clack, zap)
6.	Print awareness activities. (discussing book title/cover, tracking text, discussing letter shapes, counting letters) Training: Direct attention to print features.
7.	Instruction of letter sounds and their symbol. Training: Print is involved. Connects an oral sound with its letter.
8.	Word strategies for decoding. (blending, chunking, look at letter, write for sound) Training: The implied intention is to provide explicit ways to read words from its parts. Examples: • Blending: d – o – g • Look at a letter in a word: Reminder to look at a specific letter • Chunking: yes – ter – day • Write for sound: Write the sounds they hear. Example lfnt for elephant
9.	Attention to accuracy, rate or expression while the teacher or students read connected text (e.g., poem, sentence, story) Training: To select this, the teacher makes comments about accuracy, rate or expression. Connected text examples a sentence, poem, passage. • Accuracy: Attention to ensuring the words are read correctly. • Rate: Attention to noting the time taken to read the text. • Expression: Attention to matching how it was read to the meaning of the text. Example "He opened the door slowly", reading the word slowly, stretched out.
10.	Activities to learn the meaning of words. Training: This can happen before, during or after engaging with text. Examples: demonstrate word meaning with gestures, objects, images or examples in context.
11.	Activities or discussions about the meaning of connected text (teacher read alouds or student read text). Training: This can happen before, during or after engaging with text. Examples include: • Before reading: Background knowledge about the topic is discussed; making a prediction. • During reading: Clarification of words, events or conversation about text. • After reading: Comprehension activities (questions, summaries, role playing, written response).
	Application: Students practice skills.

12.	Students practiced skills without the teacher that were demonstrated earlier by the teacher.
	Training: This could be alone, with a few other students or more.
13.	Students read letters, individual words or connected text on their own or with another student.
	Training: This could be alone, with a few other students or more.
14.	Students discussed the text read. (give an opinion, evaluate a character)
	Training: This could be in an activity with classmates such as "turn and talk". The student has a chance to apply their ideas.
15.	Students wrote or drew (letters, words, or sentences) without copying.
	Training: The writing could be a student idea or the teacher could be dictating.

	Responsiveness: The teacher adapts to student behaviours.
16.	The teacher used a visual method to check understanding. (thumbs up, writing answers on slate, point in book)
	Training: This could be once or ongoing throughout the lesson.
17.	The teacher asked questions to check for individual understanding.
	Training: The teacher asks different questions to a few students about something that was taught.
18.	The teacher helped students to achieve the correct answer after an initial incorrect response. (hints, scaffolds)
	Training: The teacher supports the students. She gives enough support so they can get the correct answer.
19.	The teacher circulated, leaned in, redirected, to monitor student progress.
	Training: Most likely that the students are doing independent work.
20.	The teacher rephrased explanations when students did not understand.
	Training: Examples: Connect the to the student lives. Switch to a local language after trying in language of instruction.
21.	The teacher included students who did not volunteer to answer.
	Training: To capture who is included beyond those eager to answer.

	Resources: Materials that are needed for implementation are used.
22.	Students read from a book, tablet, small chalkboard, or exercise book that is dedicated for their individual use.
	Training: The intention is that students have their eyes on text to read to create.
23.	Instruction used a fixed item for all to reference (alphabet chart, chalkboard, graphic organizer, posters).
	Training: The teacher is explicit in her use of the resource that is always on the wall. An example might be saying, "Let's look at images on the poster to find something that starts with the sound of /s/."
24.	Instruction included an item designed for interaction between teacher and students or between students. (read aloud, pocket chart or similar, phonics wheels, storyboards)
	Training: This is an item that supports the lesson. Examples:
	• Read aloud: The teacher shows a picture and asks students to name something that begins with sound /k/.
	• Pocket chart: Building words with letter cards or creating sentences with word cards.
	• Phonic wheels: Provide a chance to read words that change by part (b – ack, bl -ack, s – ack, t- ack, tr – ack)
	• Storyboards: With cut out pictures or objects to explain the events of the stories.

	Efficiency: Instructional time is maximized.
25.	The teacher had materials ready at the start of an activity or the class.
	Training: Students not waiting for long periods while she organizes materials or writing on the board.
26.	The pace of the instruction is lively.
	Training: The tempo of the activities gives enough time for multiple students to engage. The activities do not drag.

	Student-Directed
27.	The content was student directed.
	Training: The intention is to observe when content is directed from student ideas. Example:
	• Students working in small groups on content they decided.
	• Students select from a choice of items of what to learn next.

Observer Statement:

After completing the checklist, observers respond to four prompts. The intention is to provide a short description of the lesson and to compensate for deficiencies in the classroom binary checklist.

Prompts
1. What did the teacher teach?
2. How were teaching and learning materials used by the teacher or students?
3. How did the students react/respond during the lesson? (instructional or behaviour)
4. What is your opinion of this lesson?

*Example Observer Statement:

The teacher spent most of the time saying the names of letters and their sounds. Later, the students named objects in their books. She wrote on the chalkboard and showed letter cards. The students repeated what she said but many looked bored. The teacher's explanation of the letters and sounds was clear but there was too much of the same instruction.

To train the observers to write these statements, we will show them the categories in the column

Observer Statement	What the statement provides
The teacher spent most of the time saying the names of letters and their sounds. Later, the students orally named of objects in their books.	What the teacher taught
She wrote on the chalkboard and showed letter cards.	Materials
The students repeated what she said but many looked bored.	Student response
The teacher's explanation was clear but there was too much of the same instruction.	Opinion

Annex E. Program Data and Reports

Cambridge Education
Cambridge Education. (2020, June). *Learning at Scale cost capture template.* Learning at Scale.
Mbiti, I., & Schipper, Y. (2020). *Teacher and parental perceptions of performance pay in education: Evidence from Tanzania.* Research on Improving Systems of Education (RISE) Working Paper 20/037.
Rawle, G., Binci, M., Pettersson Gelander, G., Harb, J., Jasper, J., Khan, S., . . . Ruddle, N. (2019). *EQUIP-Tanzania impact evaluation—Endline quantitative technical report, volume I.* Washington, DC: USAID.
Rawle, G., Binci, M., Pettersson Gelander, G., Harb, J., Jasper, P., Khan, S., . . . Ruddle, N. (2019). *EQUIP-Tanzania impact evaluation—Endline quantitative technical report, volume II.* Oxford Policy Management.

Chemonics International
Chemonics International. (2020, October). *Data quality checklist.* Learning at Scale.
Chemonics International. (2020, July). *Learning at Scale cost capture template.* Learning at Scale.
Mount-Cors, M., Rousseau, M., & de Galbert, P. (2019). *Lecture Pour Tous/All Children Reading Senegal Early Grade Reading Assessment (EGRA) midline report-Second draft.* Washington, DC: USAID.

Creative Associates
Creative Associates. (2020, November). *Data quality checklist.* Learning at Scale.
Creative International. (2020, October). *Learning at Scale cost capture template V2.* Learning at Scale.
Creative International. (2020, June). *Learning at Scale cost capture template.* Learning at Scale.
Creative International. (2018). *Northern Education Initiative Plus Early Grade Reading Assessment midline report.* Washington, DC: USAID.

FHI 360
Darko Osei, R., Adobea Owusu, G., Asem, F., & Afutu-Kotey, R. (2009). *Effects of capitation grant on education outcomes in Ghana.* Global Development Network 1999–2009.
FHI 360. (2020, October). *Data quality checklist.* Learning at Scale.
FHI 360. (2020, September). *Learning at Scale cost capture template.* Learning at Scale.
FHI 360. (2019). *Ghana Early Grade Reading Program impact evaluation endline report.* Washington, DC: USAID.

International Rescue Committee (IRC)
IRC. (n.d.). *Cost analysis approach: Cohort 3 cost-effective analysis (PRP).*
IRC. (2020, June). *C3 [Cohort 3] analysis results for RTI.*
IRC. (2020, March). *Cohort 1 & 2 cost-effective analysis—Analysis workbook for RTI.*
IRC. (2020, March). *Export-PRP materials unit costs.*
IRC. (2017, July). *Cost of the PRP model.*
IRC. (2017). *Early Grade Reading Assessment Urdu endline study.* Washington, DC: USAID.

Pratham
ASER. (2018). *Annual Status of Education Report 2012-2018.* New Delhi, India: ASER.
Pratham. (2020, September). *Learning at Scale cost capture template.* Learning at Scale.
Pratham. (2020). *Odu Karnataka annual report, 2019-2020.* New Delhi: Pratham.

Room to Read

Joddar, P. (2019). *Impact evaluation of Room to Read's literacy program under partnership approach in Madhya Pradesh and Uttar Pradesh, India.* Washington, DC: USAID.

Joddar, P. (2018). *Impact evaluation of the literacy program-partnership approach under Scaling-up Early Reading Intervention (SERI) funded by USAID.* Washington, DC: USAID.

Room to Read. (2020, November). *Data quality checklist.* Learning at Scale.

Room to Read. (2020, May). *Learning at Scale cost capture template.* Learning at Scale.

RTI International

Freudenberger, E., & Davis, J. (2017). *Tusome external evaluation—Midline report.* Washington, DC: USAID.

RTI International. (2020, November). *Data quality checklist.* Learning at Scale.

RTI International. (2020, August). *Learning at Scale cost capture template.* Learning at Scale.