Cultivating Dynamic Educators: Case Studies in Teacher Behavior Change in Africa and Asia

Edited by
Sarah Pouezevara

RTI Press

RTI INTERNATIONAL

Library of Congress Control Number: 2018959618

ISBN 978-1-934831-22-9
(refers to print version)

RTI Press publication No. BK-0022-1809
https://doi.org/10.3768/rtipress.2018.bk.0022.1809
www.rti.org/rtipress

Cover design: Dayle Johnson

The RTI Press mission is to disseminate information about RTI research, analytic tools, and technical expertise to a national and international audience. RTI Press publications are peer-reviewed by at least two independent substantive experts and one or more Press editors.

RTI International is an independent, nonprofit research institute dedicated to improving the human condition. We combine scientific rigor and technical expertise in social and laboratory sciences, engineering, and international development to deliver solutions to the critical needs of clients worldwide.

This publication is part of the RTI Press Book series.

RTI International
3040 East Cornwallis Road, PO Box 12194
Research Triangle Park, NC 27709-2194, USA
rtipress@rti.org
www.rti.org

CHAPTER 5

Short Message Service (SMS)–Based Remote Support and Teacher Retention of Training Gains in Malawi

Timothy S. Slade, Scott Kipp, Stirling Cummings, and Kondwani Nyirongo

Introduction

This chapter critically reviews the design, implementation, and evaluation of an attempt to study an exploratory short message service (SMS)–based intervention conducted under the auspices of the United States Agency for International Development's (USAID's) Malawi Early Grade Reading Activity (EGRA).[1] The overall EGRA program, which was implemented from July 2013 to October 2016 in 1,614 schools across 11 educational districts, was designed to support Malawi's Ministry of Education, Science, and Technology (MOEST) to improve reading outcomes in both Chichewa and English languages among children in grades 1–3. The program addressed this goal through the development, production, and distribution of high-quality teaching and learning materials; regularly occurring training of teachers in phonics-based instructional methodologies; support for instructional coaching provided by MOEST's primary education advisors (PEAs); community-mobilization efforts focused on developing improved out-of-school literacy environments; and engagement of MOEST leadership to shape the educational policy environment in a way that would favor improved literacy instruction.

[1] In the interest of space and in accordance with the activity's branding requirements, the authors of this chapter have used the abbreviation *EGRA* to indicate the USAID activity named in this chapter. We recognize that the same abbreviation is in widespread use for *Early Grade Reading Assessment*, which is an oral assessment of foundational literacy skills in developing countries. To avoid confusion, we have spelled out the name of the assessment when it occurs in the chapter.

Early Reading Outcomes in Malawi

Early Grade Reading Assessments conducted during the EGRA's predecessor contract, the USAID Malawi Teacher Professional Development Support program, painted a dismal picture of reading and education in Malawi. In November 2010, a nationally representative sample of 996 students in grades 2 and 4 found that 72.8 percent of grade 2 children and 41.9 percent of grade 4 children could not read a single word on an oral reading fluency subtask; the proportions of zero scores on the associated reading comprehension subtask were 97.1 percent in grade 2 and 69.3 percent in grade 4 (RTI International, 2010). As of 2011, the official pupil–teacher ratio in Malawi was 76:1 (United Nations Educational, Scientific and Cultural Organization [UNESCO], 2014), although there is reason to believe the actual figure was substantially higher.[2]

Recent pre-service teacher-training curricula in Malawi have paid scant attention to reading pedagogy. In a 2015 study commissioned by the German aid agency Deutsche Gesellschaft für Internationale Zusammenarbeit, Jackson, Jere, Togher, and Webb noted, "there is no clear and progressive programme for teaching of reading," and further indicated that, from a cross-cutting perspective, there is "no evidence of progression in pedagogical practice" and "inadequate teaching methodology" to support the practicum experience teacher trainees are expected to undertake in the second and final year of the program (2015, pp. 10–11).

Many graduates of Malawi's pre-service Initial Primary Teacher Education program have thus been deployed to schools despite being ill-equipped to teach the subjects that are explicitly covered by the curriculum, let alone the complex process of learning to read, which is not covered by the curriculum. The EGRA implementation team recognized that, without some ongoing external support for more effective lesson delivery, merely introducing termly teacher training on reading instruction would be unlikely to cause teachers to alter their own pedagogical practices enough to drastically change reading outcomes.

[2] Malawi's national media outlets decried the "ghost teacher" phenomenon (see, for example, Chauwa, 2015), which artificially inflated the number of teachers serving as the denominator in the pupil–teacher ratio. Anecdotally, one of the authors observed lessons in dozens of classrooms between July 2013 and July 2016 and never saw fewer than 90 children being taught by a single teacher—this despite pupil absenteeism averaging over 30 percent (relative to official enrollment) over the entire period of the EGRA's implementation.

Addressing Professional Development Through Coaching

Altering teachers' standing pedagogical practice by providing routine, qualified support through an instructional coach has been shown elsewhere to be by far the more efficacious, though arduous, route for effecting such change (Fisher, Frey, & Nelson, 2012; Kraft, Blazar, & Hogan, 2017). The EGRA program team elected to provide that support by underwriting the cost of PEAs' visits to schools to observe teachers' lesson delivery and provide feedback. Indeed, analyses of EGRA coaching data found that, among teachers who received multiple coaching visits, the teachers' performance improved with each subsequent visit from a coach (Pouezevara, Pflepsen, Nordstrum, King, & Gove, 2016). The metric for "performance" in this case consisted of how often a teacher exhibited behaviors and practices that were known to support reading acquisition, such as actively engaging individual students in oral reading practice activities and using varied modeling techniques for phonics instruction.

Unfortunately, separate analyses of the EGRA's monitoring and evaluation data indicated that coaching frequency (visits to schools by PEAs) was substantially below both the EGRA program's and MOEST's expectations (Slade, 2016; MOEST, 2015). Furthermore, visitation patterns suggested that PEAs' decisions regarding which schools and teachers to support were driven less by strategic concerns—such as the quality of teaching or student reading performance in a given school, the duration since the last visit to the school, or the arrival of newly transferred teachers—and more by pragmatic ones—such as proximity to the Teacher Development Center (the PEAs' office), ease of access from a paved road, etc. (See Table 5-1 and Figure 5-1, reproduced from Slade, 2016.)

Similar issues have been documented elsewhere. Piper & Mugenda (2012) found that instructional coaching levels in Kenya were low (albeit higher than in Malawi), with more than 50 percent of teachers indicating they had been observed only once per term. Piper & Zuilkowski (2015) identified a negative relationship between the number of teachers in a zone and the number of coaching visits received per teacher. RTI's qualitative analysis of outcomes in Liberia under the EGRA Plus: Liberia program (Piper & Korda, 2011) found strong relationships between learning outcomes and the distance traveled by the coach to reach the school.[3]

[3] Piper and Korda (2011) noted that it was less likely that a marginal coaching visit was what drove the stark differences in learning outcomes, and more likely that the distance traveled was related to other latent factors that may have influenced school-level outcomes.

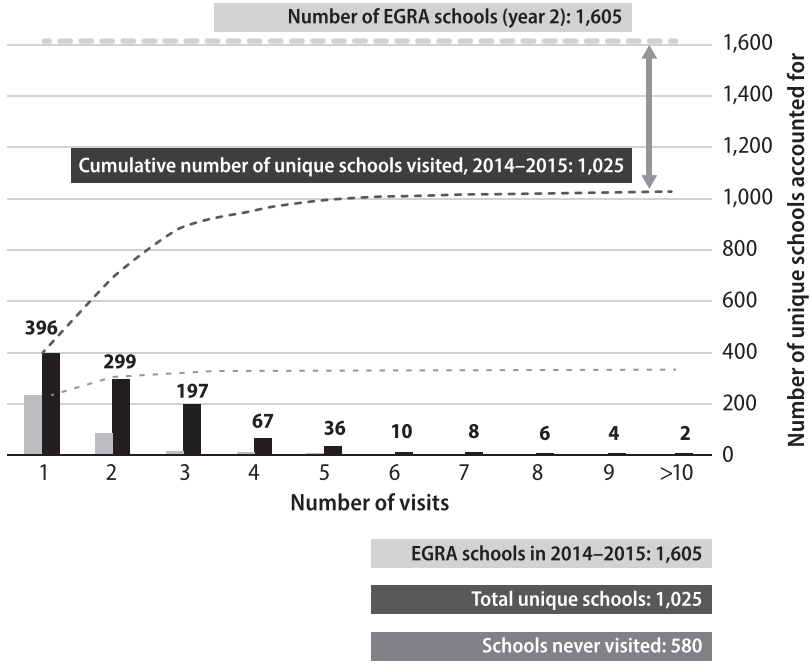
Table 5-1. Unique intervention schools receiving coaching in years 1 and 2 of the EGRA

Intervention school year	2013–2014	2014–2015
No. of schools in intervention	1,188	1,605
No. of unique schools coached (%)	330 (27.8%)	1,025 (63.9%)
No. of schools not coached (%)	858 (72.2%)	580 (36.1%)

EGRA = Early Grade Reading Activity; no. = number.

Source: Slade (2016). Used by permission.

Figure 5-1. Distribution of coaching visits by EGRA PEAs, 2013–2015



EGRA = Early Grade Reading Activity; no. = number; PEA = primary education advisor.
Source: Slade (2016). Used by permission.

Incentive structures that RTI has developed under the reading interventions it has implemented on USAID's behalf in Liberia, Malawi, and Kenya to encourage regular coaching visitation have had to balance simplicity (the coaches must be able to easily understand and respond to the incentives), power (the incentives must be strong enough to overcome coaches' tendencies to focus on other administrative or income-generating tasks), and

affordability (projects must be able to afford the investment throughout implementation, and preferably not create unrealistic expectations once a project has closed). It is not an easy task to create an incentive structure that appropriately and successfully balances those competing priorities, let alone one that facilitates and encourages strategic decision-making as well.

Given the relatively high financial and opportunity costs of providing in-person coaching, instructional coaching in any educational intervention would ideally be strategically distributed, for example, by increasing coaching intensity to teachers who are underperforming while reducing coaching to teachers who are already performing well. An example of such an approach might be to prioritize an initial coaching visit each month (or term) to all schools within a PEA's zone, and then to prioritize the weaker-performing teachers and schools for additional visits within that month (or term). For the EGRA, the PEAs from three districts were trained in the last year of the activity on precisely the approach of targeting weaker schools and teachers in planning their coaching visits.

Introducing SMS as a Component of Professional Development

Understanding that severely uneven coaching across schools was the norm, and in search of cost-efficient means for providing more regular outreach and support between coaching visits, the EGRA team designed and conducted an SMS coaching intervention with a diverse set of primary teachers, as a means to sustain and extend pedagogical and motivational support for educators following face-to-face training.

At the same time, the EGRA team was looking to assess the professional development support it was providing to teachers by way of zone-level trainings. A questionnaire designed to gauge teachers' familiarity with and attitudes toward the pedagogical practices and beliefs promoted by the MOEST and EGRA activities was administered before and after professional development workshops and again after the SMS campaign. This method, described later in this chapter, was intended to allow the EGRA team to test the following hypothesis: An SMS campaign for teachers could help extend and sustain the support provided by in-person trainings and coaching visits.

Unfortunately, due to several critical flaws in the study's design, instrumentation, and implementation, the hypothesis could be neither confirmed nor refuted. Therefore, rather than focusing primarily on reporting the study's findings with respect to the intended research question,

this chapter presents a critical review of the study process, including costs. The intent of this process review is to enable other implementers of similar programs to learn from its shortcomings and more effectively embed research on an SMS intervention into an existing program design.

Design of the SMS Intervention

The SMS campaign was an addition to the EGRA's standard three-level teacher-training cascade, which was delivered on a termly (thrice-yearly) basis. At the initial stage of the cascade, project staff (including the literacy specialist; teaching and learning materials coordinator; disability, gender, and vulnerable populations specialist; and national training coordinator) led a group of master trainers drawn primarily from local teacher-training colleges, universities, and the MOEST headquarters. At the intermediate stage, the master trainers would deploy to several locations around the country to train the PEAs and highly skilled "key teachers" assigned to assist the PEAs in delivering the final level of the training cascade.

At the final level, the PEAs would pair up with key teachers and deliver trainings to the teachers from their zone. Training content and duration were linked to the broader curriculum rollout strategy: longer, more intensive three- to five-day initial training sessions during the first term and year of a given subject's materials being present in a given grade and shorter, one- to three-day refresher trainings in subsequent terms and years. Moreover, the training content was informed by the findings of a weeklong period of intensive classroom observation conducted jointly by EGRA program staff and MOEST headquarters personnel.

Delivery and assessment of the SMS campaign were designed around the cascade. To evaluate its effectiveness using a comparison group, it was introduced only to a portion of the teachers. Program staff developed an assessment and administered it to all master trainers, all PEAs and key teachers, and a sample of teachers at the outset and conclusion of their respective training sessions, as described in the Methodology section.

The SMS content was developed by the authors of this chapter in collaboration with the program's monitoring and evaluation specialist, who assisted with the messages' translation into Chichewa. The 49 messages were grouped into eight themes, as indicated in Table 5-2.

The bulk of the messages—34 of the 49, or roughly 69 percent—were grouped within three themes: encouraging specific practices, reminders about student behavior, and encouragement and motivation. The messages

Table 5-2. Message themes, counts, and examples

No.	Theme	No. of messages	Examples of messages
1	Introduction/ welcome	2	Hello {Teacher_Name}! You have been selected to receive messages from the EGRA team.
			We appreciate your dedication to your learners, your school and your community.
2	Feedback for students	4	Be positive and remind learners to do their best. Even if the learner spells part of the word correctly, that is better than no correct part at all.
			Find a learner you have not heard read aloud in a long time today and check on them. If they need help, practice with I do / You do.*
			Have some of your learners made great improvements this year? If so, be sure to tell them today.
3	Encouraging specific practices	10	Remember, you can explain vocabulary using a picture, gestures, a song, or simple familiar words.
			The WORD EXPLORATION PROCESS has 3 steps: read the word, use the word in a sentence, find another word that means the same or the opposite.
			Any time you read words or sentences from the board, be sure to model how to use finger-pointing to read the sentence using letters as clues.
4	Reminders about student behavior	10	Every learner should be reading from their own pupil book during reading lessons.
			Remember: It is ok if your learners do not always face the front of the class. Some activities will be easier if they can work together reading in groups.
			It is ok for your learners to make mistakes. It is part of learning. Praise them for trying; encourage them to try on their own and listen to themselves.
			Even if your learners say no one at home can read, still encourage them to take home books. They may still be encouraged and might find someone who can help.
5	Lesson plans/ curricular reminders	4	Today, please make sure you use I do, We do, You do on each activity to help learners be successful with the new skill and build learners' confidence.
			Good morning! Please try to review your reading lesson plan carefully for 15 minutes today before beginning. Your learners will benefit greatly.
			Plans and preparation are very important. It will make your job easier and it will help your learners. Please check on your notes and plans for today's lesson.

(*Continued*)

Table 5-2. Message themes, counts, and examples (*Continued*)

No.	Theme	No. of messages	Examples of messages
6	Reflection reminders	4	Think about today's reading lesson. Focus on moments when your learners paid keen attention. Try to make this moment happen again tomorrow.
			Was anything difficult for you in today's reading lesson? If so, ask your head teacher about it—Asking for help is how we all learn!
			After your reading lesson today, look over the lesson plan for 15 minutes and ask yourself if you followed the plan. If you didn't, please try to tomorrow.
7	Encouragement and motivation	14	If one of your fellow teachers is a strong example for their peers in embodying the spirit of EGRA, let them know today that you are thankful.
			Remember: Your learners are individuals, and each one looks up to you for guidance. They may not thank you today, so let us say THANK YOU!
			Teacher, you are a hero making the world a better place, one child at a time.
8	End of term/ campaign	1	The EGRA team hopes you have enjoyed these messages and found them useful. Thank you for your hard work this year!

EGRA = Early Grade Reading Activity; no. = number.

* I do/We do/You do: The Malawi EGRA used a gradual-release, direct instructional technique that involved teacher modeling ("I do"), full-class practice ("We do"), and then individual student practice ("You do").

within the theme of lesson plans/curricular reminders were not specifically linked to the units and lessons the teachers were delivering on a given date, due largely to the complexity of (1) establishing with high reliability the specific lesson a teacher would be teaching on a given day and (2) efficiently routing the specific message appropriate to that lesson to the correct teachers. Future iterations of this kind of campaign and accompanying research should consider testing whether messages tightly linked to the teacher's lesson for the day would have a different impact.

By design, each teacher was to receive 49 SMS messages (an average of roughly seven messages per week) over a seven-week span following their trainings. The substantial majority of messages were sent on weekdays at 7 a.m.; in eight cases, an additional message (the second for the day) was sent at 10 a.m., and in five cases, a third message was sent at 3 p.m. (on the date

of the campaign launch, three explanatory messages were sent out at 10 a.m., 12 p.m., and 3 p.m.). In total, the intervention delivered 20,689 messages to 485 teachers (an error in configuring the messaging list led to the first 14 messages, sent on May 25–June 8, reaching only 196 of the 485 SMS treatment teachers). Of these 485 SMS-receiving teachers, the project's monitoring and evaluation staff were able to follow up with 407 to assess the effects of the messaging campaign. To better explain the background and intentions of the SMS campaign, the following section describes only the literature relevant to the use of various technologies for behavior change and professional development.

Literature Review

Technologies of various types have long been investigated as a means of delivering or enhancing training focused on changing behavior. In some cases, a chosen technology is intent on replacing an in-person training (examples include distance education and massive open online courses); whereas in other cases, a technology is used during or after an in-person event as a supplemental tool. As astutely noted by Gaible and Burns (2005), however, the use of information and communication technology for such efficiencies or enhancements should never be conceived as a *de facto* improvement for an educational program designed to improve instructional quality; many technologies, whether by poor design or by way of a steep learning curve, can cost teachers and other users precious time and effort for little in return.

As the global imperative to improve instructional quality continues to increase, so do the capabilities and prevalence of powerful technologies. Many low-cost technologies, most notably mobile phones, have become particularly widespread and popular over the past 20 years in developing-country contexts. For this reason, educational practitioners continue to explore whether such technologies can be effective in situations in which traditional training methods have proven insufficient or infrequent.

What Gaible and Burns (2005) provided in their review of technologies used for teacher professional development was a historical review of such cases, clearly indicating a powerful predictor of success for a given technological channel or tool: familiarity. They found that by virtue of context (that is, low-income developing countries), a tool or communication channel that was widely familiar to teachers (or, more broadly, to the adult

population of that context) was very likely to have a low cost as well, given the scarcity of disposable income in the population.

The rapid ascent of mobile phones and SMS throughout sub-Saharan Africa provided the backdrop of user familiarity for the Malawi EGRA's messaging campaign, and several prior studies provided further encouragement for the design undertaken. Pouezevara (2015) presented an overview of many of the ways in which mobile devices were being used in low-income contexts for networking and learning, and it was the concept of spaced repetition (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006) that was intended to be tested by this study.

At the core of the activity's intentions with the SMS campaign were two complementary goals: remind and reinforce. That is, the content of the SMS campaign itself was neither designed nor intended to *replace* any portion of in-person training activities but merely served as a timely, regular extension of training efforts and as a means to reinforce concepts and remind recipients to integrate them into their practice. While the authors did not conceive of the SMS messages as a *substitute for* face-to-face coaching, the messages were seen as a potentially effective *supplement to* face-to-face coaching that teachers could benefit from between visits from their designated coach.

In the health sector, interventions for medicinal adherence often share similar goals. As da Costa et al. (2012) demonstrated, the use of targeted SMS campaigns as a post-clinical outreach mechanism led to dramatic increases in patients' consistency in adherence to medication schedules and to increases in self-reported quality of care (e.g., reduced feelings of isolation). Fjeldsoe, Marshall, and Miller (2009) found similar results in a review of 14 studies evaluating the ability of SMS messages to positively affect behavior change with respect to medicinal adherence and clinical visitation schedules.

Aker, Ksoll, and Lybbert (2010) as well as Beltramo and Levine (2012) similarly demonstrated (in Niger and Senegal, respectively) that the familiarity of both the mobile phone and SMS messaging specifically could make them an effective modality for improving learning outcomes in adults participating in education programs. Researchers at Stanford University found positive effects on engagement and student learning when parents received text-message guidance for supporting literacy development (York & Loeb, 2014). The researchers described the

behavior-change aspect of the program as targeting "behavioral barriers to good parenting by breaking down the complexity of parenting into bite-sized pieces and providing continuous encouragement and support over long periods of time" (p. 31). These behavioral barriers are analogous to those that teachers experience when they are trying to learn new instructional methods.

More recently and more directly relevant for early grade teachers in Malawi, Jukes et al. (2016) identified SMS outreach as a key programmatic input of the Health and Literacy Intervention (HALI) project in Kenya. For the HALI project and similarly under the Primary Math and Reading (PRIMR) initiative in Kenya, the incremental value of SMS was not tested in an isolated control but as one component of a wider intervention (Jukes et al., 2016; RTI International, 2014). In Papua New Guinea and India, the SMS Story program provided daily teaching content to teachers via simple SMS, resulting in positive learning gains on early reading measures (Kaleebu et al., 2013; Pratham Education Foundation & Voluntary Service Overseas, 2015).

Ramos and Trinoña (2009), in their analysis of Project MIND (Mobile Technology Initiatives for Nonformal Distance Education), did present controlled evidence for the relative effectiveness of including SMS messaging for increasing the learning gains of an educational program—not, however, as specifically concerned with teachers.

The Malawi EGRA efforts to use SMS messages after in-person trainings and between in-person coaching visits with teachers were intent on reinforcing the in-person training and support. The SMS approach thereby served as a tertiary bridging tool; it was meant to strengthen other training inputs and, holding other activities constant, test the added value of a targeted SMS support program to help translate training into teacher behavior change in the classroom.

Methodology

Measurement Instruments

The EGRA program team developed and deployed an assessment tool dubbed the Learning Gains instrument, with the goal of measuring the effectiveness of professional development activities. The instrument was divided into three sections (see Table 5-3), each focusing on one of three

Table 5-3. Features of the Learning Gains instrument

Section	Fixed-response items	Free-response items
1: The lesson cycle	5	3
2: Classroom management in a context of large classes	7	1
3: Supplementary readers	6	1

salient aspects of the EGRA intervention and the Malawian educational context: teaching according to the lesson cycle, classroom management in a context of large classes, and utilization of supplementary readers. The lesson cycle section included one item requiring respondents to place the three stages of the EGRA lesson cycle ("I do," "We do," and "You do") in order and four multiple-choice questions. The classroom management and supplementary readers sections included pairs of contrasting statements; respondents were required to indicate the statement with which they most agreed. Each section also included free-response items (not included in this chapter's analysis).

The Learning Gains instruments used during the pre-training (P_0) and post-training (P_1) rounds of assessment were identical. The version of the instrument used during the follow-up (P_2) assessment included three additional yes/no questions to record whether the respondent had received SMS messages directly from the EGRA team; if so, whether they had shared them with peers; and, for those who had not received an SMS message directly, whether their peers had shown them the SMS messages to some extent. During each round of assessment, respondents completed the instrument independently of their peers and were not allowed to share their responses with other respondents.

The Learning Gains measurement instrument was administered both before and after the training workshops at the second level (training of trainers) and third level (zonal training workshops) of the EGRA's teacher-training cascade, described previously.

Sample Selection

All 350 facilitators (PEAs and key teachers) who were receiving training from the master trainers during the training-of-trainers stage of the EGRA's cascade completed the P_0 and P_1 assessments. (We incorporated their scores into some of the regression models that we built to analyze the SMS treatment

Table 5-4. Allocation of sampled respondents to treatment and control groups

Performance category per results on P_0 and P_1 assessments	Number of teachers		
	Population	Treatment (50%)	Control (50%)
Perfect score on both P_0 and P_1	122	61	61
Score improved from P_0 to P_1	518	259	259
No score improvement from P_0 to P_1	193	96	97
Overall score decreased or remained the same, despite new answers	274	137	137
Total	**1,107**	**553**	**554**

P_0 = pre-training; P_1 = post-training.

effect; see Table 5-10 in the Findings section.) During the zonal training workshop stage of the cascade, 1,534 grades 1–3 teachers were selected to participate in the study via probability proportional to size (PPS) sampling, with the school as the sampling unit. This means that bigger schools had a higher chance of being included in the sample. Schools also were sampled using PPS, based on the number of grades 1–3 teachers in the school. Where the number of teachers required from the school was fewer than the number of grades 1–3 teachers on staff, participating teachers were selected at random. These teachers completed both P_0 and P_1 assessments at the training workshop.

Of the 1,534 teachers initially assessed at P_0, the project monitoring and evaluation team was able to match P_1 assessment records for 1,454 teachers. The EGRA's master database of teacher information contained telephone numbers for 1,107 of the 1,454 matched respondents; these 1,107 respondents were grouped based on their performance on the P_0 and P_1 assessments, and the groups were evenly allocated to SMS treatment or control. Table 5-4 illustrates the allocation of respondents.

Approximately 10 weeks[4] after the conclusion of the teacher training, EGRA personnel followed up with the teachers sampled at the zonal training workshop stage to administer the P_2 assessment. These final P_2 assessments were collected and matched at the 10-week follow-up point. This phase of the

[4] Teachers were trained in waves based on the classes they taught, starting on March 28, 2016; training had fully concluded by April 8. P_2 follow-up data began to be collected on June 20, with most having been collected by July 1 but some lagging until July 14. As a result, most respondents' P_2 scores would have been collected 10–11 weeks following the training, but some may have been collected as late as 15 weeks after training had concluded.

Table 5-5. Sample sizes and rates of attrition, by training cascade level

Cascade level	Sample: Targeted or achieved	Training of trainers: No. of PEAs and key teachers	Zonal training workshop: No. of grades 1–3 teachers	Teacher attrition relative to initial sample: Count/%
Pre-training (P_0)	Targeted	350	1,534	0 (0%)
	Obtained	350	1,534	
Post-training (P_1)	Targeted	350	1,534	80 (5.22%)
	Matched: P_0–P_1	350	1,454	
Follow-up (P_2)	Targeted	N/A	1,107	280 (25.3%)
	Matched: P_1–P_2	N/A	827	
Overall	Targeted	N/A	1,534	707 (46.1%)
	Matched: P_0–P_1–P_2	N/A	827	

No. = number; PEA = primary education advisor.

study obtained results for 827 of the 1,107 teachers; the final-stage attrition rate of 25.3 percent was not statistically significantly different between treatment and control groups. Table 5-5 provides details of the sample size and attrition across the various levels of assessment.

As Table 5-5 indicates, attrition was significant, both between successive assessments and across the entire series. As a low-level measure to protect the respondents' anonymity, the respondents were requested to use a unique identifier (their participant number on the training session's pre-printed attendee registration sheet) at the top of each survey, rather than their names. Some respondents provided inaccurate information (such as entering the wrong participant number, established by determining that it matched a participant of a different sex whose records were successfully matched), while others omitted the item or wrote it illegibly. An inability to match some respondents' P_1 scores to their P_0 scores reduced the sample size from 1,534 to 1,454 (a loss of 80 respondents, or 5.22 percent of the initial sample).

Because the P_2 assessment period did not coincide with a training workshop, it was not possible to use attendance register numbers as unique identifiers. Therefore, the team generated a second unique identifier that concatenated each participant's sex, grade level of instruction, district, zone, and school identity code; the resultant string of characters was both unique and persistent across time periods, as none of its component

parameters changed from P_0 through P_1 and P_2. This new identifier was provided to each PEA who administered the P_2 follow-up assessment for their teachers to mark on their assessment forms. Of the 1,107 teachers for whom phone numbers were available and who were thus assigned to SMS treatment or control conditions, an inability to match respondents' P_2 scores to their P_1 and P_0 scores further reduced the sample size from 1,107 to 827 (a loss of 280 respondents, or 25.3 percent of the sample). Overall, 707 respondents did not ultimately participate, or 46.1 percent of the initial sample.

Findings

Effect of SMS on Teacher Knowledge Retention

The Learning Gains measurement methodology and sampling were intended to evaluate the effectiveness of the additional SMS support on teachers' ability to retain information imparted during the training workshop about the activity's expected classroom teaching methods. If teachers in the treatment group scored significantly better on the P_2 assessment than teachers in the control group, we could reasonably attribute that improvement to the effect of the messaging. However, teachers recorded high scores on the Learning Gains instrument at *all three* assessment points, with very little variation overall and no differences at the $p < .05$ level between treatment groups. (As will be discussed in the Limitations section, this lack of variability in outcomes posed an obstacle in detecting intervention effects.) While scores were high overall, there was some variation within and across sections. Table 5-6 shows the percentage of the respondents providing correct or positive answers to each of the assessment items.

The items under *Section 1: The lesson cycle* tended to target more factual responses, whereas items under *Section 2: Classroom management* and *Section 3: Supplementary readers* tended to target more attitudinal responses. When considering chi-square significance, only one of the five items in Section 1 yielded scores that were greater at P_2 than at P_1, with one additional item yielding scores that were improved at P_1 and P_2 relative to P_0. In contrast, scores improved from P_0 to P_1 for five of the seven items in Section 2; all scores on all items improved from P_0 to P_2; and scores on five out of seven items improved from P_1 to P_2. For Section 3, scores improved from P_0 to P_1 for two out of six items; scores on all items improved from P_0 to P_2; and scores on five out of six items improved from P_1 to P_2.

Table 5-6. Respondent scores for each assessment item across three assessments, for teachers who recorded Learning Gains scores at each assessment round (n = 827)

Item		Percent correct at P_0	Percent correct at P_1	Percent correct at P_2
Section 1: The lesson cycle				
1	Order of the three stages of the lesson cycle	89%	94%*	94%*
2	Purpose of the "I do" stage	80%	81%	82%
3	Purpose of the "We do" stage	96%	96%	98%
4	Purpose of the "You do" stage	86%	89%	92%*†
5	During "You do," teacher should . . .	94%	95%	95%
	The lesson cycle subtotal	**4.46**	**4.55***	**4.61*†**
Section 2: Classroom management				
1	More than half the class is able to read	70%	76%*	79%*
2	Learners in the infant classes can be trained	95%	97%	99%*†
3	Different activities require different kinds of learner seating	71%	76%*	81%*†
4	Lesson preparation is essential for classroom management	93%	95%	97%*†
5	Young learners learn better when they have books in their hands	65%	71%*	78%*†
6	Allowing children to make mistakes is helpful	90%	93%*	97%*†
7	I will definitely implement	95%	98%*	98%*
	Classroom management subtotal	**5.81**	**6.07***	**6.29*†**
Section 3: Supplementary readers				
1	Supplementary readers are for all the learners	93%	95%	97%*
2	Parents who are illiterate can help	76%	83%*	88%*†
3	My learners need to have a chance to read	89%	93%*	96%*†
4	Parents are very keen for their children to borrow books	85%	86%	91%*†
5	Use the supplementary readers for group and individual reading	79%	81%	86%*†
6	Will definitely allow learners to borrow books	95%	96%	99%*†
	Supplementary readers subtotal	**5.17**	**5.35***	**5.56*†**
	Overall total	**15.43**	**15.96***	**16.46*†**

P_0 = pre-training; P_1 = post-training; P_2 = follow-up.

* = $p < .05$ when comparing (P_0 and P_1) or (P_0 and P_2); † = $p < .05$ when comparing P_1 and P_2.

Table 5-7. Scores of respondents matched across three assessments, by treatment stage

	Treatment				Control			
Treatment stage	***n***	**Mean score**	**SD**	**SE**	***n***	**Mean score**	**SD**	**SE**
P_0	414	15.38	2.41	0.12	413	15.49	2.16	0.11
P_1	414	15.84	2.18	0.11	413	16.08	1.92	0.09
P_2	414	16.44	1.69	0.08	413	16.48	1.69	0.08

P_0 = pre-training; P_1 = post-training; P_2 = follow-up; SD = standard deviation; SE = standard error.

Tables 5-7 and 5-8 detail respondent scores for P_0, P_1, and P_2 by treatment status (teachers who received EGRA SMS messages versus those who did not). There were no statistically significant differences in scores between treatment and control teachers at any of the points of assessment, either before the intervention or after.

SMS Sharing Among Teachers

Over the course of the SMS intervention, the project team (including the chapter authors) became aware that some teachers who had been selected into the treatment group were sharing SMS messages with their peers. As the campaign designers, we had not anticipated this behavior and as such we neither encouraged nor discouraged it; as a result, the P_2 instrument was modified from the P_1 and P_0 instrument by the addition of three simple questions that aimed to quantify the phenomenon. Table 5-9 summarizes these yes/no and other responses that were provided by the teachers whose scores could be matched across all three assessments.

Interpretation of these data requires caution because the data appear somewhat inconsistent with respect to skip logic. For instance, the number of total respondents ("Yes" respondents plus "No" respondents) who answered item #9 ("If yes, did you share them with your peers?") does not tally with the number of respondents who had acknowledged receiving EGRA SMS messages; the number of respondents who replied to item #10 ("If no in *8* above, did a peer who received these messages share them with you?") likewise does not tally with the number of respondents who indicated they had *not* received EGRA SMS messages. These results pose a limitation in understanding the extent of SMS-sharing behaviors among teachers receiving the messages. While the numbers suggest that the overwhelming majority of teachers who received SMS messages (322 out of 329, or 97.9 percent) shared

Table 5-8. Respondent scores for each assessment item across three assessments, by treatment status

Item		Treatment (received SMS directly) (n = 413)			Control (did not receive SMS directly but received the same training) (n = 414)		
		% correct at P_0	% correct at P_1	% correct at P_2	% correct at P_0	% correct at P_1	% correct at P_2
Section 1: The lesson cycle							
1	Order of the three stages of the lesson cycle	91%	96%	94%	88%	92%	95%
2	Purpose of the "I do" stage	79%	81%	83%	80%	81%	81%
3	Purpose of the "We do" stage	97%	97%	99%	96%	96%	97%
4	Purpose of the "You do" stage	85%	88%	92%	88%	89%	92%
5	During "You do," teacher should . . .	94%	95%	96%	94%	94%	95%
	The lesson cycle subtotal	**4.46**	**4.57**	**4.63**	**4.45**	**4.52**	**4.59**
Section 2: Classroom management							
1	More than half the class is able to read	70%	77%	80%	71%	76%	78%
2	Learners in the infant classes can be trained	97%	97%	99%	94%	97%	99%
3	Different activities require different kinds of learner seating	74%	79%	83%	68%	74%	79%
4	Lesson preparation is essential for classroom management	92%	94%	97%	94%	95%	97%
5	Young learners learn better when they have books in their hands	63%	71%	73%	67%	72%	82%
6	Allowing children to make mistakes is helpful	91%	93%	98%	90%	93%	96%
7	I will definitely implement	97%	99%	99%	94%	97%	98%
	Classroom management subtotal	**5.84**	**6.09**	**6.28**	**5.78**	**6.04**	**6.3**
Section 3: Supplementary readers							
1	Supplementary readers are for all the learners	94%	96%	97%	92%	94%	96%
2	Parents who are illiterate can help	76%	85%	88%	75%	81%	87%

(*Continued*)

Table 5-8. Respondent scores for each assessment item across three assessments, by treatment status (*Continued*)

Item		Treatment (received SMS directly) (*n* = 413)			Control (did not receive SMS directly but received the same training) (*n* = 414)		
		% correct at P_0	% correct at P_1	% correct at P_2	% correct at P_0	% correct at P_1	% correct at P_2
3	My learners need to have a chance to read	90%	95%	96%	88%	91%	95%
4	Parents are very keen for their children to borrow books	84%	86%	91%	85%	86%	92%
5	Use the supplementary readers for group and individual reading	79%	82%	85%	79%	81%	87%
6	Will definitely allow learners to borrow books	95%	98%	99%	95%	95%	98%
	Supplementary readers subtotal	**5.19**	**5.41**	**5.56**	**5.15**	**5.28**	**5.56**

P_0 = pre-training; P_1 = post-training; P_2 = follow-up; SMS = short message service.

Table 5-9. Self-reported SMS receipt and sharing practices

Item	Frequency	Percent	Cumulative frequency	Cumulative percent
8. Over the past several weeks, EGRA sent frequent SMS messages to a random sample of teachers. Did you receive these messages directly from EGRA?				
No response: Missing	28	3.4%	28	3.4%
No	470	56.8%	498	60.2%
Yes	329	39.8%	827	100.0%
9. If yes, did you share them with your peers?				
No response: Missing	100	12.1%	100	12.1%
No	216	26.1%	316	38.2%
Yes	322	38.9%	638	77.2%
No response: Skipped	189	22.9%	827	100.0%
10. If no in *8* above, did a peer who received these messages share them with you?				
No response: Missing	53	6.4%	53	6.4%
No	333	40.3%	386	46.7%
Yes	249	30.1%	635	76.8%
No response: Skipped	192	23.2%	827	100.0%

EGRA = Early Grade Reading Activity; SMS = short message service.

them with peers, and a significant majority of the teachers in the control group may have received them from peers (333 out of 470, or 70.9 percent), it is not clear that these figures are wholly trustworthy.

We fit several regression models in an attempt to understand whether the SMS intervention had an effect when we controlled for other factors such as sex, educational district, the quality of the trainers delivering the workshop (measured using the Learning Gains instrument), the number of coaching visits received, and the teachers' most recent scores on the EGRA teacher observation instrument, which was used to create a teacher performance index. (Details regarding the construction of the performance index can be found in Pouezevara et al., 2016). Table 5-10 presents the results of several such models.

First, we established a "base" model using the respondents' P_2 scores as the dependent variable. The base model included the respondents' P_1 scores, their sex, their treatment status, and the district in which they were trained and were working. In some variants, the number of coaching visits the respondents had received over the lifetime of the project was included as well; where the coaching visits were included, the respondents' latest score on an index of teacher practices that supported reading acquisition was also included. However, since including the coaching visitation variable significantly reduced the number of respondents, we ran some of the models without it as well as coaching-related factors.

Generally speaking, the regression models suggested that the only factor that was consistently predictive of a teacher's Learning Gains score at P_2 was that teacher's score at P_1. Model 1, which incorporated the proportion of trainers (PEAs and key teachers) who had earned perfect scores on their own P_0 and P_1 assessments, found it to be significant at $p < .05$. A regional (geographic) factor was significant at $p < .05$ in six of nine districts, but only in Model 4, which included the number of trainers present at the training as an additional factor. While the presence of regional (geographic) variation in scores tallied with EGRA team members' anecdotal observations that the quality and degree of implementation of the EGRA's overall reading intervention varied substantially from district to district, the fact that it is present in only one model of the four suggests it may have been a spurious result.

The alternative "base" model, which incorporated coaching-related factors, is not presented here for the sake of parsimony. With the addition of the coaching factors, the respondents' sex became significant at $p < .05$ for all models; however, nothing else was significant apart from the P_1 score. It did,

Table 5-10. Results of regression models controlling for non-intervention factors

Factor	Model 1	Model 2	Model 3	Model 4
Analysis population	768	768	768	768
R^2	0.15	0.15	0.15	0.15
Model *p*-value (*F*-test)	<.0001	<.0001	<.0001	<.0001
Intercept	11.54***	11.79***	11.41***	12.52***
P_1 total score	0.30***	0.30***	0.30***	0.31***
Female teacher	0.11	0.1	0.08	0.06
Treatment group	0.07	0.03	0.1	0.16
District				
Mzimba North	1.57	1.57	1.35	0.44
Ntchisi	–0.29	–0.19	–0.36	–1.23*
Salima	0.06	0.08	0.14	–0.78
Lilongwe Rural East	–0.1	–0.21	–0.11	–0.98*
Lilongwe Rural West	–0.16	–0.12	–0.13	–0.99*
Ntcheu	0.02	0.02	0.02	0.27
Balaka	–0.14	–0.11	–0.12	–1.12*
Machinga	–0.04	–0.08	–0.02	–0.95*
Zomba Rural	–0.17	–0.16	–0.14	–0.97*
Proportion of "perfect-score" facilitators	0.36*			
Average facilitator difference from perfect score		–0.07		
Proportion of "perfect" or "improved" facilitators			0.32	
Number of trainers present at workshop/training				–0.1

P_1 = post-training.

* = $p < .05$, *** = $p < .001$.

however, increase the amount of variance explained in P_2 scores (R^2) to 24 percent for all models, as compared with the 15 percent that was explained by the original base model.

In sum, the intervention appears to have had no effect on P_2 scores. The only factor consistently correlated with P_2 scores was P_1 scores.

Costs

The cost of the SMS intervention was modest. In this section, we detail costs for two reasons: to provide clarity on the specifics of our fieldwork and to

Table 5-11. EGRA SMS traffic during SMS campaign, with worst-case cost projections and actual costs

Month	Total EGRA SMS messages	Messages sent for SMS study	SMS messages payable at marginal rate	Real cost in USD/kwacha	Projected cost in USD/kwacha absent CSR pricing*
May	1,372	1,372	0	USD 0.00/0 kwacha	USD 52.68/ 32,298 kwacha
June	13,913	12,042	3,913	USD 38.52/ 25,825 kwacha	USD 462.41/ 289,008 kwacha
July	8,403	7,275	0	USD 0.00/ 0 kwacha	USD 279.36/ 174,600 kwacha

CSR = corporate social responsibility; EGRA = Early Grade Ready Activity; SMS = short message service.

* This estimate is based on a worst-case assumption that none of the teachers would have had Airtel Malawi phone numbers, so the SMS messages would have been charged at out-of-network rates (24 kwacha/USD 0.038) rather than at in-network rates (15 kwacha/USD 0.024). In fact, roughly 50 percent of EGRA teachers used Airtel, and roughly 50 percent used Telekom Networks Malawi (TNM), Malawi's second-largest telecom firm at the time.

voice our support for future related work to track and report the costs associated with the use of SMS messaging or other technologies in education systems because the financial implications of any research-based intervention must be clear if the research activity is to inform a discussion about systemic uptake or policy reform.

Through a corporate social responsibility (CSR) arrangement the EGRA team had negotiated with Airtel Malawi two years prior, the cost of the EGRA's first 10,000 SMS messages in any month was borne entirely by the telecom company; beyond 10,000 SMS messages, the EGRA incurred a unit cost of 6.60 kwacha (approximately USD 0.010).[5] Table 5-11 shows the volume of SMS traffic and the EGRA's associated costs, as well as projections of what the costs would have been absent its CSR arrangement with Airtel Malawi.

Table 5-12 details the setup and infrastructure maintenance costs.

Table 5-13 compares the actual costs incurred by the EGRA's SMS campaign and a worst-case projection of costs had the CSR arrangement not been in place.

[5] Malawi's currency, the kwacha, depreciated substantially over the lifespan of the EGRA intervention. Costs under the CSR arrangement remained fixed in kwacha terms throughout the program, but the equivalent costs in US dollars plummeted. For this chapter's purposes, we estimated infrastructure fees in dollar terms using the exchange rate of 489.8 kwacha per USD 1.00 that was in place on November 27, 2014, when the CSR arrangement was formalized. We estimated per-SMS costs using the exchange rate of 670.5 kwacha per USD 1.00 that was in place on May 25, 2016, when the campaign was launched.

Table 5-12. Breakdown of SMS campaign costs and estimate of per-participant expense (in USD)

Expense item	Cost	Recurrence	Cost per month
Telerivet* annual fee	$2,304.00	Yearly	$192.00
Airtel integration	$120.00	Monthly	$120.00
Airtel one-time setup	$408.32	One-off	$34.03
RTI one-time setup	$476.00	One-off	$39.67
Short code** maintenance	$100.00	Yearly	$8.33
Infrastructure costs for seven-week campaign:			**$479.55**

SMS = short message service.

* Telerivet (http://telerivet.com) operates an international SMS web portal that facilitates voice calls for surveys and mass-messaging campaigns. It is roughly akin to its somewhat better-known competitor Frontline SMS.

** Short codes are brief sequences of digits (shorter than typical phone numbers) that mobile networks use for addressing SMS messages.

Table 5-13. Comparison of actual costs and projected costs absent CSR pricing (in USD)

	Actual costs incurred		Projected costs absent CSR pricing	
Campaign costs	**Cost**	**Percent of total**	**Cost**	**Percent of total**
Infrastructure	$689.55	94.7%	$689.55	48.2%
SMS delivery	$38.52	5.3%	$740.55	51.7%
Total	$728.07	100.0%	$1,430.09	100.0%
Cost per teacher receiving messages (485 participants)	$1.50		$2.95	

CSR = corporate social responsibility; SMS = short message service.

The cost of the intervention is thus estimated to have been USD 728.07, with USD 689.55 (94.7 percent) of that expense being infrastructure costs and only USD 38.52 (5.3 percent) being delivery costs. For the 485 teachers who received the SMS messages in the seven-week campaign, the cost per teacher was approximately USD 1.50. Absent a CSR arrangement, the cost would have nearly doubled, and the SMS delivery costs would have accounted for 51.7 percent of the cost of the campaign.

These figures represent direct costs only and as such do not attempt to factor in the costs of labor spent in the SMS system setup or SMS content management. That said, the Telerivet contact-management, message-creation, and scheduling interfaces were all fairly simple to master—requiring an investment of a few hours, not days—and once the initial configuration was

complete, the campaign was an entirely hands-off affair. That is, the project team created recipient lists (based on the assignment of respondents to the "SMS treatment" groups), typed in the messages, and scheduled them for delivery at specific times on later dates. After those tasks had been completed, the campaign required no further input; staff simply performed periodic verifications that the number of system-confirmed recipients matched the number expected upon initial setup.

Limitations

As indicated previously, this study faced several limitations. First and foremost, the instrument used to assess teacher content mastery was hampered by ceiling effects and thus appears to have been unable to distinguish between higher-performing teachers and lower-performing teachers. For 7 of the 18 items, over 90 percent of the teachers assessed answered the item correctly at the P_0 time point; for 5 more of the items, over 84 percent of respondents answered the item correctly at P_0. The combination of such a high initial score and a limited number of discrete items (18) resulted in limited scope for improvement. A more sophisticated assessment instrument featuring items tightly tied to the broader EGRA intervention design, in more subtle or less obvious ways—and relying on a harder-to-game evaluation approach (such as a principal components analysis) rather than simple scoring—might have revealed different outcomes by way of being a more genuine and precise measurement.

Equally important, the particular constructs being assessed—PEAs' and teachers' knowledge of certain concepts related to reading instruction, classroom management in a context of extreme class sizes, and the use of supplementary readers to complement curricular materials—are distinct from teacher *practice*. A teacher may be highly knowledgeable, yet incapable of delivering a high-quality lesson to his or her students. Similarly, a teacher may have performed poorly on the particular items included on this assessment and still have been an effective instructor in the classroom. While some of the regression models incorporated data from classroom observation records, the lack of observation data from the P_2 time point impeded a rigorous analysis of the relationships between the assessment scores and teaching practice. Conducting nuanced observations of the teachers' classroom practices would have added significant value to understanding the relationships among the SMS campaign, the assessment instrument, and potential classroom outcomes.

Because the EGRA SMS campaign and Learning Gains assessment represented an exploratory and observational study, differences in teachers' scores in the P_1–P_2 interval could not be conclusively attributed to the SMS intervention. Furthermore, as described previously, a degree of contamination was observed in the SMS intervention. Brief telephone calls were placed to a handful of teachers within the first week of the SMS campaign to confirm that messages were being received and that message timing was appropriate. During those calls, some teachers indicated that they were sharing the messages they received with their peers. While the EGRA program had not explicitly forbidden this behavior, it also had not encouraged it. Still, upon learning of the SMS sharing, the EGRA team did not intervene to attempt to promote or discourage this practice. Rather, the EGRA team sought simply to measure the prevalence of the SMS sharing, and for this reason the P_2 version of the assessment instrument was modified to include the three questions intended to understand the extent of these sharing practices.

However, two factors impeded a rigorous analysis of the data on the sharing of SMS messages among teachers. First, the questions did not specifically ask about sharing *with peers who were included in the study*; as a result, we have limited insight into the extent of contamination. (For instance, the receiving teacher may have shared the SMS messages with peers who were not included in the study's control group; in such a case, there would be no contamination.) The study would have been strengthened by foreseeing the potential for such behavior from the outset and designing the sampling and analysis plan accordingly. For example, had we assigned entire schools—rather than individual teachers—into treatment or control groups, it is conceivable that we would have minimized contamination effects because teachers would have been less likely to share SMS messages with their closest peers. Second, the instrument was not designed either to ascertain whether recipients had retained any of the messages or themes included in the SMS messages or to probe their thoughts and feelings about receiving the messages.

The study of the SMS campaign would have been strengthened by including a qualitative element (such as focus group discussions) to examine the nature of teachers' interactions with the SMS messages themselves. In a similar vein, the EGRA program made no attempt to test out the effectiveness of different SMS campaign types; for example, could more frequent messages,

of a different tone and directly relevant for a geographic area, have been more impactful? As noted in the discussion of Table 5-2, messages were not specifically linked to the content of the lessons the teachers were teaching. Anecdotally, on any given day, the teachers participating in the EGRA could have been teaching very different lessons; some might have been far ahead of the expected pace, others far behind. In the absence of a sophisticated and adaptive messaging system, targeting SMS content to a teacher's lesson for the day would have required a substantially greater investment in labor than the EGRA program could afford. Our analysis is therefore unable to provide insight into the possible impact of a similar SMS intervention that would tightly link message content with lesson content for each teacher.[6]

For this chapter, we did make an effort to accurately capture all historical classroom observation data (rather than those from just the visits during the P_1–P_2 interval). However, a review of the visit data suggested that the number of coaching visits between the treatment and control groups was unbalanced, with control group teachers receiving, on average, more coaching visits than the treatment group teachers. Table 5-14 presents the number of coaching visits received by the respondents.

Only 249 of the teachers who were assigned to the control group ever received a coaching visit under the EGRA; the substantial majority of teachers sampled for this study were visited only once or twice, as Figures 5-1 and 5-2 underscore. As a result, incorporating the teacher observation instrument data into regression models added only limited additional value and did so at the cost of a substantial reduction in the number of respondents upon which the model could draw.

Another limitation of the SMS campaign related to its timing. When the trainings for teachers at the P_1 time point concluded, the participating teachers' mastery and retention of training concepts was presumably at its peak. Due to some logistical constraints and the time needed to draw the sample frame for the study, the SMS campaign did not begin until approximately five weeks after the P_1 assessment. Had the campaign begun immediately after the conclusion of zonal trainings, or shortly thereafter, it is

[6] Since 2015, RTI has been using a software system called Gooseberry to support multistep, interactive registration of teachers via SMS during trainings for the Tusome Early Grade Reading Activity. Gooseberry is highly configurable and relatively simple to manage. A replication of this study could be strengthened by using Gooseberry to provide teachers with a toll-free telephone number to which they could send their lesson day and week and from which they would subsequently receive a response based on that information.

Table 5-14. Number of teachers of each treatment status receiving the listed number of coaching visits, over the entire lifespan of the EGRA's implementation

	Control teachers		Treatment teachers		Cumulative
No. of visits	**%**	***n***	**%**	***n***	**total**
0	39.7%	164	57.5%	238	402
1	39.0%	161	25.1%	104	667
2	15.3%	63	13.0%	54	784
3	4.8%	20	3.4%	14	818
4	1.2%	5	0.7%	3	826
5	—	—	—	—	—
6	0%	0	0.2%	1	827
Total	**100.0**	**413**	**100.0**	**414**	**827**

EGRA = Early Grade Reading Activity; no. = number.

conceivable that the SMS content would have more clearly facilitated knowledge retention.

Conclusions

As our review of the literature indicates, there is strong reason to believe that an SMS-based intervention could be a powerful mechanism to support behavior change. The causal theory at play—that familiar technology can help reinforce learning and behaviors with short and timely reminders—is well documented in education and elsewhere. Unfortunately, our study design and implementation contained flaws that precluded us from drawing evidence-based conclusions about such a program's effectiveness. We believe, however, that a critical examination of our missteps may prove instructive for others who would seek to add to the research base regarding the use of SMS to support behavior change in development programs. We use this Conclusions section to provide actionable guidance to those who would succeed and improve upon our efforts.

The results of our study suggest that further investigation is warranted, using a better-calibrated instrument allowing for greater discrimination in teacher knowledge across a wider range of scores. While our analyses found no impact from the SMS campaign on their own, given the substantial limitations noted earlier, it would be premature to conclude that a more-tailored, better-targeted SMS campaign whose impact is measured with a better assessment tool may not demonstrate a meaningful impact.

Avenues for Further Exploration of Teachers' SMS-Sharing Behaviors

One of the most interesting findings—and one that also warrants more investigation, given its potential implications for social behavior–change communication campaigns with teachers—relates to the SMS recipients' sharing behaviors. As mentioned in the Limitations section, teachers who were selected to receive the SMS treatment were neither instructed to share the SMS messages with peers nor explicitly requested to avoid doing so. However, nearly all of the teachers in the treatment group (97.9 percent) appeared to have done so of their own volition. With the caveat that the figures reported by teachers were inconsistent with the expected skip logic, it appeared that up to 70.9 percent of the teachers assigned to the SMS control group may have been "contaminated" by such well-meaning sharing of the messages. This finding led us to believe that teachers found value in the messages, even if the effect of messaging on teacher content knowledge could not be directly linked.

This behavior presents multiple opportunities for further exploration. A carefully designed improvement on this study could qualitatively explore the factors that led to certain teachers choosing to share SMS messages with peers while others did not; whether they passed the messages to teachers at their own school or at others; whether they shared only with teachers in the EGRA grades (1–3) or even with colleagues in higher grades; and so on. It could also quantitatively explore whether certain messages (or message *types*) were more or less likely to be shared; whether those sharing patterns were consistent across all teachers or whether they varied according to teacher characteristics; and whether the act of sharing spurred peer discussions about teaching or classroom management practices. Including a treatment arm in which teachers received only generic "encouraging" SMS messages completely unrelated to the technical substance of the reading intervention, rather than messages focused on instructional issues, would enable a more nuanced understanding of teachers' sharing behaviors.

SMS and Social Learning as an Entry Point for Changing Teacher Practice

As suggested in the Literature Review section, the motivation to use technology in education programs must be based on the added value of the technology to do something that traditional methods as yet have not been able to do. In many cases, this added value comes in the form of increasing access to or receipt of information—although it may depend on what the user

does with the information as a result. The ease with which teachers received and shared simple instructions via text messages seems to have facilitated ongoing social learning to some extent, but as the teachers' SMS sharing was unplanned and as such unstructured, future research might consider how the SMS medium (or related tools) could be used to change pedagogical practice as a precursor to influencing changes in teachers' attitudes and beliefs about instruction.

The purpose of teacher professional development programs is to improve student outcomes through some effect on teachers' knowledge, pedagogical practices, beliefs, and attitudes. As Guskey (2002) pointed out, the presumption among designers of professional development programs has historically been that changing a teacher's beliefs about a particular instructional innovation will lead to the adoption of that innovation, which in turn will lead to improved student outcomes. Guskey noted, however, that at least in the case of experienced teachers, this presumption likely does not hold. Instead, Guskey (1986) proposed a model in which teachers' behaviors and attitudes are the *last* thing to change. That is, before teachers will update their belief systems about what works, they must first see concrete improvements in their students' outcomes as a direct result of specific pedagogical innovations they have introduced and tested in their classrooms.

A crucial element in Guskey's model is thus teachers' willingness to at least pilot the pedagogical innovations in their classrooms. If the piloting bears fruit in terms of improved student outcomes, updated beliefs and attitudes will follow, and long-term shifts in teacher practice may take hold. If the piloting never happens in the first place, however, there is no opportunity for student outcomes to increase as a result of the pedagogical innovation, and teacher practice will not change.

Future studies could be explicitly designed to explore the relationship between teachers' willingness to share the SMS messages they received, social learning, and the first crucial step in Guskey's model of teacher behavior change: willingness to try small changes to classroom practice. For instance, SMS messages could be used to nudge teachers to pilot in their classrooms the new techniques on which they were trained and to discuss with peers their sense of whether the techniques "worked." Different formulations of the SMS nudges could be tested out to see whether any were more consistently effective in catalyzing action on the teachers' part and whether the type of action (e.g., piloting new techniques, discussing the results with peers, attempting to

fully adopt new techniques) varied at all as a function of the message content or message frequency. If certain formulations seemed to lead to greater discussion among peer teachers, would that social element lead to greater piloting and/or long-term uptake of the new techniques?

Avenues for Further Exploring the Role of SMS Interventions in the Design of Teacher-Training Strategies

Given the low marginal cost of the EGRA SMS messages (albeit after a moderately substantial initial investment), the finding that 10 weeks of SMS as a medium could influence teachers' retention or consolidation of factual knowledge (or increase their alignment with desired attitudes and beliefs) would have significant implications for educational interventions attempting to reach teachers at scale and across large physical distances. Although the EGRA team's SMS campaign did benefit from a subsidized SMS rate from Airtel Malawi, it is worth noting that, absent this corporate support, the per-teacher campaign cost was still estimated to have been less than USD 3.00. Although the CSR support that the EGRA received should be acknowledged and sought in any replication of this study, the incremental costs without this support would not drastically change the business case for including similar campaigns as part of a broader literacy intervention, as evidenced by the HALI and PRIMR projects described in this chapter.

A tightly designed replication of the study that rigorously isolates and identifies effects attributable to SMS would point to several additional avenues of investigation. For instance, might there be a limit to the duration of the effect of an SMS campaign or an inflection point at which additional messaging might cease to generate a return? Sending a teacher too many SMS messages in a given period might lead to a sort of saturation in which the teacher might begin tuning out the messages; alternatively, even if a saturation point were not reached, the SMS messages might have initially been a welcome novelty but cease to be exciting after four months or so of continual receipt. In either case, the impact of the messaging intervention might then diminish or disappear altogether. Further research might also investigate other noncognitive effects of the SMS program, such as maintaining motivation and engagement with a professional development program (Dubeck, Jukes, Brooker, Drake, & Inyega, 2015).

Understanding such effects would have actionable implications for the deployment of an SMS campaign in places like Malawi, where each academic

term is often 12–14 weeks in length and teacher-training programs often happen during inter-term breaks. If the influence of an SMS messaging campaign were found to be persistent over a substantial period of time—say, 10–12 weeks—the SMS medium might prove to be an effective way of ensuring that a Term 2 training program is building upon a more solid foundation (with respect to retention of Term 1 training content) than would otherwise be the case. This would, in turn, have implications for the design and delivery of training content, in that less time might need to be spent on remediation or review, and more time could be spent on delivering and reinforcing new content.

Besides the implications for training design, an SMS-based teacher support campaign would be intriguing from a cost perspective. As reported previously, the direct cost per teacher of the SMS campaign using the EGRA setup was USD 1.50; under the EGRA's reimbursement model, the amount budgeted to support PEAs' coaching visits each month was 25,000 kwacha/ USD 37.29. Obviously, the value of that trade-off would depend largely on how much *coaching* actually happened. Because analyses of the EGRA's coaching data suggested that the substantial majority of EGRA teachers received no coaching visits in any given six-week span, the project would have realized only minimal savings because the actual expenditures on coaching did not approach the budgeted allocations for coaching. For a program that enjoys greater success in mobilizing instructional coaches, however, an SMS intervention might provide some of the same benefits as a coaching visit in terms of teachers' consolidation of knowledge gained from training workshops.

Integrating SMS Interventions Into Teacher Coaching Processes

We are not implying that an SMS-based system, were it shown to be legitimately impactful, could or should be construed as an adequate substitute for in-person coaching. While in-person coaching is expensive in direct costs, opportunity costs, and administrative complexity, it is probable that more substantial additional benefits accrue through that relational experience than can be transmitted in a 160-character impersonal text message. Considering as well that coaching visits are rarely limited to classroom instructional support—they frequently provide valuable opportunities for administrative discussions with head teachers, inspection-type functions, and conversations with parents or school management committees, all of which can help

improve school quality in both tangible and intangible ways—there would not be a strong argument for *replacing* coaching with SMS-based support.

One could imagine, however, designing a coaching system that attempted to use SMS-based "nudges" to *complement* in-person coaching. In a mixed-methods study of teachers' classroom practices conducted earlier in the EGRA's period of performance, Mattos and Sitabkhan (2016) made note of teachers' on-the-fly modifications to the prescribed lesson plan and then interviewed them to understand the reasoning behind those modifications. The authors found that over half of the modifications affected the *content* of the lesson (52 percent), with nearly another quarter affecting the *structure* of the lesson (23 percent); their expert judgment was that 52 percent of the content modifications and 79 percent of the structural modifications were detrimental to the goal of the lesson. Such changes to lesson structure and content could be identified and recorded by non-expert coaches using classroom observation software such as RTI's *Tangerine:Tutor*® platform. Integrating *Tutor* with an SMS delivery mechanism could allow a program, in the weeks and months following an observation, to send the teacher SMS messages tailored specifically to address the motivations, points of confusion, or areas of weakness that led him or her to make those changes in the first place.

For those SMS messages to be maximally effective, however, the program would need to understand teachers' reactions to such messages. Would they welcome them and find them helpful or consider them intrusive and frustrating? Would teachers recognize the intended link between the SMS messages and their own classroom practices, or would they see the two as fundamentally unrelated? Even if they recognized the link, there is a difference between knowing about the effectiveness or importance of a given practice and actually implementing it within a lesson. Would even highly targeted SMS messages translate into observable differences in pedagogy that persisted between coaching visits? We would recommend that researchers interested in the use of SMS to support teacher behavior change invest prudently in up-close, expert qualitative inquiry to understand how teachers engage with the SMS medium and the messages' content, as well as whether and how that engagement affects classroom practice.

However, sound evidence regarding the mechanism by which instructional coaching drives improved student outcomes in the developing world remains far too limited. If the mechanism of change is merely a function of a

Hawthorne or observer effect—in which the simple awareness that someone is paying attention to the teachers' actions leads them to be more conscientious and diligent in their work, or encourages them to care more and thus invest more time and effort in preparing their lessons, etc.—then an SMS-based system might have a useful role as a low-cost reminder to the teachers that their work matters and is being tracked. Given what we have observed and reported with respect to the teachers' sharing of SMS messages among peers, it is quite plausible that teachers themselves valued the SMS content to some degree, and this alone merits further testing as to how best to leverage this cost-efficient, familiar medium for more sustained professional development efforts targeting early grade teachers.

Acknowledgments

The authors wish to thank the education office of the United States Agency for International Development (USAID)/Malawi for its leadership and support throughout the implementation of the Malawi Early Grade Reading Activity (EGRA) covered in this case study. Special appreciation goes to Kevin Roberts (Contracting Officer's Representative [COR]), Ramsey Sosola (Alternate COR), and Christine Djondo (Education Office Chief for the majority of EGRA's period of performance). All four authors were assigned to EGRA during its implementation and played major roles in the short message service (SMS) campaign described in this chapter. The contents of this chapter are the responsibility of the authors and do not necessarily reflect the views of USAID or the US government.

References

Aker, J., Ksoll, C., & Lybbert, T. J. (2010). *ABC, 123: The impact of a mobile phone literacy program on educational outcomes* (Working Paper 223). Washington, DC: Center for Global Development. Retrieved from https://www.cgdev.org/publication/abc-123-impact-mobile-phone-literacy-program-educational-outcomes-working-paper-223

Beltramo, T., & Levine, D. I. (2012). *Do SMS text messaging and SMS community forums improve outcomes of adult and adolescent literacy programs?* (Working Paper No. 015). Berkeley, CA: Center for Effective Global Action, University of California. Retrieved from http://escholarship.org/uc/item/1c31c2m4

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*, 354–380. https://doi.org/10.1037/0033-2909.132.3.354

Chauwa, A. (2015, March). Malawi losing billions on ghost teachers, nationwide strike looming. *Nyasa Times*. Retrieved from http://www.nyasatimes.com/malawi-losing-billions-on-ghost-teachers-nationwide-strike-looming/

da Costa, T. M., Barbosa, B. J. P., Gomes e Costa, D. A. G., Sigulem, D., de Fátima Marin, H., Filho, A. C., & Pisa, I. T. (2012). Results of a randomized controlled trial to assess the effects of a mobile SMS-based intervention on treatment adherence in HIV/AIDS-infected Brazilian women and impressions and satisfaction with respect to incoming messages. *International Journal of Medical Informatics*, *81*(4), 257–269. https://doi.org/10.1016/j.ijmedinf.2011.10.002

Dubeck, M. M., Jukes, M. C. H., Brooker, S. J., Drake, T. L., & Inyega, H. N. (2015). Designing a program of teacher professional development to support beginning reading acquisition in coastal Kenya. *International Journal of Educational Development*, *41*, 88–96. https://doi.org/10.1016/j.ijedudev.2014.11.022

Fisher, D., Frey, N., & Nelson, J. (2012). Literacy achievement through sustained professional development. *The Reading Teacher*, *65*(8), 551–563. https://doi.org/10.1002/TRTR.01082

Fjeldsoe, B. S., Marshall, A., & Miller, Y. (2009). Behavior change interventions delivered by mobile telephone short-message service. *American Journal of Preventive Medicine*, *36*(2), 165–173. https://doi.org/10.1016/j.amepre.2008.09.040

Gaible, E., & Burns, M. (2005). *Using technology to train teachers: Appropriate uses of ICT for teacher professional development in developing countries*. Washington, DC: infoDev/World Bank. Retrieved from http://www.infodev.org/articles/using-technology-train-teachers

Guskey, T. R. (1986). Staff development and the process of teacher change. *Educational Researcher*, *15*(5), 5–12. https://doi.org/10.3102/0013189x015005005

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching*, *8*(3), 381–391. https://doi.org/10.1080/135406002100000512

Jackson, D., Jere, M., Togher, A., & Webb, D. (2015). Initial Primary Teacher Education (IPTE) Programme: An independent review of the IPTE Curriculum undertaken on behalf of the Malawi Institute of Education and the Department of Teacher Education and Development. Lilongwe, Germany: GIZ (Deutsche Gesellschaft für Internationale Zusammenarbeit).

Jukes, M. C. H., Turner, E. L., Dubeck, M. M., Halliday, K. E., Inyega, H. N., Wolf, S., Zuilkowski, S. S., & Brooker, S. J. (2016). Improving literacy instruction in Kenya through teacher professional development and text messages support: A cluster randomized trial. *Journal of Research on Educational Effectiveness,* 1–33. https://doi.org/10.1080/19345747.2016.1221487

Kaleebu, N., Gee, A., Maybanks, N., Jones, R., Jauk, M., & Watson, A. H. A. (2013). SMS story: Early results of an innovative education trial. *DWU Research Journal, 19*(November), 50–62. Retrieved from http://www.dwu.ac.pg/en/images/Research_Journal/2013_Vol_19/5_Kaleebu_Gee_Maybanks_Jones_Jauk__Warson_SMS_Story.pdf

Kraft, M. A., Blazar, D., & Hogan, D. (2017). *The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence* (Working Paper). Providence, RI: Brown University and Harvard University. Retrieved from https://scholar.harvard.edu/files/mkraft/files/kraft_blazar_hogan_2016_teacher_coaching_meta-analysis_wp_w_appendix.pdf

Mattos, M., & Sitabkhan, Y. (2016). *Malawi Early Grade Reading Activity: Scripting study report.* Prepared for USAID/Malawi under the Malawi Early Grade Reading Activity, Contract AID-612-C-13–0002. Research Triangle Park, NC: RTI International. Retrieved from http://pdf.usaid.gov/pdf_docs/pa00mdzf.pdf

Ministry of Education, Science, and Technology (MOEST) [Malawi]. (2015). *Handbook for inspectors and advisors: September 2015*. Lilongwe, Malawi: Author.

Piper, B., & Korda, M. (2011). *EGRA Plus: Liberia. Understanding the causal mechanisms: EGRA Plus's effectiveness.* Prepared for USAID/Liberia under the Education Data for Decision Making (EdData II) project, Early Grade Reading Assessment (EGRA): Plus Project, Task Order No. EHC-E-06–04–00004–00 (RTI Task 6). Research Triangle Park, NC: RTI International. Retrieved from http://pdf.usaid.gov/pdf_docs/pnaea386.pdf

Piper, B., & Mugenda, A. (2012). *The Primary Math and Reading (PRIMR) Initiative: Baseline report.* Prepared under the USAID EdData II project, Task Order No. AID-623-M-11–00001 (RTI Task 13). Research Triangle Park, NC: RTI International. Retrieved from http://pdf.usaid.gov/pdf_docs/pa00hx75.pdf

Piper, B. P., & 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

Pouezevara, S. (2015). Revisiting the 'm' in m-learning: Making the most of mobile environments for teaching and learning in developing countries [Chesapeake, VA: Association for the Advancement of Computing in Education] [AACE]. *Proceedings of E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, *2015* (pp. 1350–1360). Retrieved from https://www.learntechlib.org/p/152173

Pouezevara, S., Pflepsen, A., Nordstrum, L., King, S., & Gove, A. (2016). *Measures of quality through classroom observation for the Sustainable Development Goals: Lessons from low- and middle-income countries.* Background paper prepared for the 2016 Global Education Monitoring Report. Education for people and planet: Creating sustainable futures for all. Paris, France: UNESCO. Retrieved from http://unesdoc.unesco.org/images/0024/002458/245841E.pdf

Pratham Education Foundation & Voluntary Service Overseas (VSO). (2015). *SMS Story Project: Bundi, Rajasthan* (Impact Assessment Report). Mumbai, India: Author. Retrieved from https://www.vsointernational.org/sites/default/files/sms_report_final_v1_4.pdf

Ramos, A. J., & Trinoña, J. P. (2009). Mobile technology in nonformal distance education. In J. Baggaley & T. Belawati (Eds.), *Distance education technology in Asia* (pp. 231–256). Lahore, Pakistan: University of Pakistan. Retrieved from http://www.pandora-asia.org/downloads/Book-2/PANdora-book2_v6.pdf

RTI International. (2010). *USAID funded Malawi Teacher Professional Development Support (MTPDS) Program. 2010 Early Grade Reading Assessment: National baseline report.* Prepared for Creative Associates International and USAID under ABE/LINK Task Order No. EDH-I-04–05–00026–00. Research Triangle Park, NC: Author. Retrieved from https://globalreadingnetwork.net/sites/default/files/eddata/Malawi_National_Baseline_EGRA_2010.pdf

RTI International. (2014). *USAID/Kenya Primary Math and Reading (PRIMR) Initiative: Final report*. Prepared for USAID under the EdData II project, Task Order No. AID-623-M-11–00001 (RTI Task 13). Research Triangle Park, NC: Author. Retrieved from http://pdf.usaid.gov/pdf_docs/PA00K282.pdf

Slade, T. S. (2016, March). *Getting instructional coaching right is hard—Why do we bother? Observations from USAID's Malawi Early Grade Reading Activity.* Delivered as part of the panel "Coaching Support for Teachers to Improve Early Grade Reading: Experiences, Results and Future Directions," 2016 Annual Conference of the Comparative and International Education Society (CIES), Vancouver, Canada.

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2014). *Teaching and learning: Achieving quality for all. Malawi: Fact sheet* (EFA Global Monitoring Report 2013/4). Paris, France: UNESCO. Retrieved from http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/ED/GMR/pdf/Malawi_Factsheet.pdf

York, B., & Loeb, S. (2014). *One step at a time: The effects of an early literacy text messaging program for parents of preschoolers* (NBER Working Paper Series, No. 20659). Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w20659