Fadeout and growth: Long-term impacts of an effective early grade reading intervention in South Africa

Jonathan M. B. Stern[a]*, Matthew C. H. Jukes[b], Jacobus Cilliers[c], Brahm Fleisch[d], Stephen Taylor[e], and Nompumelelo Mohohlwane[f]

[a]RTI International, Research Triangle Park, NC, USA; jstern@rti.org; [b]RTI International, Research Triangle Park, NC, USA; mjukes@rti.org; [c]Georgetown University, Washington, DC, USA; ejc93@georgetown.edu; [d]University of Witwatersrand, Johannesburg, South Africa; Brahm.Fleisch@wits.ac.za; [e]National Department of Basic Education, Pretoria, South Africa; Taylor.S@dbe.gov.za; [f]National Department of Basic Education, Pretoria, South Africa; Mohohlwane.N@dbe.gov.za

Abstract

This paper presents findings from a long-term, randomized evaluation of a structured pedagogy program implemented in South Africa for students in grade 1 through grade 3. This study tracked students over a period of 7 years, from the start of grade 1 into grade 7. We find evidence of a sustained impact of the intervention on Setswana (home language) literacy of 0.17 SD in grade 7, four years after students participated in the program. This constituted a fadeout for impacts on basic literacy skills, but an expansion for higher-order skills not targeted by the program, such as English written comprehension (as compared with estimated impacts from grade 4). There were also positive effects on improved grade attainment and the reduction of nonreaders, which provides evidence that the program positively affected some of the most at-risk or lowest-performing students. Overall, these results are promising for the potential impacts of well-implemented structured pedagogy interventions and support the idea that investing in foundational skills in early grades can lead to long-term improvements in literacy skills, as well as other outcomes (such as grade attainment).

Keywords: impact evaluation; early grade reading; structured pedagogy; long-term outcomes

1. Introduction

Foundational literacy and numeracy (FLN) skills have gained significant interest from the international donor community in recent years. These skills are seen as crucial building blocks for higher-order cognitive abilities. For instance, Sustainable Development Goal 4.4 aims to achieve universal literacy and numeracy, and the World Bank now regularly tracks "learning poverty"—the number of children who cannot read with comprehension by age 10. Several large initiatives sponsored by donor agencies, such as the Bill and Melinda Gates Foundation's Global Education Program, prioritize basic literacy and numeracy skills. The international education donor community also coalesced around support for FLN as their top priority for advocacy at the United Nations' Transforming Education Summit held in September 2022.

Although it is plausible that investments in FLN during early childhood may yield higher returns in the long run, it is also possible that early gains fade over time without complementary investments at a later age (Cunha et al., 2006; Johnson & Jackson, 2019;). Additionally, although empirical evidence has established a correlation between FLN at a young age and subsequent cognitive skills and earnings (Carter et al., 2020; Valerio et al., 2016), it remains unclear whether this association is driven by other factors, such as ability, a supportive home environment, or school quality. And socioeconomic status is a much stronger determinant of college attendance than are primary school test scores (Das et al., 2022). In fact, a recent review by Evans and Hares (2021) concluded that causal evidence in developing countries is lacking on the long-term impact of early investments in basic numeracy and literacy skills.

This paper contributes to this literature by presenting findings from a long-term follow-up of a randomized evaluation of a structured pedagogy program implemented in South Africa between 2015 and 2017. The program provided teachers with a combination of lesson plans,

teaching materials, and one-on-one coaching, to improve their teaching of early grade home language (Setswana) literacy. The Early Grade Reading Study (EGRS I) intervention was implemented over a period of 3 years, targeting consecutive grades each year (grade 1 in 2015, grade 2 in 2016, and grade 3 in 2017). Prior studies of EGRS I found a 0.24 standard deviation (SD) increase in student literacy after 2 years (Cilliers et al., 2021) and also some sustained impacts on teaching practices (Cilliers et al., 2022a).

Importantly, our current study has now tracked and assessed the same group of students over a period of 7 years, from the start of grade 1 in 2015 into grade 7 in 2021. We found that the learning gains from the original cohort of students—who participated in the interventions from grades 1 through 3—were sustained into grade 7,[1] 4 years after participating in the program (albeit with smaller effect sizes). When comparing the same group of students who were assessed in both 2018 and 2021 (grades 4 and 7), we found a 45 percent reduction in effect size for basic literacy skills, suggesting substantial fadeout over time for lower-order skills. However, effect sizes expanded significantly over time in English literacy, which was not prioritized in the original intervention. This suggests the possibility of spillover of FLN skills into higher-order cognitive skills.

2. Literature Review

Over the past decade or so, there has been a dramatic increase in the amount of evidence about both the learning crisis and program models that can improve learning at scale. The models can

[1] Grade 7 was the intended grade for on-track students, but some students repeated grades and were assessed in their 2021 grade.

be crudely divided into those that use a combined set of ingredients around structured pedagogy and those that focus on remediation (using differentiation, such as Pratham's Teaching at the Right Level (TaRL) approach). There is now a great deal of high-quality evidence of when and under what circumstances the models "work" and we even have crude but meaningful estimates of how much learning gains we could expect should these models be implemented with fidelity (Banerjee et al., 2016; Cilliers, 2022b; Piper et al., 2018; Stern et al., 2021).

Much of the research in LMICs has been conducted in India, Kenya, and South Africa. Building on the work of Pratham in India, researchers have provided a solid research base on TaRL (Banerjee et al., 2016; Lakhsman, 2019). In Kenya, the success of the randomized controlled trials in the Primary Mathematics and Reading (PRIMR) Initiative led to the implementation of the Tusome ("Let's Read") system-wide intervention (Gove et al., 2017; Piper et al., 2018; Piper & Zuilkowski, 2015). The South African research emerged in a partnership between government and universities (Cilliers et al., 2020; Cilliers et al., 2022b; Fleisch et al., 2017). These studies, along with a series of systematic reviews (Besharati et al., 2021; Snilstveit et al., 2016; Stern et al., 2021) have provided a coherent and consistent body of evidence to guide program developers and policy makers.

What we know less about is the extent to which these scalable models, and programs that use them system-wide, have had enduring impacts on learning beyond the initial gains in early grade learning outcomes. More specifically, do gains achieved in the second or third year of formal schooling in foundational reading skills translate into better reading comprehension at the end of the basic education phase and into secondary education and beyond? Clearly, this is a critical question. Improvement in early literacy skills may be important in and of itself, but the underlying and often unstated purpose of most early grade interventions is to ensure that children

are able to gain mastery of the knowledge and skills associated with the school curriculum and to build on that success through their schooling career.

This review of the literature begins with a summary of the evidence of persistence—that is, whether and to what degree gains made by schoolchildren in successful early grade learning interventions persist into later years of schooling. The second part of the review is a conceptual mapping of possible mechanisms associated with persistence. Although not the focus of this empirical study, this mapping is critical in helping to interpret the findings and directing questions for future research.

Evidence of persistence

The most prevalent and compelling evidence around persistent effects in education interventions has been generated from studies in high-income countries (predominantly the United States). For example, Bai et al. (2020), in a large-scale regression analysis study, found significant positive impacts of funding early childhood education interventions on reading and math test scores and reductions in special education placement and grade retention. Rather than fading out, these programs appear to have continued to improve children's prospects in the middle years of schooling.

One of the earlier studies of long-term effects of early grade reading interventions was from Project Follow Through (Meyer, 1984). Years after the initial intervention, the study found positive outcomes on ninth-grade reading and mathematics scores, higher high school graduation rates, and higher rates of acceptance at postsecondary educational institutions.[2] Similarly, using

[2] This study was not a randomized evaluation, and included a sample of two schools.

observational data, Borman and Hewes's (2002) study of the long-term effects of three learning intervention models found that on average, students who were enrolled in schools that participated in interventions completed grade 8 at a younger age, had higher learning outcomes, and experienced fewer special education placements and fewer repetitions.

The evidence on persistence of impacts in LMICs is far more limited. Chen et al. (2022) showed that dosage can play a key role in persistent effects. In the case of China, rural children who attended 3 years of preschool performed significantly better than rural children who attended only 2 years, and the 3-year cohort performed as well as their urban counterparts. But we were unable to identify any long-term follow-ups studies randomized evaluations of foundational numeracy or literacy programs.

Hypotheses for persistence

Along with greater concern about whether, and to what degree, interventions have continuing positive effects on student learning, as well as their institutionalization in teacher practice, there is a recognition that interventions' impacts on students differ based on their prior levels of academic achievement (Glewwe et al., 2009). More recently, Kaffenberger and Pritchett (2021) used structured modeling to make a compelling case that for interventions to be successful, they have to be pitched at the correct level. Core to this argument is notion that the profile of students' academic achievement varies from poor to relatively academically strong, but that for interventions to be successful, they need to be calibrated to meet the learning needs at the right level. The learning profile research has altered the nature of the debates about effective interventions to differentiated learning and alignment of interventions to meet the learning levels of the majority of children.

These two emerging trends in the study of effective intervention models have shifted some research toward long-term effectiveness on the one hand and differentiated outcomes on the other. However, little research has put these thoughts together to study not only the persistence of effects, but also how persistence may vary among different groups of students based on their original learning profiles. To pursue this avenue of research would require some theoretical models to help develop new hypotheses.

Pfost et al. (2014) posited three hypotheses associated with the relationship between early reading performance and later reading achievement. The first hypothesis, the Matthew effect, or fan-spread hypothesis, assumes a positive relationship between successful early reading achievement and later reading proficiency. That is, students who read better in the first two grades continue to show higher achievement later in their school careers, due to the mechanisms of bootstrapping and the bidirectional and reciprocal relationships between better reading and greater learning across the curriculum. A second hypothesis, sometimes referred to as the developmental-lag model, suggests that the relationship between initial reading performance and later reading levels could be negative. In other words, students who perform well in the early grades may continue to improve, but at a lower rate than students who initially performed poorly. This outcome would be the case if there were a learning threshold that all students eventually reached. The third hypothesis suggests a stable proficiency difference between high- and low-performing students.

Although these three hypotheses were specifically framed to explore the patterns of achievement of a student population over time, they apply equally to tracking the sustainability of the impact of early grade reading interventions over time. If the early grade intervention is initially successful, then under the first hypothesis, the student population sample that has

benefited should make greater gains over time. With the second hypothesis, even if the intervention benefited students in the short term, the students who had not been part of the intervention would catch up over time, thereby reducing the learning gaps between those who had originally benefited from the intervention and those who had not. This may happen when an intervention teaches skills that children would have learned anyway (Bailey et al., 2017). If an increased focus on these skills takes time away from nurturing other critical academic or socioemotional skills, the net long-term effect of an intervention may even be negative (Durkin et al., 2022). The third hypothesis assumes a stable pattern of performance between those who initially benefited and those who did not (i.e., a sustained impact equal in magnitude to the original).

This study was designed to examine the persistence of impacts from the EGRS I intervention, framed by the three hypotheses described above.

3. Program Description

The EGRS I program was designed as a randomized controlled trial with three intervention arms and one control arm, aimed at improving early grade reading in Setswana language. These interventions were implemented in 230 schools across two districts (Ngaka Modiri Molema and Dr. Kenneth Kaunda) in North West Province of South Africa. The EGRS I intervention began with the teachers of a cohort of students in grade 1 in 2015 and continued with the teachers of the cohort through grade 3 in 2017.

Treatment 1: Training, scripted lessons, graded readers

Treatment 1 (a.k.a. the "training" intervention) provided teachers with training based on scripted

lesson plans that were aligned to the national Curriculum and Assessment Policy Statements for home-language literacy in grades 1–3. The lesson plans incorporated the use of learning support materials, including government-provided workbooks as well as certain additional materials (graded reading booklets, flashcards, posters, etc.), which were provided through EGRS I. The graded reading booklets supported teachers in using group-guided readings and individual work.

Under this treatment arm, teachers were trained on how to use the lesson plans and accompanying materials through centralized training sessions, each lasting 2 days, and occurring twice yearly.

Treatment 2: Reading coaches, scripted lessons, graded readers

Under Treatment 2 (a.k.a. the "coaching" intervention), teachers were given the same set of instructional materials as in the "training" intervention (i.e., scripted lesson plans, graded reading booklets, and other materials). However, the teachers also received a one-day training/orientation at the start of each term, accompanied by ongoing support to teachers consisting of regular (monthly) in-school coaching from specialist "reading coaches" (in lieu of centralized training sessions). In addition to these on-site visits, the program offered occasional needs-based workshops with the coach and a small cluster of nearby Treatment 2 schools.

Treatment 3: Parental involvement

Treatment 3 (a.k.a. the "parental involvement" intervention) was designed to promote parents' involvement to support their children's reading progress. This intervention arm ended prematurely in 2016, after 2 years of implementation showed little impact. The parents of grade 3 students were not part of the intervention and this intervention arm is therefore not

incorporated into the current sustainability study.

This paper focuses only on Treatment 2 ("coaching"), which was found in prior evaluations of EGRS I to have had the largest and most cost-effective impacts on student literacy (Cilliers et al., 2022b). The motivation for highlighting a single intervention arm is that we were interested in the persistence of early gains of a successful program.

4. EGRS I Evaluation Design

As noted in Section 3 above, EGRS I was designed as a randomized controlled trial with three treatment arms and one control arm. Overall, 230 schools were randomly assigned to one of the intervention groups (50 schools each) or to the control group (80 schools). To evaluate the effectiveness of the interventions, the evaluation team selected a random sample of 20 students per school in grade 1 in 2015 (in all 230 schools). (See Cilliers et al., (2020) for details on treatment assignment and random selection for EGRS I.) These same students were tracked longitudinally through four additional waves of data collection, culminating in the final Wave 5 activity in 2021. Students' reading proficiency was assessed at each wave.

Baseline data (i.e., Wave 1) were collected at the start of 2015, when students had just begun grade 1. This stage was followed by Wave 2 (end of 2015, when students had completed grade 1), Wave 3 (end of 2016, when most of the students were in grade 2), and Wave 4 (in 2018, when students were then expected to be in grade 4). The EGRS Wave 4 evaluation showed that the initial impacts of the EGRS I on students' ability to read continued for the coaching and training arms, one year beyond the end of the intervention.

The results presented in this paper are based on Wave 5 data collected in September 2021, when most students were expected to be in grade 7.

Sampling for EGRS I sustainability impact evaluation

After the start of EGRS I, one school closed, leaving 229 of the original 230 schools. These 229 schools formed the proposed sample for data collection for Wave 5. Data were expected to be collected from all students for whom data were collected in grade 1 and who were present at the school on the day of assessment. The original sample was 20 students per school. In 2018, on average, 14.5 students were available per school for Wave 4. The expectation was that this sample would be further reduced to 12 students per school in 2021. The minimum detectable effect size for this reduced sample was estimated to be 0.23 SD.

Data collection and final sample

Data collection for this evaluation occurred September 7–30, 2021. Two main types of data were collected for this study: (1) individual and written student assessment data; and (2) contextual data (including a student asset list). The contextual data provide background information on students, teachers, principals, parents, and schools.

Although the target sample for the overall evaluation was 229 schools, the team was able to successfully collect data from students in only 216 schools.[3] The final sample for the present study consisted of 1,401 students in 78 control schools and 46 coaching schools, with sample sizes per type of instrument displayed in Table 1.

[3] The team was unable to collect data from 13 schools due to a combination of reasons: lack of grade 7 at the schools, uncooperative principals, rotational schedules with grade 7 out during data collection, permanent school closures, temporary school closures, and/or no matched students from the linking sheet in grade 7.

Table 1: Final student sample sizes for EGRS evaluation by instrument

Instrument	Number of observations
Grade 7 oral assessment	1,401
Grade 7 Setswana written assessment	1,312
Grade 7 English written assessment	1,289
Learner asset list	1,369

Data collection instruments

A battery of reading assessments was developed for this evaluation, as shown in Table 2. The English assessment tasks administered for this evaluation were adapted from the Early Grade Reading Assessment (EGRA; Dubeck & Gove, 2015; RTI International, 2016), and from publicly available items from the Progress in International Reading Literacy Study (PIRLS). The assessments were adapted to account for the more advanced reading ability of students in grade 7. The Setswana home-language assessment tasks developed for this evaluation were based primarily on an EGRA approach.

Table 2: List of student assessment subtasks

Setswana-language subtasks	
Setswana Oral Reading Passage 1 *(Ditshwanelo tsa Botho)*	Setswana Written Text Comprehension *(Perele)*
Setswana Oral Reading Text Comprehension 1	Setswana Written Vocabulary
Setswana Oral Reading Passage 2 *(Bopelokgale jwa ga Bonolo)*	

Setswana Oral Reading Text Comprehension 2	
English-language subtasks	
English Oral Reading Passage *(An Unbelievable Night)*	English Written Text Comprehension *(The Life Cycle of Plants)*
English Oral Reading Text Comprehension	English Written Vocabulary

All instruments were piloted and adjusted through several iterations, thereby reducing both the length of the reading passages and the difficulty level of the comprehension questions asked about the passage. The internal consistency of the overall test scale for the entire battery, measured by Cronbach's alpha, was 0.95.[4]

After they completed the reading assessments, students were asked questions about the assets in their household, as a proxy measure for their socioeconomic status; and about their well-being.

Three other questionnaires were administered to collect background information from each school. The principal responded to questions about the school context, including about school facilities, teacher absenteeism, and the school's response to the COVID-19 pandemic. A questionnaire administered to two grade 7 teachers in each school gathered information about teacher experience, challenges in the classroom, and support for teachers. Finally, a fieldworker completed a school functionality observation schedule to assess school environment, note the

[4] When estimated separately for each language, the test scale was $\alpha = 0.92$ for Setswana and $\alpha = 0.89$ for English.

timetable for foundational learning, and produce a subjective assessment of overall school quality.

5. Empirical Strategy

The main evaluation question for this sustainability impact evaluation was: What is the impact of the EGRS I interventions on reading outcomes in Setswana (home language) and English (second language),[5] for students in grade 7 (3 years after the completion of all interventions)? More specifically, we were interested in examining whether the initial impact of the program, which was found to be sustained for the coaching intervention through grade 4, was further sustained until grade 7.

For the analysis of EGRS I impact, we conducted longitudinal analyses among a sample of students who were available during the Wave 1 and Wave 5 data collections. Our initial analyses were unadjusted for any covariates, whereas our final models all included covariates controlling for student- and school-level characteristics that were unrelated to treatment assignment but might have impacted student performance. The aim of creating these final models was to increase precision and statistical power, while accounting for any incidental differences that may have existed between the treatment groups.

Our main estimating equation, Equation 1, is:

$$y_{isb1} = \beta_0 + \beta_1 \text{Coach}_s + X'_{isb0} + \rho_b + \varepsilon_{isb1} \ (1) \ ,$$

[5] Second language: The South African basic education system refers to English as First Additional Language (EFAL).

where y_{isb1} is the outcome indicator of interest for student i in school s and strata b; Coach is the treatment dummy, which is equal to 1 for the coaching treatment arms; $\rho_b$refers to strata fixed effects; X'_{isb0} is a vector of control variables;[6] and ε_{isb1} is the error term clustered at the school level.

We also examined the data for evidence of differential attrition. Attrition would threaten the validity of the evaluation if were to differ between the treatment and control groups (e.g., if struggling students in the intervention group were less likely to drop out of school than those in the control group). We therefore assessed attrition in two stages. In the first stage, we assessed whether students included in the final sample differed in baseline characteristics from those who were included at baseline. In the second stage, we looked for evidence of differential attrition.

The primary outcomes of interest for this evaluation were reading proficiency in Setswana and English. For reading proficiency, we measured a range of reading skills: vocabulary, reading fluency, and reading with comprehension. Because a single composite score across subtasks would not allow for meaningful interpretation, results throughout this study are presented for a simplified composite score (for orally administered tasks only), as well as for individual subtasks. The composite score has been standardized, with a control mean of zero and standard deviation of 1. Therefore, all coefficients for this score can be interpreted as effect sizes of the intervention.

---

[6] Controls included: vector of students' baseline reading scores; student gender, parental education, and assets and books at home; school district and average performance on 2014 Annual National Assessments; and average wealth index and attendance rates in the community surrounding the school.

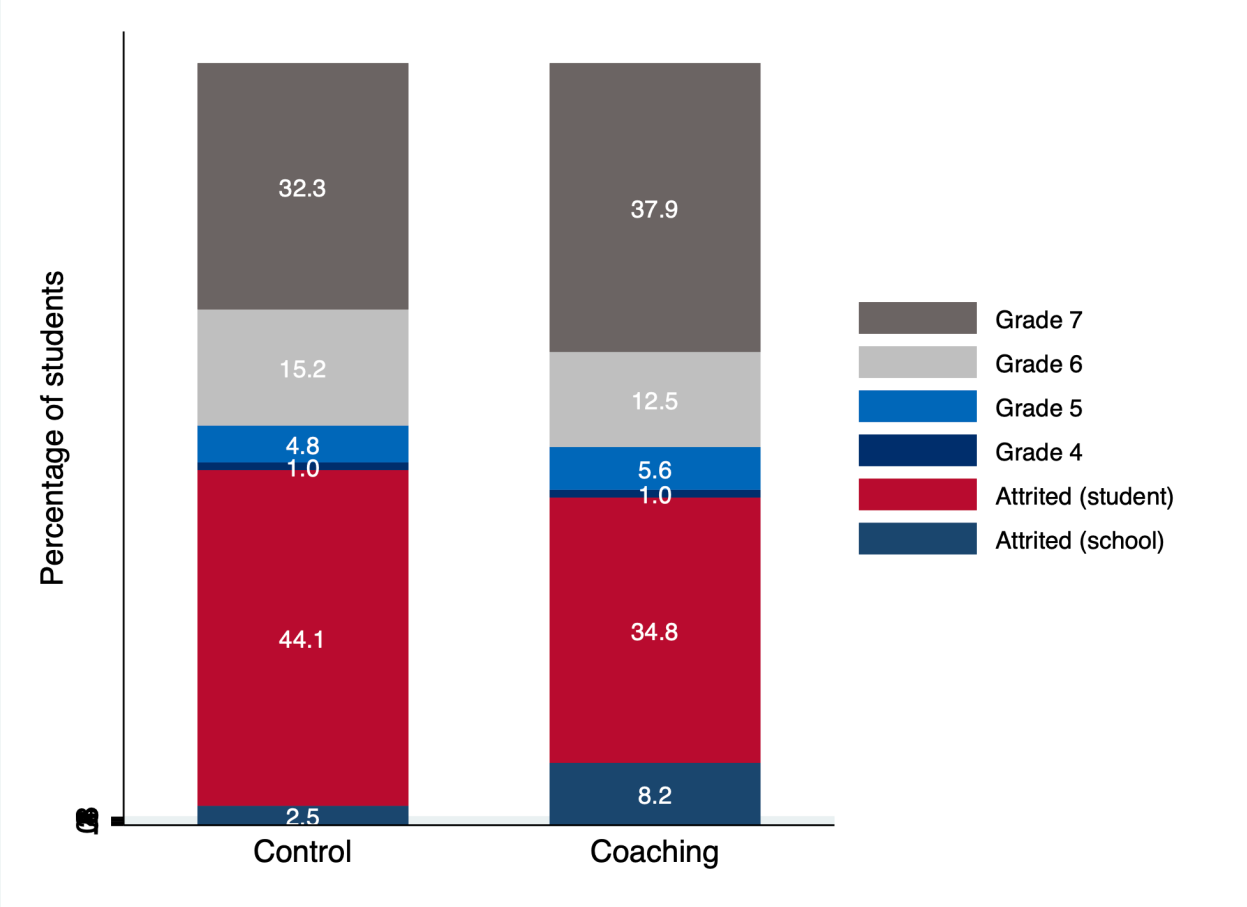
6. Results

This section presents the main findings of the long-term evaluation of EGRS I.

Attrition analysis and balance

Figure 1 shows grade repetition and attrition results from 2021, distinguishing between school-level and student-level attrition. Attrition rates were high: 43 and 47 percent in the treatment and control groups, respectively. A small fraction of this attrition was due to the researchers' inability to access schools that had closed, but the majority consisted of students who could not be assessed on the day the school was visited. Figure 1 further shows that 38 percent of the original sample reached grade 7 by their seventh year of primary school compared to 32 percent in the control.

Figure 1: Proportion of attrition and grade repetition by treatment group



Note. Attrition and grade attainment data are from Wave 5 of data collection, conducted in 2021.

To further investigate the implications of attrition for this study, we ran additional analyses and balance tests, as shown in Table 3. Each column in the table represents a separate regression that was run on attrition, student age, student gender, and student baseline score, respectively.

Column 1 in Table 3 shows that the difference in attrition rates between the evaluation arms was small and not statistically discernible from zero. Moreover, the coefficients on "attrite" in Table 3 (in columns 2, 3, and 4) show that students who were older, male, or performed worse on the baseline reading assessments were more likely to attrite. Because attrition was (slightly)

larger in the control group, this difference could (slightly) bias the results upward. Indeed, the coefficient on "coaching" in Table 3 shows that non-attriters in the treatment group had ~0.1 SD higher levels of learning at baseline compared to non-attriters in the control group (as shown in column 4). But this difference was not statistically significant, and we controlled for baseline learning in all our main regressions. Finally, the coefficient on "attrite × coaching" (i.e., the interaction effect for attrition and coaching) shows that treatment did not change the composition of attriters, which is encouraging for the internal validity of the study.

Table 3: Attrition analysis and balance tests for difference in age, gender, and baseline composite oral reading score

	(1)	(2)	(3)	(4)
VARIABLES	Attrite	Age	Female	Baseline learning
Coaching	–0.037	–0.014	–0.016	0.096
	(0.038)	(0.056)	(0.028)	(0.164)
Attrite		0.151***	–0.064**	–0.147**
		(0.044)	(0.028)	(0.061)
Attrite × coaching		–0.012	0.002	–0.084
		(0.063)	(0.043)	(0.104)
Observations		2,556	2,547	2,556
R-squared		0.009	0.020	0.006
Mean attrition		0.466		

*Notes***.** Each column represents a separate regression, including strata fixed effects. The sample includes all students who were assessed at baseline. *Attrite* is a dummy variable equal to 1 if a student was not assessed in Wave 5, when

nonrepeaters were in grade 7. ***$p < .01$; **$p < .05$; *$p < .1$. Standard errors are clustered at the school level and reported in parentheses.

Lastly, although we did not collect data on the reason for attrition, it is possible to further explore some aspects of attriting students. For example, after we removed schools with no grade 7 data, 1,132 students remained in the sample who were assessed in Wave 1 but who were labeled as attriters in Wave 4. Of those, 893 (79 percent) were also attriters in Wave 5, while 239 (21 percent) had returned to the Wave 5 sample. This fluidity is evidence that at least one-fifth of the attrition in Wave 4 was likely due to absenteeism or another nonpermanent attrition status.

Descriptive statistics

Mean estimates for each subtask are presented in Table 4. It is clear from these estimates that the average scores among students in the coaching group were slightly higher than those in the control group, for nearly all subtasks.

Table 4: Average reading performance by subtask and treatment group (mean scores)

	Control	Coaching	Number of observations
Setswana Text 1 Reading Fluency	56.9	59.0	1,370
Setswana Text 2 Reading Fluency	75.6	80.4	1,368
Setswana Text 1 Comprehension (% correct)	29.7	31.9	1,370
Setswana Text 2 Comprehension (% correct)	41.2	44.3	1,369
Setswana Written Comprehension (% correct)	49.2	51.6	1,308
Setswana Vocabulary (% correct)	62.0	65.8	1,309
English Text Reading Fluency	77.9	79.5	1,366

	Control	Coaching	Number of observations
English Text Comprehension (% correct)	52.2	53.7	1,366
English Written Comprehension (% correct)	40.3	42.1	1,288
English Vocabulary (A)[7] (% correct)	25.7	25.0	1,287

A similar trend appears in Table 5, which shows the percentage of students who were unable to correctly identify a single item on a given subtask (i.e., zero scores). Students in the coaching group showed a lower proportion of zero scores than control students for 7 of the 10 subtasks displayed in Table 5 (with larger reductions in reading comprehension than in text reading—i.e., fluency).

Table 5: Percentage scoring zero, by subtask and treatment group (mean percent scores)

	Control	Coaching	*Number of observations*
Setswana Text 1 Reading	6%	6%	1,370
Setswana Text 2 Reading	10%	10%	1,368
Setswana Text 1 Comprehension	27%	23%	1,370
Setswana Text 2 Comprehension	17%	13%	1,369
Setswana Written Comprehension	10%	8%	1,308
Setswana Vocabulary	6%	5%	1,309
English Text Reading	10%	9%	1,366
English Text Comprehension	13%	11%	1,366
English Written Comprehension	8%	7%	1,288

[7] The vocabulary subtask consisted of three sections but the majority of students only received Section A. Therefore, only this section is included in the analyses throughout the report.

English Vocabulary (A)	18%	25%	1,287

Impact of the coaching intervention on literacy skills

The descriptive analyses suggest that the coaching intervention had small, positive impacts on literacy. Although random assignment to intervention groups theoretically allows for mean comparisons to measure impact, precision can be increased through the use of regression models that account for factors which differed at baseline. Therefore, we used regression models to additionally control for factors that could have impacted scores but that were not affected by the interventions. All regression models throughout this section include controls at the student level (a vector of baseline reading scores, gender, parental education, assets at home, and books at home), school level (location of school and average performance on 2014 Annual National Assessments), and community level (wealth index and attendance rates). These regression models apply separately for all reading outcomes in Tables 6 through 9, with the coefficients on coaching signifying the estimated impact of the intervention relative to the control group.

The first column in Table 6 shows that students' overall proficiency in home language—as measured by the composite score—increased by 0.17 SD, relative to the control (p = .016).[8] The remaining columns show improvements across most of the indicators (which were components in the composite score). Notably, performance on the written comprehension test

[8] The Setswana composite score was created using the first component loadings from a principal component analysis of the four individually administered Setswana reading subtasks.

improved by 4 percentage points (as shown in column 6 in Table 6). This test was not administered in the first three waves of data collection.

Impact estimates for zero scores in Setswana are presented in Table 7. Because the "zero score" outcome variables in each of these models are binary, logistic regression models were estimated. The coaching coefficients in the table therefore represent odds ratios (with numbers smaller than 1 signifying a lower likelihood of scoring zero and numbers larger than 1 signifying a greater likelihood of scoring zero). The intervention reduced the odds of scoring zero in the oral comprehension tests, signaling a specific impact on the lowest-performing students.

We ran the same models for all English outcomes, with results presented in Table 8 and Table 9. The impacts were smaller in magnitude, but remained statistically significant at the 10 percent level. The intervention improved English literacy by 0.13 SD, even though the program targeted only the teaching of home-language literacy. English written comprehension improved by 4 percentage points (as shown in column 4 in Table 8).

Table 6: Ordinary least squares (OLS) regression estimates for impact of intervention on Setswana reading outcomes, with controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	Setswana Composite	Setswana Text 1 Reading	Setswana Text 2 Reading	Setswana Text 1 Comprehen-sion	Setswana Text 2 Comprehen-sion	Setswana Written Com-prehension	Setswana Vocabulary
Coaching	0.167**	4.040**	7.293***	3.306	4.305**	4.144**	4.692**
	(0.0683)	(1.913)	(2.610)	(2.201)	(2.067)	(1.857)	(2.290)
Observations	1,367	1,369	1,367	1,369	1,368	1,282	1,283
R-squared	0.224	0.208	0.165	0.161	0.190	0.181	0.191
Control mean	0	56.86	75.61	29.67	41%	49%	61.95

Notes. Each column represents a separate regression, estimated using Equation 1. Control variables are explained in footnote 6 above. The dependent variables in columns 2 and 3 are the number of words read in a minute; the dependent variables in columns 4 through 6 are the percentage of questions correctly answered in the comprehension test. ***$p < .01$; **$p < .05$; *$p < .1$. Standard errors are clustered at the school level and are reported in parentheses.

Table 7: Logistic regression estimates for impact of intervention on Setswana zero scores, with controls

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Setswana Text 1 Reading	Setswana Text 2 Reading	Setswana Text 1 Comprehension	Setswana Text 2 Comprehension	Setswana Written Comprehension	Setswana Vocabulary
Coaching	0.930	0.576**	0.706**	0.648**	0.860	0.840
	(0.258)	(0.161)	(0.124)	(0.128)	(0.222)	(0.206)
Observations	1,369	1,367	1,369	1,368	1,282	1,283
Control mean	0.0559	0.0681	0.272	0.166	0.0579	0.100

Notes. Each column represents a separate logistic regression, estimated using the same controls as in Table 7. Estimates are odds ratios. ***$p < .01$; **$p < .05$; *$p < .1$. Standard errors are clustered at the school level and reported in parentheses.

Table 8: OLS regression estimates for impact of intervention on English reading outcomes, with controls

	(1)	(2)	(3)	(4)	(5)
VARIABLES	English Composite	English Text Reading	English Text Comprehension	English Written Comprehension	English Vocabulary (A)
Coaching	0.129*	5.277*	3.855*	4.068**	0.895
	(0.0675)	(2.966)	(2.206)	(1.932)	(1.755)
Observations	1,366	1,366	1,366	1,263	1,262
R-squared	0.269	0.237	0.256	0.201	0.195
Control mean	0	77.94	52.19	40.28	25.66

Notes. ***$p < .01$; **$p < .05$; *$p < .1$. See Table 7 for additional notes.

Table 9: Logistic regression estimates for impact of intervention on English zero scores, with controls

	(1)	(2)	(3)	(4)
VARIABLES	English Text Reading	English Text Comprehension	English Written Comprehension	English Vocabulary (A)
Coaching	0.794	0.722	0.799	1.331
	(0.191)	(0.160)	(0.207)	(0.276)
Observations	1,366	1,366	1,263	1,262
Control mean	0.0973	0.125	0.0760	0.183

Notes. ***$p < .01$; **$p < .05$; *$p < .1$. See Table 8 for additional notes.

Sustained effects (compared to prior waves)

In order to measure the degree of fadeout of effect sizes over time, we restricted the sample to students who were assessed in *both* the fourth and fifth waves of data collection, which were conducted in 2018 and 2021, respectively. The odd-numbered columns in Table 10 shows the effect sizes when students were at the end of grade 4, one year after the program ended; the even-numbered columns show the effect sizes 3 years later.

The degree of fadeout varied by the outcome measure considered. Using the composite literacy score as the outcome, an overall fadeout of roughly 45 percent emerged over a period of 3 years (from 0.24 to 0.13, as shown in columns 1 and 2 of Table 10). This implies an annualized depreciation rate of roughly 26 percent. There was fadeout in the effect on the number of words read in the first Setswana reading text, mostly because the control group had caught up with the treated students. In contrast, there was an increase in the effect on the number of words in the Setswana and English written comprehension tests, with the magnitude of the effect size in English increasing ten-fold with time (from 0.42 to 4.55, as shown in columns 9 and 10).

These results are consistent with the idea that investing in foundational skills could eventually have spillover effects in other higher-order cognitive skills.

Table 10. Impact of coaching on a sample of students who were assessed in both Wave 4 and Wave 5

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Composite Score		Setswana Text 1 Reading		Setswana Text 2 Reading		Setswana Written Comprehension		English Written Comprehension	
	Wave 4	Wave 5	Wave 4	Wave 5	Wave 4	Wave 5	Wave 4	Wave 5	Wave 4	Wave 5
Coaching	0.240***	0.132*	7.032***	3.230	6.038**	5.929**	4.309***	4.654**	0.423	4.552***
	(0.0792)	(0.0745)	(2.430)	(2.045)	(2.826)	(2.843)	(1.536)	(1.948)	(1.480)	(1.634)
Observations	1,200	1,200	1,200	1,200	1,200	1,200	1,166	1,166	1,147	1,147
R-squared	0.201	0.215	0.189	0.206	0.158	0.153	0.061	0.187	0.048	0.205
Control mean	0	0	51.21	58.13	59.56	77.70	22.78	49.67	20.17	40.77

Notes. Each column represents a different regression estimated using Equation 1. See footnote 6 for a list of control variables. The sample is restricted to students who were assessed in both Wave 4 and Wave 5. Dependent variable in the first two columns is the index for reading proficiency in the two respective waves of data collection, constructed using principal component analysis and standardized to have a mean of zero and a standard deviation of 1.

Grade attainment: Predicting "on-track" status

Although the EGRS coaching intervention has been shown to have had an impact on learning outcomes, in this section we explore the association between receiving the intervention and progressing through the first six grades of schooling "on track."

Column (1) in Table 11 shows that the coaching intervention increased the probability that a student would reach grade 7 by the 7th year of school. More specifically, being in the coaching group was associated with students having 1.28 times the odds of being in grade 7, as compared with control schools.[9] As shown by column 2, the effect size did not depend on whether or not we included the non-attriters in the sample, although the results were more precisely estimated with the larger sample.

Based on the final model with a full set of controls (shown in column 3), we estimated that students in the coaching group were 1.6 times more likely to be in on-track status 4 years after the intervention ended, as compared with students in control schools.

Table 11. Associations with student-level on-track status

	(1)	(2)	(3)
VARIABLES	On track	On track	On track
Coaching	1.28*	1.29	1.59**
	(0.186)	(0.232)	(0.331)
Sample	All	Non-attriters	Non-attriters
Controls	None	None	Full
Observations	2,556	1,401	1,369

[9] Although there was a significant relationship between baseline learning and on-track status, we found no statistically significant difference in that relationship between coaching and control schools.

	(1)	(2)	(3)
VARIABLES	On track	On track	On track
R-squared	0.010	0.017	0.121
Mean control	0.323	0.605	0.605

Notes. Estimates are odds ratios. ***$p < .01$; **$p < .05$; *$p < .1$.Standard errors are clustered at the school level and reported in parentheses. See footnote 6 for the list of control variables.

7. Limitations

There are several limitations to this study. First, just under 50 percent of the original sample was not available during the Wave 5 data collection. Although attrition is unavoidable in a long-term, longitudinal study, this high level of attrition does reduce external validity. In other words, it is possible that the effect sizes would have been different for the subset of students who were unavailable in 2021. This limitation implies that our findings are no longer generalizable to the original population of participating schools in the North West province.

Our analyses suggest, however, that the attrition in this study was similar in the treatment and control groups. We are therefore confident that the impact estimates in the paper are not the result of differential attrition, based on observable characteristics. However, there may have been unobservable characteristics related to attrition and treatment status that we were unable to account for.

Lastly, our study measured only Setswana and English literacy skills, but a large motivation for prioritizing FLN is that there could be positive spillovers onto higher-order cognitive skills, and eventually improved labor market outcomes. It is too early for us to measure the impact on other outcomes such as college attendance and labor market outcomes. Future research will investigate this.

8. Discussion and Conclusion

This sustainability impact evaluation of EGRS I provided a rare opportunity to track the progress of individual students for a span of nearly 7 years, from grade 1 in 2015 through grade 7 in 2021. More importantly, it allowed us to examine the long-term impacts of a successful early grade reading intervention, approximately 4 years after the intervention's completion. To the best of our knowledge, there are no longer-term follow-ups of randomized evaluations of programs that improved FLN skills in early grades.

We found evidence of a sustained impact of the coaching intervention on Setswana (home language) literacy of 0.17 SD, and on improved grade attainment: roughly 10 percent more students reached grade 7 by the 7th year of primary school. Although there were positive effects for mean performance on nearly all Setswana literacy subtasks, there were also impacts on the reduction of nonreaders. In other words, the program produced sustained impacts for all students on average but also had specific impacts on some of the most at-risk or lowest-performing students.

Additionally, while there was fadeout over time on the effect of the Setswana composite reading score for orally-administered tasks (i.e., when we compared the same group of students who were assessed both in 2018 (one year after the intervention) and 2021 (4 years after the intervention), we found that the effect size decreased by 45 percent over a period of 3 years), this overall fadeout masked important differences on the program's impact by the type of literacy skill.

For example, the fadeout in home-language oral reading fluency (as evinced by the composite score), most likely occurred because the control group caught up with the treatment group over time. This outcome is to be expected, as there are inherent ceiling effects in reading fluency—and it aligns with the developmental-lag hypothesis for persistence that was laid out at the start of this paper.

Perhaps more importantly, when it comes to written reading comprehension, there was a small increase in the magnitude of the impact on Setswana scores and a ten-fold increase in the impact on *English scores*. This finding is particularly salient, because the program did not directly target English literacy, and written comprehension is a more complex skill than oral language fluency or anything that was directly taught as a part of the intervention. Moreover, because all instruction in higher grades in South Africa takes place in English, students with improved English literacy skills will conceivably also perform better in other subjects and have improved labor market outcomes. In other words, these compounding effects show that once students began to have an advantage (i.e., stronger literacy skills by grade 4), they built upon them to gain further advantages down the line. This theory aligns with the first hypothesis for long-term impacts of early grade interventions laid out in the literature review—that an existing divide continues and may expand over time.

These results thus support the idea that investing in FLN could lead to improvements not only in literacy and numeracy but also in other, higher-order cognitive abilities, as well as in non-literacy outcomes (such as improved grade attainment). They also highlight the importance of measuring different types of cognitive skills when investigating long-term outcomes.

Overall, these results are very promising for the potential impacts of well-implemented early grade reading structured pedagogy interventions—and more specifically for the EGRS I coaching intervention approach, particularly in light of the fact that few other studies have shown such a sustained, long-term impact of an early grade reading program.

Ideally, future research will allow us to continue to follow these students as they progress through school, in order to measure the long-term EGRS impact on other academic competencies, as well as high school graduation and college placement rates.

References

Bai, Y., Ladd, H. F., Muschkin, C. G., & Dodge, K. A. (2020). Long-term effects of early childhood programs through eighth grade: Do the effects fade out or grow? *Children and Youth Services Review*, *112*, Article 104890. https://doi.org/10.1016/j.childyouth.2020.104890

Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and Fadeout in the Impacts of Child and Adolescent Interventions. *Journal of Research on Educational Effectiveness*, *10*(1), 7–39. https://doi.org/10.1080/19345747.2016.1232459

Bannerjee, P. A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukherji, S., Shotland, M., & Walton, M. (2016). *Mainstreaming an effective intervention: Evidence from randomized evaluations of "Teaching at the Right Level" in India.* NBER Working Paper No. 22746. U.S. National Bureau of Economic Research. https://doi.org/10.3386/w22746

Besharati, N. A., Fleisch, B., & Tsotsotso, K. (2021). Interventions to improve learner achievement in South Africa: A systematic meta-analysis. Chapter 2 in F. Meringe (Ed.), *Systematic Reviews of Research in Basic Education in South Africa* (pp. 27–67). African Sun Media. https://doi.org/10.18820/9781991201157/02

Borman, G. D., & Hewes, G. M. (2002). The long-term effects and cost-effectiveness of Success for All. *Educational Evaluation and Policy Analysis*, *24*(4), 243–266. https://doi.org/10.3102/01623737024004243

Carter, R. A., Jr., Rice, M., Yang, S., & Jackson, H. Al. (2020). Self-regulated learning in online learning environments: Strategies for remote learning. *Information and Learning Sciences, 121*(5–6), 321–329. https://doi.org/10.1108/ILS-04-2020-0114

Chen, S., Zhao, C., Chen, C., Wu, Z., Snow, C. E., & Lu, M. (2022). Does one more year matter? Dosage effect of the one-village-one-preschool intervention in rural China. *Journal of Research on Educational Effectiveness*, *15*(2), 217–242. https://doi.org/10.1080/19345747.2021.2006383

Cilliers, J., Fleisch, B., Kotze, J., Mohohlwane, M., & Taylor, S. (2022a). The challenge of sustaining effective teaching: Spillovers, fade-out, and the cost-effectiveness of teacher development programs. *Economics of Education Review*, *87*, Article 102215. https://doi.org/10.1016/j.econedurev.2021.102215

Cilliers, J., Fleisch, B., Kotze, J., Mohohlwane, N., Taylor, S., & Thulare, T. (2022b). Can virtual replace in-person coaching? Experimental evidence on teacher professional

development and student learning. *Journal of Development Economics*, *155*, Article 102815. https://doi.org/10.1016/j.jdeveco.2021.102815

Cilliers, J., Fleisch, B., Prinsloo, C., & Taylor, S. (2020). How to improve teaching practice? An experimental comparison of centralized training and in-classroom coaching. *Journal of Human Resources*, *55*(3), 926–962. https://doi.org/10.3368/jhr.55.3.0618-9538r1

Cunha, F., Heckman, J. J., Lochner, L., & Masterov, D. V. (2006). Interpreting the evidence on life cycle skill formation. Chapter 12 in E. Hanushek & F. Welch (Eds.), *Handbook of the Economics of Education*, Vol. 1 (pp. 697–812). North-Holland.

Das, J., Singh, A., & Chang, A. Y. (2022). Test scores and educational opportunities: Panel evidence from five low-and middle-income countries. *Journal of Public Economics, 206,* Article 104570. https://doi.org/10.1016/j.jpubeco.2021.104570

Dubeck, M. M., & Gove, A. (2015). The Early Grade Reading Assessment (EGRA): Its theoretical foundation, purpose, and limitations. *International Journal of Educational Development*, *40*, 315–322. https://doi.org/10.1016/j.ijedudev.2014.11.004

Durkin, K., Lipsey, M. W., Farran, D. C., & Wiesen, S. E. (2022). Effects of a Statewide Pre-Kindergarten Program on Children's Achievement and Behavior Through Sixth Grade. *Developmental Psychology*, *58*(3), 470–484. https://doi.org/10.1037/dev0001301

Evans, D. K., & Hares, S. (2021). *Should governments and donors prioritize investments in foundational literacy and numeracy?* [Working paper]. Center for Global Development. https://www.cgdev.org/publication/should-governments-and-donors-prioritize-investments-foundational-literacy-and-numeracy

Fleisch, B., Taylor, S., Schöer, V., & Mabogoane, T. (2017). Failing to catch up in reading in the middle years: The findings of the impact evaluation of the Reading Catch-Up Programme in South Africa. *International Journal of Educational Development*, *53*, 36–47. https://doi.org/10.1016/j.ijedudev.2016.11.008

Glewwe, P., Kremer, M., & Moulin, S. (2009). Many children left behind? Textbooks and test scores in Kenya. *American Economic Journal: Applied Economics, 1*(1), 112–135. https://doi.org/10.1257/app.1.1.112

Gove, A., Korda Poole, M., & Piper, B. (2017). Designing for scale: Reflections on rolling out reading improvement in Kenya and Liberia. *New Directions for Child and Adolescent Development*, *2017*(155), 77–95. https://doi.org/10.1002/cad.20195

Graham, J., & Kelly, S. (2019). How effective are early grade reading interventions? A review of the evidence. *Educational Research Review*, *27*, 155–175. https://doi.org/10.1016/j.edurev.2019.03.006

Johnson, R. C., & Jackson, C. K. (2019). Reducing inequality through dynamic complementarity: Evidence from Head Start and public school spending. *American Economic Journal: Economic Policy*, *11*(4), 310–349. https://doi.org/10.1257/pol.20180510

Kaffenberger, M., & Pritchett, L. (2021). A structured model of the dynamics of student learning in developing countries, with applications to policy. *International Journal of Educational Development*, *82*, Article 102371. https://doi.org/10.1016/j.ijedudev.2021.102371

Lakhsman, S. (2019). Improving reading and arithmetic outcomes at scale: Teaching at the Right Level (TaRL), Pratham's approach to teaching and learning. *Revue International d'éducation de Sèvres.* https://doi.org/10.4000/ries.7470

Meyer, L. A. (1984). Long-term academic effects of the direct instruction Project Follow Through. *The Elementary School Journal*, *84*(4), 380–394. https://doi.org/10.1086/461371

Pfost, M., Hattie, J., Dörfler,T., & Artelt, C. (2014). Individual differences in reading development: A review of 25 years of empirical research on Matthew effects in reading. *Review of Educational Research, 84*(2), 203–244. https://doi.org/10.3102/0034654313509492

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., DeStefano, J., Kinyanjui, E., & Ong'ele, S. (2018). Scaling up successfully: Lessons from Kenya's Tusome national literacy program. *Journal of Educational Change, 19*(3), 293–321. https://doi.org/10.1007/s10833-018-9325-4

RTI International. (2016). *Early Grade Reading Assessment (EGRA) toolkit, Second Edition.* USAID. https://pdf.usaid.gov/pdf_docs/pa00m4tn.pdf

Snilstveit, B., Stevenson, J., Menon, R., Phillips, D., Gallagher, E., Geleen, M., Jobse, H., Schmidt, T., & Jimenez, E. (2016). *The impact of education programmes on learning and school participation in low-and middle-income countries*. International Initiative for Impact Evaluation (3ie). https://www.3ieimpact.org/sites/default/files/2019-05/srs7-education-report.pdf

Stern, J., Jukes, M., Piper, B., DeStefano, J., Mejia, J., Dubeck, P., Carrol, B., Jordan, R., Gatuyu, C., Nduku, T., Punjabi, M., Harris Van Keuren, C., & Tufail, F. (2021). *Learning at Scale: Interim report.* Prepared for the Center for Global Development and the Bill & Melinda Gates Foundation. RTI International. https://shared.rti.org/content/learning-scale-interim-report#

Suggate, S. P. (2016). A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities*, *49*(1), 77–96. https://doi.org/10.1177/0022219414528540

Valerio, A., Sanchez Puerta, M. L., Tognatta, N. R., & Monroy Taborda, S. (2016). Are there skills payoffs in low- and middle-income countries? Empirical evidence using STEP data. Policy Research Working Paper No. 7879. The World Bank. http://documents.worldbank.org/curated/en/231391478194467051/pdf/WPS7879.pdf