CENTER FOR GLOBAL DEVELOPMENT

Persistence and Emergence of Literacy Skills

LONG-TERM IMPACTS OF AN EFFECTIVE EARLY GRADE READING INTERVENTION IN SOUTH AFRICA

Jonathan M. B. Stern, Matthew C. H. Jukes, Jacobus Cilliers, Brahm Fleisch, Stephen Taylor, and Nompumelelo Mohohlwane

Abstract

Developing countries strongly emphasize enhancing foundational reading skills, with the expectation that such improvements will catalyze the development of new cognitive abilities and lead to improved life outcomes. However, the relationship between early educational gains and later-life achievements remains theoretically complex and evidence is lacking, particularly from low- and middle-income countries (LMICs). This paper examines the long-term impacts of an early grade literacy intervention that was implemented in South Africa in Setswana language from 2015 to 2017. A unique feature of our study is the longitudinal tracking and assessment of the same group of students over a 7-year period, extending up to 4 years post-intervention. We find sustained improvements in Setswana oral reading fluency, indicating the lasting impact of the intervention. Moreover, treated students exhibited enhanced Setswana and English written comprehension skills, signifying the emergence of new skills. Grade progression also improved, potentially reducing dropout rates, and facilitating transitions to secondary education. This study contributes to the limited literature on the long-term causal impacts of early literacy investments in LMICs, suggesting potential for enduring educational benefits and improved later-life outcomes.

KEYWORDS
impact evaluation, early grade reading, structured pedagogy, long-term outcomes

WORKING PAPER 672 • DECEMBER 2023

Persistence and Emergence of Literacy Skills: Long-Term Impacts of an Effective Early Grade Reading Intervention in South Africa

Jonathan M. B. Stern
RTI International (jstern@rti.org)

Matthew C. H. Jukes
RTI International (mjukes@rti.org)

Jacobus Cilliers
Georgetown University (ejc93@georgetown.edu)

Brahm Fleisch
University of Witwatersrand (Brahm.Fleisch@wits.ac.za)

Stephen Taylor
National Department of Basic Education, South Africa (Taylor.S@dbe.gov.za)

Nompumelelo Mohohlwane
National Department of Basic Education, South Africa; Center for Global Development (Mohohlwane.N@dbe.gov.za)

Jonathan M. B. Stern, Matthew C. H. Jukes, Jacobus Cilliers, Brahm Fleisch, Stephen Taylor, and Nompumelelo Mohohlwane. 2023. "Persistence and Emergence of Literacy Skills: Long-Term Impacts of an Effective Early Grade Reading Intervention in South Africa." CGD Working Paper 672. Washington, DC: Center for Global Development. https://www.cgdev.org/publication/persistence-and-emergence-literacy-skills-long-term-impacts-effective-early-grade

CENTER FOR GLOBAL DEVELOPMENT
2055 L Street, NW Fifth Floor
Washington, DC 20036

1 Abbey Gardens
Great College Street
London
SW1P 3SE

www.cgdev.org

Contents

List of Figures

List of Tables

Introduction

Foundational literacy skills have gained significant interest from the international community in recent years, given that reading is considered an essential building block for learning that could lead to the development of other skills over time. Yet by some estimates, more than half of 10-year-olds in low- and middle-income countries (LMICs) cannot read for meaning (World Bank, 2019). These children will struggle to learn in any subject that requires reading, and risk falling behind the curriculum, repeating grades, or dropping out completely. It is thus reasonable to believe that early investments in children's literacy skills could yield large returns later in life.

Despite this enthusiasm, however, it is possible that the initial positive impacts of early investments in literacy skills will fade over time. Students who received a treatment could forget what they learned, or students assigned to a control group could eventually catch up (Bailey et al., 2017). Importantly, even if there is no fadeout in the original skills impacted, new skills might not emerge, if the original gains were not large enough to pass certain thresholds for developing higher-order skills, or if complementary investments are required later in life (Cunha et al., 2006).

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 first phase of the Early Grade Reading Study (EGRS I) program provided teachers with a combination of lesson plans (see additional information under "EGRS I Design" below), teaching materials, and one-on-one coaching, to improve their teaching of early grade home language (Setswana) literacy. The program was implemented over 3 years, targeting consecutive grades each year (grade 1 in 2015, grade 2 in 2016, and grade 3 in 2017) and students were followed longitudinally for the evaluation. Prior studies of EGRS I found a 0.24 *SD* increase in student literacy after 2 years (Cilliers et al., 2020) as well as sustained impacts on some teaching practices (Cilliers et al., 2022a). Our current study tracked and assessed these same students over 7 years, from the start of grade 1 in 2015 to grade 7 in 2021 (with approximately 55 percent of students available at both time points).[1]

We document four main results. First, comparing learning outcomes in years 4 and 7, we found persistence of the initial gains in Setswana oral reading fluency (ORF) and written comprehension. Second, treated students' English written comprehension skills improved, which was not observed in prior years' comparison to the baseline, thus serving as evidence of the emergence of new skills. The ORF passages were leveled literary texts that varied by grade; the written comprehension tasks were consistent across the two time points for each language. Third, whereas all students improved in Setswana ORF because of the intervention, only the top-performing students improved their comprehension skills, suggesting that there may be threshold effects in the emergence of new skills. Fourth, grade progression improved, which could have knock-on effects on reducing dropout rates and transitioning to secondary and tertiary education.

1 The National Department of Basic Education (DBE) in South Africa was the lead for EGRS I. Therefore, this study falls within the broader legislative mandate of the DBE and has received a waiver for approval of this research with human participants.

To the best of our knowledge, this is the first study that examines longer-term causal impacts of a program that improved early grade reading skills in a low- or middle-income country.

Literature review

Global evidence on persistence

This paper contributes to the nascent literature on the causal impacts of early investments in education quality in developing countries on long-term outcomes. 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 (Banerjee et al., 2016; Cilliers et al., 2022b; Evans & Mendez Acosta, 2021; Evans & Popova, 2016; Evans & Yuan, 2022; Graham & Kelly, 2019; Piper et al., 2018; Snilstveit et al., 2016; Stern et al., 2021). But there is less evidence of the enduring impact on learning beyond the initial gains.

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), which has shown mixed results (Jesson & Limbrick, 2014; Suggate, 2016). For example, Borman and Hewes (2002) found that an early learning intervention in the United States improved grade attainment and reduced dropout rates, whereas Hurry and Sylva (2007) found fadeout of the initial gains 3 years after the intervention ended. However, these results have little applicability to LMICs, given the substantial differences in school quality, opportunities, and home environments between developed and developing nations.

Turning to LMICs, there is a tradition of research on long-term impacts of preschool interventions (Gertler et al., 2014; Jukes, 2005) and considerable research on the long-term impacts of conditional and unconditional cash transfers on learning and post-school outcomes (Araujo et al., 2016; Barham et al., 2013; Barrera-Osorio et al., 2019). Recent studies have also exploited long-term panel data to examine correlations between early grade learning, cognitive skills, and later-life outcomes (Danon et al., in press; Das et al., 2022). Most relevant to our study, Das et al. (2022) showed that socioeconomic status is a much stronger determinant of secondary school education outcomes than a child's cognitive skills at age 12.

However, a recent review by Evans and Hares (2021) concluded that causal evidence in LMICs is lacking on the long-term impact of early investments in basic numeracy and literacy skills.

Mechanisms for persistence and fadeout

Because evidence is scarce or absent in LMICs on long-term impacts of early learning interventions, it is important to consider why early interventions in reading skills may or may not translate into long-term gains of a treatment group over the control group, as a basis for understanding

the results of our study. Several frameworks exist to understand the mechanisms involved in reading development specifically (Pfost et al., 2014) and the long-term impact of early childhood interventions more generally (Bailey et al., 2017; Cunha & Heckman, 2007; Pages et al., 2020). Even though our study cannot explicitly test all these mechanisms, they are useful for framing the hypotheses for this study and in interpreting our findings.

Three classes of mechanisms are identified in the literature on persistence. The first class is termed the "skills beget skills" hypothesis (Cunha and Heckman, 2007) or the "Matthew effect" in reading development (Stanovich, 2009). Under this hypothesis, the development of cognitive skills leads to the further development of other cognitive skills. Bailey et al. (2017) identified a trifecta of conditions required for this mechanism to operate. The skills must be malleable—i.e., amenable to change in response to the intervention; they should be the foundational skills upon which more advanced skills are built; and they should be skills that would not develop anyway in the absence of the intervention.

The literacy skills targeted by an early grade program are likely to be foundational: children's ability to match letters to sounds supports their learning to decode written words, which in turn supports their vocabulary learning, which then supports their reading comprehension (LaBerge & Samuels, 1974). Reading comprehension, in turn, can aid learning in all subjects that require comprehension of text. Additionally, a successful early literacy program should benefit both reading skills in the target language and acquisition of a second language. The largest body of evidence on language transfer has noted that language structure knowledge, phonemic awareness, and mastery of decoding in letter and word reading in the first language may spill over to a second language if certain conditions are met (Branum-Martin et al., 2006; Commissaire et al., 2011; Koda & Reddy, 2008). Emerging evidence from South Africa and Kenya on transfer between an indigenous home language and English confirms the same (Mohohlwane et al., in press; Probert & De Vos, 2016).

However, a transition to high-level literacy skills might depend on meeting certain thresholds of proficiency. For example, moving from correctly reading two words in a minute to five words might not translate into comprehension, whereas moving from two words to 20 words might (Hasbrouck & Tindal, 2017; Mohohlwane et al., 2022)

Moreover, a key consideration is how much the intervention improves constrained vs. unconstrained skills (Paris, 2005; Snow & Matthews, 2016). Constrained skills are those literacy skills which are finite, such as learning the 26 letters of the English alphabet. In contrast, unconstrained skills are those generally learned through a combination of deliberate direct instruction, self-directed learning, and gradual exposure through school subjects and outside of school. There are clear ceiling effects in the constrained skills, and it is therefore possible that the short-term effects on targeted constrained skills will fade over time, as unconstrained skills begin to play a larger role in the continued literacy development of children. However, rapid mastery of constrained skills may allow learners in the intervention to focus on unconstrained skills earlier and more effectively.

The second class of mechanisms involves "foot-in-the-door" (Bailey et al., 2017) interventions, which help either to avoid imminent risks (such as dropout) or to seize opportunities (such as scholarships or access to secondary or tertiary education) (Bailey et al., 2020). Even if the skills fade, these opportunities can lead to improved long-term outcomes. In the case of EGRS I, a likely candidate for such a binary event was the avoidance of grade repetition or dropout. Thus, if improved literacy skills resulting from the EGRS I intervention were sufficient to reduce levels of grade repetition, we could expect to see long-term effects on academic achievement. If the initial impact of the program was too small to reduce grade repetition, the gains could be lost over time.

The third class of mechanisms relates to the need for sustained environments (Bailey et al., 2017)—what Cunha and Heckman (2007) referred to as dynamic complementary investments—which are necessary to allow short-term impacts to translate into long-term gains. This is a particularly important concern in areas with high levels of poverty, weak school quality, and a limited literacy environment in students' homes. For example, if teachers are unable to teach the higher-order skills targeted in grade 4–7, or students lack the English vocabulary skills to develop English comprehension, then the initial gains in literacy skills might not translate to future skills.

A fourth mechanism, not explicitly discussed in the literature on persistence, relates to alignment between the curriculum and the learning levels of students. A large body of evidence has documented that curricula in developing countries are often overly ambitious and not tailored to students' level of ability (Kaffenberger & Pritchett, 2021). Students thus learn at a slow pace, because their level of understanding falls outside the range of skills taught in class. If an intervention can bring students into that range, they will receive the complementary investments (i.e., instruction at a level appropriate to the students) to learn at a faster pace moving forward, relative to control-group students who remain behind the curriculum. Alternatively, if an intervention does not take students up to the range of skills taught in the classroom, or accelerate students beyond them, then effects may fade after the intervention as the control group catches up.

To summarize, it is an open question whether the initial gains of EGRS I may persist and lead to the emergence of new skills. On the one hand, there are several reasons why new skills might emerge: literacy skills targeted by EGRS I are likely to be foundational, leading to the development of other skills; improvement in cognitive skills may also lead to the development of noncognitive skills that further enhance the development of literacy skills; and original gains could also compound if the treated students acquired sufficiently high literacy levels to follow the pace of the curriculum, relative to the control students. Additionally, the program could lead to reduced grade repetition and dropout rates, which would put the children on a different path of faster learning and higher likelihood of transitioning to secondary school.

But there are also a number of reasons why fadeout may occur or new skills might not emerge in the South African context. For example, gains in learning experienced by the treatment group might not have crossed thresholds for transition to new skills, or brought them up to the level of the curriculum;

students could have forgotten what they learned if the skills were not sufficiently consolidated; enabling conditions might not have been in place for new skills to development, given the low school quality and lack of literate environment in students' homes; and targeting mainly constrained skills might have allowed control-group students to catch up on unconstrained skills over time.

As previously stated, this study does not test each of these hypotheses directly, but they aid in the interpretation of our results and provide insights into areas for further exploration.

Study overview

South African curriculum and context

The South African early grade home-language curriculum is well-grounded in pedagogical theory on literacy acquisition. The primary objective is for students to build and become fluent in decoding skills. Learners are taught to master phonological awareness, letter naming, letter sounds, phoneme/grapheme relationships, and sounding out of syllables and letter blends, initially in their home language. Combined, these skills are used to decode words. These skills and strategies are taught from grade 1 and continue through grade 3.

Learners in South Africa start schooling with their home language as the language of instruction (i.e., Setswana in the case of EGRS I), while simultaneously learning English as an additional language. They then transition to English as the language of instruction in grade 4. In recognition of this progression, learner assessments for this study were completed for both Setswana and English in each wave (based on grade-level standards for each language).

EGRS I design

Under EGRS I, a structured pedagogy program aimed at improving teaching of early grade home language (Setswana) literacy was implemented in 50 schools across two districts (Ngaka Modiri Molema and Dr. Kenneth Kaunda) in the North West Province of South Africa. The province is relatively homogenous linguistically and one of the poorest provinces in the country, with weak academic performance. The program provided early grade teachers with curriculum-aligned lesson plans, as well as learning aids, such as graded reading booklets, flashcards, and posters.[2] The lesson plans were explicit in how teachers should use these resources in the classroom. The resources were distributed at a one-day training/orientation implemented at the start of each term.

2 The lesson plans developed for EGRS I aimed to help teachers develop weekly and daily routines that would appropriately use South African home-language literacy curriculum instructional methods for grades 1 through 3. The resources included explicit support for teachers to use the full repertoire of literacy strategies. The lesson plans further aimed to foster better integration of new reading materials into the instructional methods prescribed by the curriculum.

In addition to the resources and training, the teachers also received monthly visits (on average, 10 visits in a year) from specialist "reading coaches." During these visits, the coaches observed lessons, provided feedback on how to improve teaching, and demonstrated (or modeled) effective teaching practices. The coaches also offered occasional needs-based afternoon workshops (one to three per year) to teachers in small clusters of schools near to each other. There were three coaches, each serving 16–17 schools.

The program was implemented over 3 years, starting in 2015. In the first year, the grade 1 teachers participated in the program; in the second year, the grade 2 teachers participated; and in the third year, the grade 3 teachers participated. Thus, the same cohort of students (if they did not repeat a grade) benefited from the program over a period of 3 years, although different teachers participated each year.

Evaluation design

The program was evaluated as a clustered randomized controlled trial. The research team identified 130 public primary schools that used Setswana as the main language of instruction and served communities of low socioeconomic status in North West Province. They then created 10 strata of similar schools in terms of school size, socioeconomic status, and previous academic performance, and randomly assigned 50 schools to treatment and 80 to control.[3] In each school, 20 grade 1 students were randomly sampled at the beginning of 2015.

Five rounds of data collection took place over 7 years, tracking students within the sampled schools. 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 were completing grade 1), Wave 3 (end of 2016, when most of the students were in grade 2), Wave 4 (in 2018, when students were then expected to be in grade 4), and Wave 5 (September 2021, when most students were expected to be in grade 7). The results presented in this paper are based on Waves 4 and 5—1 and 4 years respectively after the completion of the program.

Evaluation instruments

A battery of reading assessments was developed for this evaluation and adapted to each wave of data collection, based on students' reading development, language, and grade-level standards. Different literacy skills were assessed across waves in one-on-one assessments for oral language and reading skills, as well as group-administered assessments for the written components. For example, the Wave 1 assessments included receptive vocabulary, phonemic awareness, and letter-sound

3 See Cilliers et al. (2020) for details on treatment assignment and random selection for EGRS I. The EGRS I evaluation also included two other treatment arms: a teacher training program that did not include the coaching, and a parental-involvement arm. Our study focused only on the treatment arm that was found to be most effective and cost-effective, based on results from Waves 3 and 4.

recognition, which are considered the building blocks of decoding skills. Waves 4 and 5 included Setswana individual word recognition and ORF, which measures the number of words read correctly per minute from a connected text. The length and level of difficulty of the words and reading passage were increased in each wave. Waves 4 and 5 also included Setswana and English written comprehension, given that South African students are expected to have developed these skills by grade 4. The written comprehension tests were the same in both Waves 4 and 5, with some minor adjustments in grade 7 (making the passage longer, with additional comprehension questions). This consistency allowed for direct comparability across waves. All instruments were piloted locally and iteratively revised based on the results and participant feedback.

The three learning assessments that were the focus of this study were Setswana ORF, Setswana written comprehension, and English written comprehension. The motivation for using these subtasks was that ORF is the culmination of decoding skills that the curriculum requires learners to master by the end of grade 3. However, learners in South Africa are not fully on track, so we still expect to see growth in these skills between grades 4 and 7. Written comprehension is a core outcome of reading for meaning and a targeted outcome from grade 4 until the end of primary school. This expectation applies for both languages. Taken together, these comparable subtasks allowed us to test for both persistence of impacts on the skills developed in earlier years (ORF) and emergence of new skills (English and Setswana written comprehension) over time.

The Setswana ORF passages for both grades were literary passages (stories), with accompanying comprehension questions. The passage length ranged from 159 words to 261 words, with up to eight questions ranging in difficulty from literal questions to inferential questions. Released passages from the Progress in International Reading Literacy Study (PIRLS) were used for the written comprehension assessment for both languages and grades. We slightly adapted these passages in grade 7 in two ways: increasing the length and adding three more questions.

Internal validity tests and descriptive statistics

Attrition and balance

In Wave 4, attrition rates were 28 percent in each evaluation arm, and half the students of the original sample reached grade 4 by the fourth year of primary school. Attrition rates in Wave 5 were higher: 43 and 47 percent in the treatment and control groups, respectively (as shown in Figure 1). Although there was some school-level attrition in Wave 5 due to the inability to collect data from two control schools and four treatment schools,[4] the majority of attrition was driven by students who could not be assessed on the day the school was visited (and resource constraints made it infeasible to track

4 Data were unavailable from these schools for reasons such as lack of grade 7 at the school, uncooperative principals, or temporary or permanent school closures. These schools had higher attrition rates in Wave 4 (38 vs. 27 percent), and fewer students who had reached grade 4 (33 vs. 51 percent).

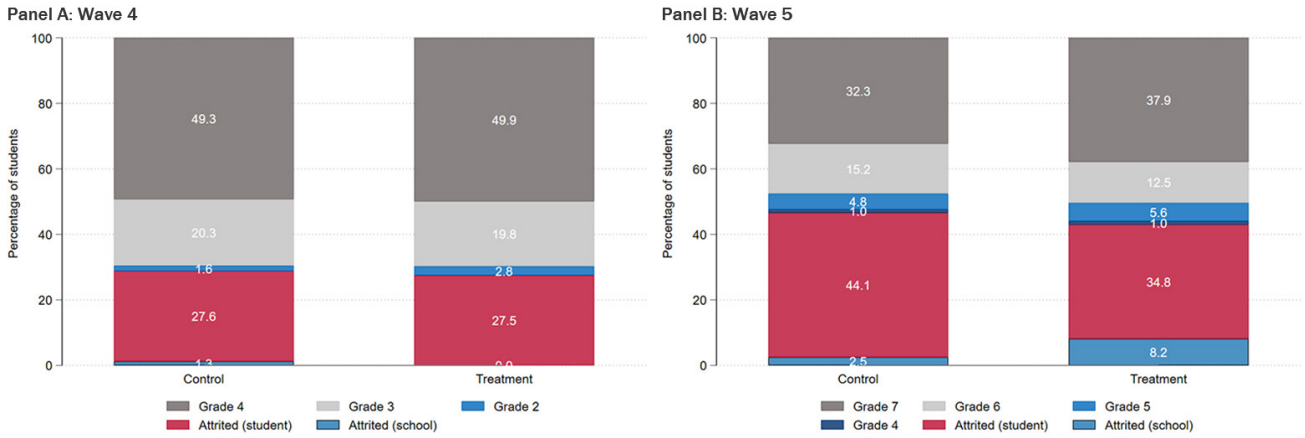
these children and assess them at home). Our final sample consisted of 1,401 students in 78 control and 46 treated schools, with sample sizes per type of instrument displayed in Table 1.

TABLE 1. Final student sample sizes for EGRS evaluation by instrument

Instrument	Number of Observations
Setswana oral reading assessment	1,401
Setswana written assessment	1,312
English written assessment	1,289

Figure 1 further shows that 38 percent of the treatment sample reached grade 7 by their seventh year of primary school, compared to 32 percent in the control. Furthermore, 49 percent of the original sample was assessed in both Waves 4 and 5 (not shown).

FIGURE 1. Proportions of attrition and grade repetition for Wave 4 and Wave 5, by treatment



Note: Attrition and grade attainment data. Panel A: Wave 5 of data collection, conducted in 2021. Panel B: Wave 4 data collection, conducted in 2018.

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, 80 percent of those who attrited in Wave 1 also attrited in Wave 5, while 20 percent had returned to the Wave 5 sample. This fluidity is evidence that at least one-fifth of the attrition in Wave 4 was due to absenteeism or another nonpermanent attrition status.

To further investigate the implications of attrition for this study, we ran additional analyses and balance tests for Wave 5, as shown in Table 2. In column 1, we regress treatment status on attrition status. In the subsequent columns, we regress treatment status, attrition status, and the interaction between the two on the different measures collected at baseline: student age, gender, and baseline score, respectively. All regressions include strata fixed effects, and the standard errors clustered at the school level.

TABLE 2. Wave 5 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**
Treatment	−0.04	−0.01	−0.02	0.10
	(0.04)	(0.06)	(0.03)	(0.16)
Attrite		0.15***	−0.06**	−0.15**
		(0.04)	(0.03)	(0.06)
Attrite × Treatment		−0.01	0.00	−0.08
		(0.06)	(0.04)	(0.10)
Observations		2,556	2,547	2,556
R-squared		0.01	0.02	0.01
Mean attrition	0.47			

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

Column 1 in Table 2 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 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 in the control group. Combined with the overall attrition rates, this information limits the generalizability of findings for the original sample of participating EGRS I schools.

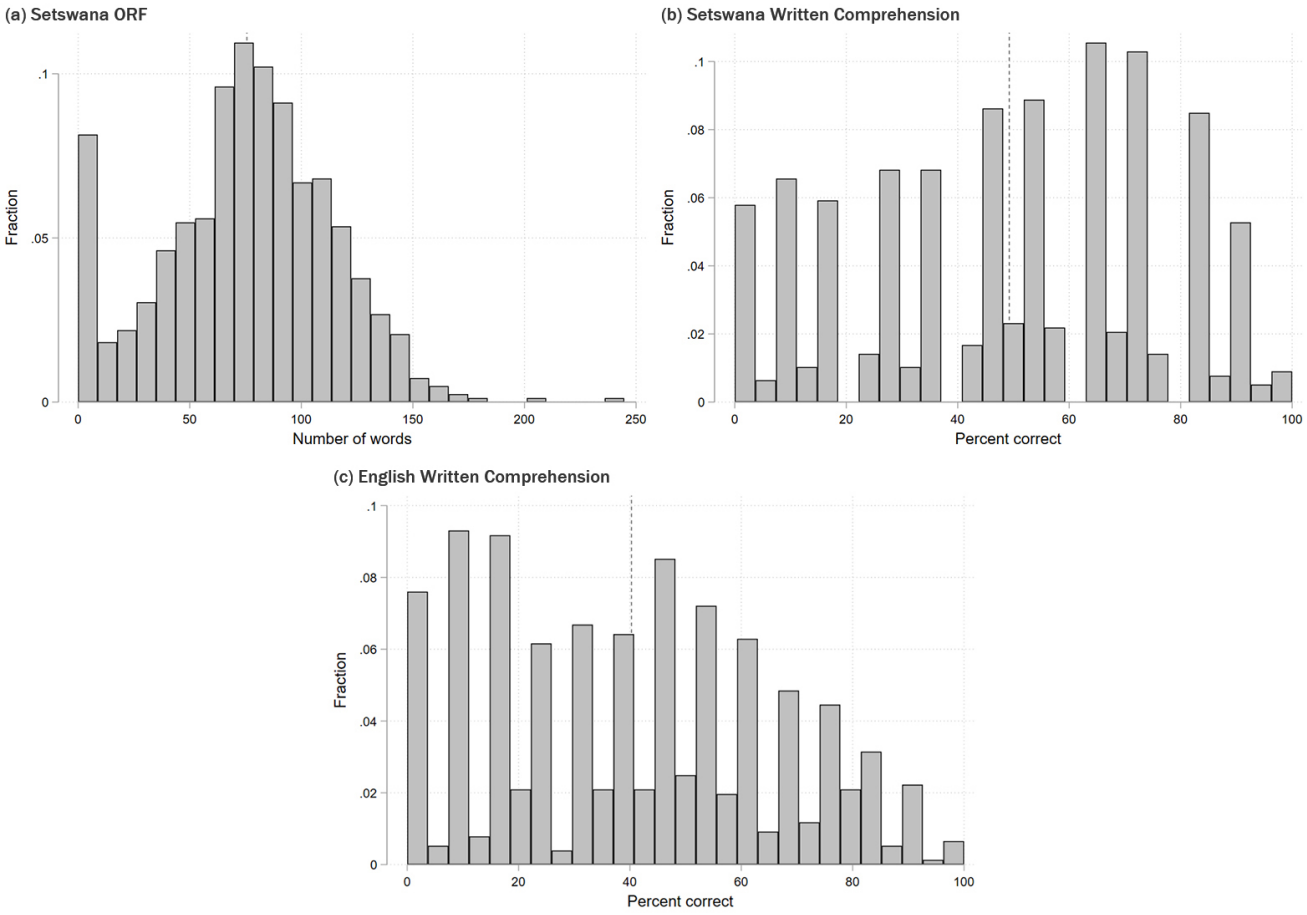
Since attrition is balanced, however, the more important test for the internal validity of the study is the coefficient on "Attrite × Treatment," which tests whether the characteristics of students who attrited were different in the treatment vs. control groups. It is thus encouraging that these coefficients are small and not statistically discernible from zero. Indeed, the coefficient on "Treatment" shows that the sample remained balanced for the sample of non-attriters.

Descriptive statistics

Figure 2 shows histogram plots of the three main learning outcomes of the study. The vertical dotted lines show the mean values. There was substantial variation in performance, with some students scoring 100 percent in the written comprehension tests, but others not correctly identifying a single item on the respective subtasks. This dissimilarity is encouraging from an analytical perspective, since the assessments captured variation in performance across most of the distribution of student ability, allowing us to measure treatment effects for all students. But it is discouraging from the perspective of student learning, because many students were performing extremely poorly. Strikingly, 7 percent of the control students were still unable to read a single word in their home language even after 7 years of school (not shown). The dotted lines show the mean scores.

On average, students were able to read 75 words in Setswana (Figure 2a), and got 49 and 40 percent of the questions right in the Setswana (Figure 2b) and English (Figure 2c) written comprehension tests, respectively.

FIGURE 2. Histogram plots for main outcomes measured at Wave 5



Notes: Histogram plots of the control group for main outcomes measured in Wave 5. Dotted lines show the mean values: 79 words, 49 percent, and 40 percent, respectively.

Results

This section presents the main findings of the long-term evaluation of EGRS I. To summarize, we found evidence of persistence in the main outcomes targeted by the program, emergence of new skills not targeted by the program, and faster progress through the school system.

Our main equation for estimating the impact of the program is:

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

where y_{isb1} is the outcome indicator of interest for student i in school s and strata b; Treatment is the treatment dummy, which is equal to 1 for the treatment arm; ρ_b refers to strata fixed effects; X'_{isb0} is a vector of control variables;[5] and ε_{isb1} is the error term clustered at the school level.

Table 3 shows a positive long-term impact of the program on the four primary outcomes of interest for this evaluation: Setswana ORF, Setswana written comprehension, English written comprehension, and grade attainment. For the first three outcomes, we report the impacts in terms of both raw scores and standard deviations, with only raw score impacts presented for the fourth outcome.

Students in the treatment group read 7.3 (9.6 percent) more correct words per minute, compared to students in the control group. Similarly, performance in Setswana written comprehension was 4.1 percentage points (8 percent) larger in the treatment group relative to the control. The impact on English written comprehension was similar in magnitude: 4.1 percentage points (10 percent). The final row reports the effect sizes in terms of standard deviations. The impacts are moderate in magnitude, ranging between 0.15 and 0.19 *SD*. For comparison, Evans and Yuan (2022) found that the median effect size for evaluations of education programs in LMICs was 0.1 *SD*. Furthermore, students in the treatment group were approximately 9 percentage points (15 percent) more likely to reach grade 7 by their seventh year of school (thus not repeating a grade).

TABLE 3. Regression estimates for impact of intervention on reading outcomes and grade attainment

	(1)	(2)	(3)	(4)
Variables	**Setswana Oral Reading Fluency**	**Setswana Written Comprehension**	**English Written Comprehension**	**Reached Grade 7 (on track)**
Treatment	7.3***	4.1**	4.1**	0.09**
	(2.6)	(1.9)	(1.9)	(0.04)
Observations	1,367	1,282	1,263	1,369
R-squared	0.17	0.18	0.20	0.15
Control mean	75.6	49.2	40.3	0.61
Effect size	0.19***	0.16**	0.15**	–

Notes: Each column represents a separate regression, estimated using Equation 1. Control variables are explained in footnote 5 above. The dependent variable in column 1 is the number of words read correctly in a minute; the dependent variable in columns 2 and 3 is the percentage of questions correctly answered on written reading comprehension subtasks; the dependent variable in column 4 is a binary variable denoting whether or not a student was in grade 7 in Wave 5 (i.e., on track). ***$p < .01$; **$p < .05$; *$p < .1$. Standard errors are clustered at the school level and are reported in parentheses.

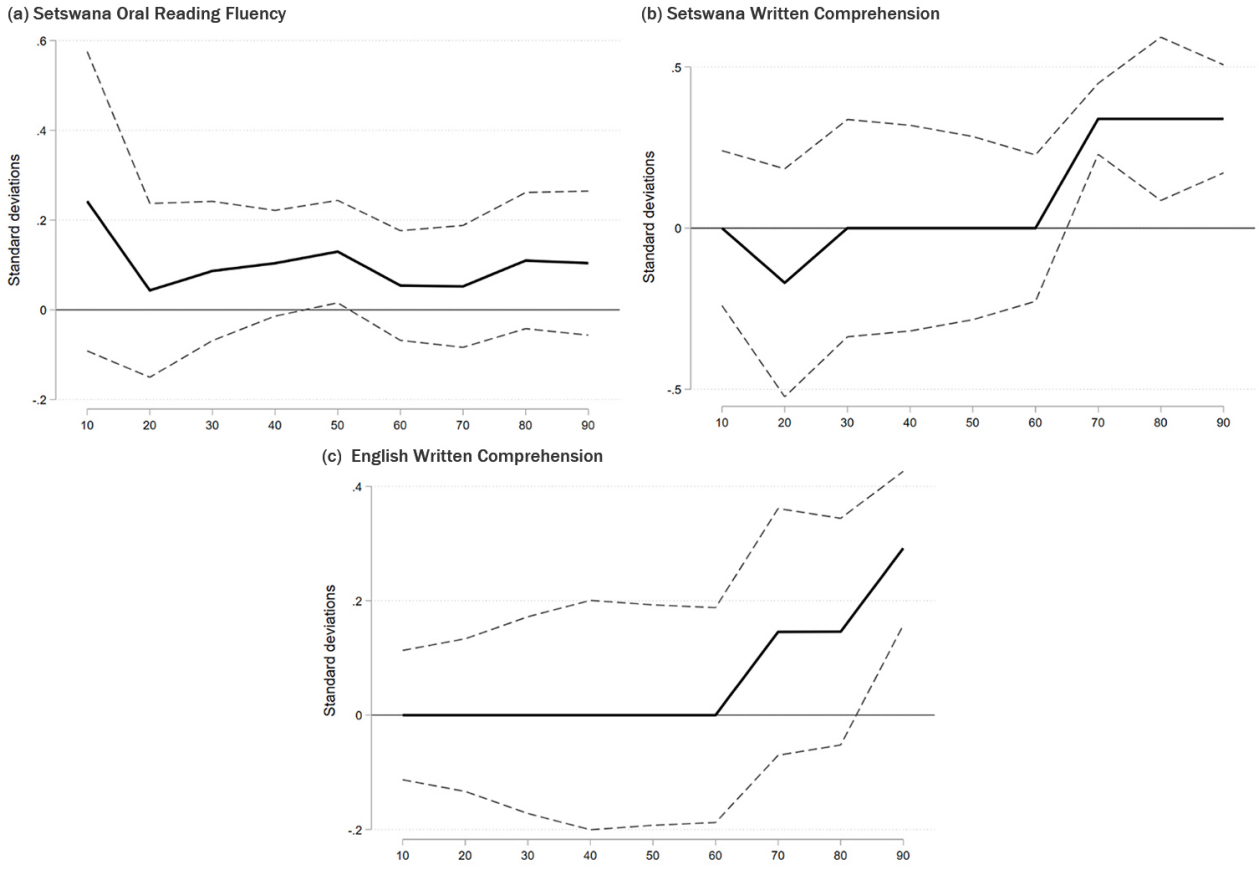
Having established lasting impacts 4 years after the end of the program, we then asked a related question: Which type of students benefited most? It was possible that the (longer-term) effect sizes would be larger for higher-performing students, either because weaker students lacked the enabling conditions that would allow for consolidation of existing skills and emergence of new skills over time,

5 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.

or because they did not pass certain thresholds required to transition to new skills, or because they fell behind the curriculum and did not progress to higher grades. Conversely, it was possible that the effect sizes would be larger for the weaker students if the stronger students in the control group were to catch up over time. This outcome would be especially likely if there were ceiling effects in the measurement indicators, such that many students obtained maximum proficiency levels. These hypothesized relationships could be different for different outcomes, depending on the difficulty of the skill.

Quantile regression analyses for Setswana ORF, as well as Setswana and English written comprehension, are displayed in Figure 3, with the main effect represented by the bolded solid line and the 95 percent confidence interval represented by the corresponding dotted lines. Effect sizes are shown in terms of standard deviations. Students across the distribution of ability seem to have improved their Setswana ORF (with the effect remaining above 0 at all parts of the distribution), even 4 years after the program's termination (Figure 3a). This finding is highly encouraging, since it means even the weaker students continued benefiting from the program. But the pattern is different for Setswana and English written comprehension. For these two outcomes, only the top 30 percent appear to have benefited in Wave 5 from the prior intervention (represented by the positive impacts at the 70th, 80th, and 90th percentiles in Figure 3b and Figure 3c).

FIGURE 3. Quantile regressions by outcome at Wave 5



Sustained effects (compared to prior waves)

Finally, we wanted to quantify the degree of persistence or fadeout over time, by comparing the effect sizes for comparable items and a comparable sample across different rounds of data collection. For this purpose, we restricted the sample to students assessed in both the fourth (2018) and fifth (2021) waves of data collection. The odd-numbered columns in Table 4 show 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.

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

	(1)	(2)	(3)	(4)	(5)	(6)
	Setswana Oral Reading Fluency		Setswana Written Comprehension		English Written Comprehension	
	Wave 4	Wave 5	Wave 4	Wave 5	Wave 4	Wave 5
Treatment	6.0**	5.9**	4.3***	4.7**	0.4	4.6***
	(2.8)	(2.8)	(1.5)	(1.9)	(1.5)	(1.6)
Observations	1,200	1,200	1,166	1,166	1,147	1,147
R-squared	0.16	0.15	0.06	0.19	0.05	0.21
Control mean	59.6	77.7	22.8	49.7	20.2	40.8
Effect size	0.15***	0.15**	0.19***	0.17**	0.019	0.17***

Notes: Each column represents a different regression estimated using Equation 1. See footnote 5 for a list of control variables. The sample is restricted to students who were assessed in both Wave 4 and Wave 5.

The coefficients for both Setswana outcomes are similar in Wave 4 and Wave 5. The effect size on Setswana ORF remained approximately the same (columns 1 and 2), and the effect size on written comprehension either slightly increased or decreased, depending on whether the variable was standardized or not (columns 3 and 4). But there was a substantial increase for English comprehension, with the magnitude of the coefficient increasing roughly tenfold (columns 5 and 6). The program had no impact on English comprehension by grade 4—the positive impact did not emerge until 4 years after the program ended.

This effect constitutes not just persistence but expansion of the impact into a new skill. These results are thus consistent with the idea that investing in foundational skills could eventually have spillover effects into other higher-order cognitive skills.

Limitations

This study had two main limitations. 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 not generalizable to the original population of participating schools in North West Province. Our analyses also suggest, however, that the attrition in this study was similar in the treatment and control groups, so we are confident in the internal validity of the study.

Second, our study measured only Setswana and English literacy skills, but a large motivation for prioritizing foundational literacy is that there could be improved academic overall and improved labor market outcomes. We did not measure whether students performed better in other subjects, nor did we measure transition to secondary school, although it seems like there would have been improvements in these 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 these aspects.

Discussion and conclusion

This sustainability impact evaluation of EGRS I was 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 aimed to improve foundational literacy skills in early grades.

Our research indicated a definitive, prolonged effect on Setswana ORF and written comprehension, alongside advancements in English written comprehension abilities by the seventh year, in contrast to the fourth-year results. These findings imply not only the durability of skills directly imparted through the intervention but also the cultivation of advanced skills. Moreover, while students across various performance levels enhanced their ORF, only the top-tier students demonstrated progress in their English and Setswana written comprehension. Finally, treated students were more likely to pass a grade and thus progress faster through the school system.

Although the data did not allow us to identify the mechanisms by which the EGRS I intervention improved children's reading skills, the finding of a long-term effect does help to rule out some mechanisms outlined in the introduction, namely that the initial impact on children's foundational skills would be too small to translate into long-term investments; that children in the control group would be taught the same skills and catch up with the intervention group; and that lack of good-quality teaching in upper primary school would fail to reinforce the skills of children in the intervention group. Our findings show that these mechanisms were not sufficiently at play to undermine the long-term impact of the program.

As a matter of fact, the observed improvement in English written comprehension and grade progression substantiates two rationales for investing in foundational literacy: (1) skills beget skills, and (2) students with basic literacy are more likely to progress through educational milestones. We did not address this point directly in our study, but it is likely that improved English comprehension will translate to overall improved academic performance across a range of subjects, given the transition to English instruction in the South African curriculum from grade 4. Moreover, children who repeat grades are more likely to drop out of school before completing secondary school. Given the literature on the returns to education, it is thus likely that the treated students will also earn more as adults. Yet, the fact that only the top students improved their comprehension skills as a result of the program suggests possible thresholds or prerequisites for the emergence of new skills.

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 intervention approach, particularly because few other studies have shown such a sustained, long-term impact of an early grade reading program. Future research should continue to follow these students as they progress through school, to measure the long-term EGRS I impact on other academic competencies, high school graduation and college placement rates, and future earnings.

References

Araujo, M. C., Bosch, M., & Schady, N. (2016). *Can cash transfers help households escape an inter-generational poverty trap?* HCEO Working Paper No. 2016-30. Chicago, IL: Human Capital and Economic Opportunity Global Working Group. http://humcap.uchicago.edu/RePEc/hka/wpaper/Araujo_Bosch_Schady_2016_cash-transfers-escape.pdf

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

Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., & Yeager, D. S. (2020). Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. *Psychological Science in the Public Interest*, *21*(2), 55–97. https://doi.org/10.1177/1529100620915848

Banerjee, 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. Washington, DC: U.S. National Bureau of Economic Research. https://doi.org/10.3386/w22746

Barham, T., Macours, K., & Maluccio, J. (2013). Boys' cognitive skill formation and physical growth: Long-term experimental evidence on critical ages for early childhood interventions. *American Economic Review*, *103*(3), 467–471. https://doi.org/10.1257/aer.103.3.467

Barrera-Osorio, F., Linden, L. L., & Saavedra, J. E. (2019). Medium-and long-term educational consequences of alternative conditional cash transfer designs: Experimental evidence from Colombia. *American Economic Journal: Applied Economics*, *11*(3), 54–91. https://doi.org/10.3386/w23275

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

Branum-Martin, L., Mehta, P. D., Fletcher, C. D., Carlson, A., Ortiz, M., Carlo, M., & Francis, D. J. (2006). Bilingual phonological awareness: Multilevel construct validation among Spanish-speaking kindergarteners in transitional bilingual education classrooms. *Journal of Educational Psychology*, *98*(1), 170–181. https://doi.org/10.1037/0022-0663.98.1.170

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

Commissaire, E., Duncan, L. G., & Casalis, S. (2011). Cross-language transfer of orthographic processing skills: A study of French children who learn English at school. *Journal of Research in Reading*, *34*(1), 59–76. https://doi.org/10.1111/j.1467-9817.2010.01473.x

Cunha, F., & Heckman, J. (2007). The technology of skill formation. *American Economic Review*, *97*(2), 31–47. https://doi.org/10.1257/aer.97.2.31

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). Amsterdam: North-Holland.

Danon, A., Das, J., De Barros, A., & Filmer, D. (in press). Cognitive and socioemotional skills in low-income countries: Measurement and associations with schooling and earnings. *Journal of Development Economics*, Article 103132 (online preprint). https://doi.org/10.1016/j.jdeveco.2023.103132

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

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

Evans, D. K., & Mendez Acosta, A. (2021). Education in Africa: What are we learning? *Journal of African Economies*, *30*(1), 13–54. https://doi.org/10.1093/jae/ejaa009

Evans, D. K., & Popova, A. (2016). What really works to improve learning in developing countries? An analysis of divergent findings in systematic reviews. *The World Bank Research Observer*, *31*(2), 242–270. https://doi.org/10.1093/wbro/lkw004

Evans, D. K., & Yuan, F. (2022). How big are effect sizes in international education studies? *Educational Evaluation and Policy Analysis*, *44*(3), 532–540. https://doi.org/10.3102/01623737221079646

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

Gertler, P., Heckman, J., Pinto, R., Zanolini, A., Vermeersch, C., Walker, S., Chang, S., & Grantham-McGregor, S. (2014). Labor market returns to an early childhood stimulation intervention in Jamaica. *Science*, *344*(6187), 998–1001. https://doi.org/10.1126/science.1251178

Hasbrouck, J., & Tindal, G. (2017). *An update to compiled ORF norms*. Technical Report No. 1702. Eugene, OR: Behavioral Research and Teaching, University of Oregon. https://intensiveintervention.org/sites/default/files/TechRpt_1702ORFNorms%20FINAL.pdf

Hurry, J., & Sylva, K. (2007). Long-term outcomes of early reading intervention. *Journal of Research in Reading*, *30*(3), 227–248. https://doi.org/10.1111/j.1467-9817.2007.00338.x

Jesson, R., & Limbrick, L. (2014). Can gains from early literacy interventions be sustained? The case of Reading Recovery. *Journal of Research in Reading*, *37*(1), 102–117. https://doi.org/10.1111/1467-9817.12017

Jukes, M. C. (2005). The long-term impact of preschool health and nutrition on education. *Food and Nutrition Bulletin*, *26*(2, suppl2), S193–S201. https://doi.org/10.1177/15648265050262s210

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

Koda, K., & Reddy, P. (2008). Cross-linguistic transfer in second language reading. *Language Teaching*, *41*(4), 497–508. https://doi.org/10.1017/S0261444808005211

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293–323. https://doi.org/10.1016/0010-0285(74)90015-2

Mohohlwane, N., Wills, G., & Ardington, C. (2022). A review of recent efforts to benchmark early reading skills in south African languages. In N. Spaull & E. Pretorius (Eds.), *Early Grade Reading in South Africa* (Vol. 1, pp. 83–108). Oxford, UK: Oxford University Press.

Mohohlwane, N., Taylor, S., Cilliers, J., & Fleisch, B. (in press). Reading skills transfer best from home language to a second language: Policy lessons from two field experiments in South Africa. *Journal of Research on Educational Effectiveness*.

Pages, R., Lukes, D. J., Bailey, D. H., & Duncan, G. J. (2020). Elusive longer-run impacts of Head Start: Replication within and across cohorts. *Educational Evaluation and Policy Analysis*, *42*(4), 471–492. https://doi.org/10.3102/0162373720948884

Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly*, *40*(2), 184–202. https://doi.org/10.1598/rrq.40.2.3

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., 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

Probert, T., & De Vos, M. (2016). Word recognition strategies amongst isiXhosa/English bilingual learners: The interaction of orthography and language of learning and teaching. *Reading & Writing – Journal of the Reading Association of South Africa*, *7*(1), a84. https://doi.org/10.4102/rw.v7i1.84

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*. London: International Initiative for Impact Evaluation (3ie). https://www.3ieimpact.org/sites/default/files/2019-05/srs7-education-report.pdf

Snow, C. E., & Matthews, T. J. (2016). Reading and language in the early grades. *The Future of Children*, *26*(2), 57–74. https://doi.org/10.1353/foc.2016.0012

Stanovich, K. E. (2009). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Journal of Education*, *189*(1–2), 23–55. https://doi.org/10.1177/0022057409189001-204

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. Research Triangle Park, NC: 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

World Bank. (2019). *Ending learning poverty: What will it take?* Washington, DC: World Bank. http://hdl.handle.net/10986/32553