I Z A Institute
of Labor Economics

DISCUSSION PAPER SERIES

The Effect of Teacher Training and Community Literacy Programming on Teacher and Student Outcome

Feliciano Chimbutane ®
Naureen Karachiwalla ®
Catalina Herrera-Almanza ®
Jessica Leight ®
Carlos Lauchande ®

The Effect of Teacher Training and Community Literacy Programming on Teacher and Student Outcome

Feliciano Chimbutane
Universidade Eduardo Mondlane

Catalina Herrera-Almanza
University of Illinois Urbana-Champaign and IZA

Naureen Karachiwalla
International Food Policy Research Institute

Jessica Leight
International Food Policy Research Institute

Carlos Lauchande
Universidade Pedagogica Maputo

ABSTRACT

The Effect of Teacher Training and Community Literacy Programming on Teacher and Student Outcome*

Motivated by extremely low levels of basic reading skills in sub-Saharan Africa, we experimentally evaluate two interventions designed to enhance students' early-grade literacy performance in rural Mozambique: a relatively light-touch teacher training in early-grade literacy along with the provision of pedagogical materials, and reacher training and materials in conjunction with community-level reading camps. Using data from 1,596 third graders in 160 rural public primary schools, we find no evidence that either intervention improved teachers' pedagogical knowledge or practices or student or teacher attendance following two years of implementation. There are some weak positive effects on student reading as measured by a literacy assessment, primarily observed in a shift away from scores of zero, and these effects are consistent across arms. Our findings are consistent with the growing consensus that more intensive school- and/or community-based interventions are required to meaningfully improve learning.

Corresponding author:
Catalina Herrera-Almanza
University of Illinois
321 Mumford Hall
Urbana-Champaign, MA 02115
USA

E-mail: cataher@illinois.edu

1 Introduction

In sub-Saharan Africa, a strikingly large number of primary school children do not master basic literacy skills in primary school, an oft-cited "learning crisis" that has shadowed the rapid expansion of primary school enrollment in recent decades and that showed further deterioration after the school closures linked to the COVID-19 outbreak (World Bank, 2017; Bold et al., 2017). According to the United Nations, 88% of children of primary or lower secondary age in the region are not proficient in reading (United Nations, 2019). Given that enrollment in lower secondary school is also notably low in sub-Saharan Africa in comparison to other low-income regions (Evans and Mendez Acosta, 2021), stubbornly poor learning outcomes in primary school also constitute a major barrier to further expanding enrollment at higher grade levels.

This paper reports on a randomized controlled trial analyzing two strategies to increase early-grade literacy in rural Mozambique, a country characterized by one of the lowest rates of human capital accumulation in the world; fewer than half of Mozambican children complete primary school (Mambo et al., 2019), and in Nampula province, the site of our study, only 3.2% of grade three students demonstrate grade level reading skills (INDE/MINEDH, 2017). A sample of 160 rural public primary schools was randomly assigned to receive status quo education programming, a short pedagogical training for teachers in conjunction with the provision of classroom materials, or training and classroom materials in conjunction with community-led reading camps designed to enhance children's engagement in reading outside of school.

A large literature has suggested that teacher training can be an effective strategy to improve learning (Arancibia et al., 2016), especially given widespread evidence that teachers' typical pedagogical techniques are often unsuited to classrooms including students with a large variation in initial skill levels (Muralidharan, 2017). However, evidence presented in Arancibia et al. (2016) suggests that successful training programs are relatively intense and subject-specific, accompanied by learning materials, and supplemented by follow-up training, advisory visits, and incentives for teachers. As discussed in more detail below, the teacher training program evaluated here was in fact generally light and characterized by limited follow-up. Community-level reading interventions are a more novel and potentially promising complementary intervention that can increase students' and parents' investment in the development of literacy skills (Banerjee et al., 2010; Björkman Nyqvist and Guariso, 2021).

This trial was conducted in collaboration with World Vision (WV) in the context

of the multi-year educational program Educating Children Together (ECT3) in rural Nampula province, Mozambique.[1] The pedagogical training in literacy methods (Unlock Literacy or UL) focused on five core early-grade reading skills and targeted teachers in grades one through three, in conjunction with the provision of supplementary teaching materials. The training was relatively light touch (only two days with a planned one day follow-up by supervisory staff). The second treatment arm supplemented this training with the creation of reading camps, led by a volunteer from the community who was mandated to conduct weekly camp sessions with students outside of school centering around literacy-focused activities, games, and stories. The volunteer was supported by a designated teacher at the school.

The evaluation employs a repeated cross-section design, surveying a sample of approximately 1,600 students who had completed grade two education at baseline (2021) and endline (2023), including the administration of the Early Grade Reading Assessment (EGRA). Both survey rounds also included surveys with deputy school principals and grade four (baseline) or three (endline) reading teachers. The pre-specified primary outcomes of interest include teachers' pedagogical knowledge, their pedagogical practices (as observed and scored by a trained enumerator using a project-specific rubric developed to reflect the training curriculum), and student literacy as measured by the EGRA score, comprised of five reading sub-tasks; the pre-specified secondary outcomes of interest include student dropout, student absenteeism, teacher absenteeism, and an index of school management practices.

Our primary findings suggest that the teacher training itself — already designed to be relatively light-touch — was not implemented with particularly high fidelity. On average, fewer than half of the teachers assigned to treatment schools report exposure to Unlock Literacy training (and conditional on participation, they report attending around two of the three days of training, on average). This could be consistent with the hypothesis that some teachers simply declined to attend or were absent from training, with high turnover of teachers such that trained teachers are no longer present in treated schools at endline, or with limited salience of a brief training that teachers do not recall in the follow-up survey. Qualitative evidence suggests that teachers were unhappy with the absence of any incentive provided for training participation, consistent with the hypothesis that they declined to attend.

---

[1] More specifically, ECT is funded through the US Department of Agriculture (USDA) McGovern-Dole Food for Education (FFE) program, and also entails the provision of school meals to all schools in the sample; given that there is no experimental variation in this dimension of programming, we generally do not discuss it in the paper.

The provision of learning materials was implemented with higher fidelity, with 70% of schools receiving materials, primarily in Portuguese.

The reading camps, by contrast, seem to have been implemented with intermediate fidelity in the target arm : all schools in the second treatment arm reported having a functioning reading camp. Student self-reported participation is less impressive, however, as only 39% of students in the target arm report membership in a reading camp (but of those, 91% attend the reading camp once or more per week). There is also some, though not high, contamination of both interventions in the control arm, where 13% of teachers report receiving training and 8% of students report membership in a reading camp. This could reflect a misunderstanding of the question posed, teacher churning, or participation in other related interventions (particularly, a closely related intervention that included teacher training and reading camps implemented by Save the Children immediately prior to the roll out of Unlock Literacy for which 63% of teachers had attended training at baseline).

Given these relatively low penetration rates for the intervention, it is perhaps unsurprising that the estimated effects on the key outcomes of interest are generally minimal. For primary outcomes, there is no effect of the training on teachers' pedagogical knowledge or practice, suggesting that teachers did not particularly benefit from the relatively light training in shifting their classroom management. The average effects on students' EGRA scores are similarly small: for the continuous scores in each sub-task (letter recognition, familiar word reading, oral vocabulary, listening comprehension, and reading comprehension), the training only arm has a positive and significant effect on a single sub-task, while the training and reading camp arm has a positive and significant effect on three sub-tasks; however, the analysis uniformly fails to reject hypothesis that the (weak positive) effects are consistent across treatment arms. The magnitude of the positive effect is around 0.145 standard deviations (for letter identification, for any treatment), and the p-value for a joint test of any treatment effect (across both arms) on any sub-task is $p = 0.071$, consistent with the hypothesis of weak positive effects. A related but exploratory analysis suggests that these effects are concentrated at the bottom of the learning distribution, as there are a range of positive effects on the probability of non-zero scores on specific EGRA sub-tasks.

Moving on to secondary outcomes, we again observe consistent null effects for dropout, absenteeism (attendance), and school management practices. There is no evidence that the interventions catalyzed any wider shift in students' or teachers'

3

engagement in school. An analysis of heterogeneous effects suggests there is relatively little evidence of heterogeneity according to baseline school characteristics (including baseline test scores) or student gender.

Overall, we interpret these findings as suggesting that a relatively light-touch teacher training intervention was in fact insufficient to generate any increase in teacher human capital or shift pedagogical or management practices within the school environment; accordingly, the school-focused intervention was primarily a simple provision of pedagogical materials. Consistent with past literature around the effects of school input interventions (Conn, 2017; Das et al., 2013; Glewwe et al., 2004, 2009), the effects of this intervention are weakly positive, though perhaps minimal. The observed effects may be somewhat enhanced by the implementation of reading camps (again, an intervention implemented with intermediate fidelity), but there is little robust evidence of larger effects in this arm.

Our paper contributes to existing literature analyzing educational and pedagogical interventions in sub-Saharan Africa as well as other low- and middle-income countries; a large literature here is summarized in Evans and Popova (2016) and Evans and Mendez Acosta (2021), building on a series of earlier reviews (Kremer et al., 2013; Glewwe et al., 2011; Murnane and Ganimian, 2014; Krishnaratne and White, 2013). More specifically, we make two contributions. First, we assess the effects of a relatively lighter-touch training model in enhancing literacy, an exploration with important implications for cost-effectiveness; however, we find that this lighter model does not seem to be effective in shifting teacher behavior or student achievement, a finding parallel to the evidence in Kerwin and Thornton (2021) that a reduced-cost pedagogical model was unsuccessful in Uganda. Popova et al. (2022) find that teacher trainings are successful when they involve several days of face-to-face training in a row and when there is substantial time for teachers to practice what they have learned. Second, we evaluate the effects of a joint in-school and out-of-school intervention, evidence that is relatively novel in sub-Saharan Africa; here, we find that positive effects observed in India (described in more detail below) are not replicated. One potential interpretation of this finding is that the overall level of community human capital is too low in contexts such as rural Mozambique for community-based interventions to be effective. Another interpretation is that the camps were simply not intensive enough in implementation, as fewer than half the students in the target arm reported attending the camp weekly (as was intended).

With respect to teacher training specifically, a growing body of evidence sug-

gests that training centering around "teaching at the right level" — a strategy that groups students by ability levels rather than by grade levels — can be extremely successful in enhancing learning, and literacy in particular (Banerjee et al., 2017; Björkman Nyqvist and Guariso, 2021). Tracking, the use of computer-aided learning, or the provision of additional teachers to conduct remedial lessons are frequent strategies employed to enable better tailoring of pedagogy to students' actual, rather than expected, level of learning (Duflo et al., 2011, 2020; Muralidharan et al., 2019). Insofar as Unlock Literacy as evaluated in this trial did not include a focus on teaching at the right level or meaningful follow-up or reinforcement in the skills provided, its failure to significantly enhance student learning is, unfortunately, consistent with the existing literature.

Our findings are also consistent with a growing literature documenting that multi-faceted educational interventions can have large effects, and much larger effects than more focused programs. In Uganda, an intensive early-grade literacy intervention providing training, complementary inputs, and ongoing support had large positive effects that were substantially reduced in a lower-cost model (Kerwin and Thornton, 2021).[2] In Tanzania, there was evidence of substantial complementarities between cash grants to schools and teacher incentives (Mbiti et al., 2019), and related evidence showed complementarity of ex-ante presence of school materials and teacher incentives in Uganda (Gilligan et al., 2022). In Guinea-Bissau, complex supply-based interventions (providing para-teachers in conjunction with teacher coaching, or running schools in lieu of the government) generated dramatic increases in learning levels (Fazzio et al., 2021; Eble et al., 2021). Again, the evidence in this paper showing that a relatively light-touch intervention was relatively ineffective is aligned with previous evidence that very intensive interventions are required to shift learning.

Finally, our paper also links to a literature around the role of community-based educational interventions, including tutoring and interventions designed to enhance parental engagement in education. An evaluation of two interventions in Mexico (providing financial support to parents' associations, and providing information about how to support children's learning) demonstrated some positive effects on parental and student behavior, but no effects on learning (Barrera-Osorio et al., 2020). In Angola, an intervention combining parents' meetings and an informational campaign positively impacted school management and a range of attitudinal variables, but

---

[2]There was also separate evidence, however, that the effects were widely heterogeneous (Buhl-Wiggers et al., 2022).

effects on learning were not assessed (Di Maro et al., 2024). In Kenya, cross-age tutoring had weak positive effects in mathematics but not in English (Romero et al., 2022), but tutoring generated large positive effects in China (Behrman et al., 2024). In Pakistan, school report cards reporting their child's school and other schools' test scores provided to parents modestly increased learning after one year (Andrabi et al., 2017). In India, parental or volunteer-run reading camps were effective in targeting students requiring remedial reading instruction (Banerjee et al., 2010) and also in implementing teaching at the right level (Banerjee et al., 2016).

Our paper is most closely related to another recent contribution in India that finds that the combination of pedagogical training (in teaching at the right level) and out-of-school study groups (similar to reading camps) increased test scores in mathematics and language, but neither intervention was effective separately (Björkman Nyqvist and Guariso, 2021). Here, we find that even the effect of the conjunction of the two interventions was quite small. One interpretation is that since the teacher training intervention itself was ineffective (perhaps due to the absence of key design elements), the required complementarity for a community-level intervention was absent. A second interpretation is that the level of human capital possessed by the community members who are supporting these supplementary community-level interventions in sub-Saharan Africa is insufficient; adult literacy is low particularly for women, only half of whom can read (Bank, 2024), and only 20% of reading camp leaders had completed high school. Our findings suggest that further exploration of the effectiveness of tutoring and other educational interventions conducted outside of school in particularly resource-poor areas characterized by very low educational attainment ex-ante would be a useful contribution to this evidence base.

The remainder of the paper proceeds as follows. Section 2 describes the intervention and the experimental design. Section 3 outlines our empirical strategy. We then present our results in Section 4. Finally, Section 5 concludes.

2 Experimental design

2.1 Intervention

Since 2013, WV has delivered a comprehensive school meals program to all rural primary schools in two districts in Nampula province, Muecate and Nacarôa, and this trial centers around two additional educational interventions that were deployed in the same target schools. The first, T1, is an early grade literacy program called

Unlock Literacy (UL) that entails teacher training in pedagogical techniques for early-grade literacy and the provision of learning materials; the second, T2, is the addition of community-level reading camps.

More specifically, UL training (delivered to all teachers in grades one to three in the target schools) focuses on five core early-grade reading skills: letter recognition, familiar word reading, oral vocabulary, listening comprehension, and reading comprehension. The training builds on the government curriculum and incorporates supplementary print materials (in Portuguese as well as Emakhuwa, the local language), and it also includes an emphasis on creating a print-rich environment in classrooms and ensuring student motivation. The initial teacher training was conducted in two days and there was a one-day refresher training; the intervention plan also called for follow-up guidance and teacher exchanges through workshops convened in local school districts.[3] UL also entails some limited engagement with communities and school stakeholders (mobilization around the importance of schooling and attendance, and training of school directors around the importance of monitoring and ensuring teacher attendance).

The second intervention (T2) evaluated in this trial is reading camps, in which a volunteer in the community meets with children weekly outside of school (usually on the weekends). Activities implemented in reading camps also focus on the five core reading skills and include the use of songs, stories read aloud, and games; the camp leaders are provided with training and learning materials (books, posters, etc.) Each reading camp leader works in collaboration with a teacher from the target school appointed as a focal point to assist in recruiting and encouraging students to attend the camps and encouraging their parents to allow them to join and attend.

2.2 Randomization

The overall experimental design was simple. All 160 rural primary schools in the target districts entered the sample and were randomly assigned to three arms: schools in the control arm received status quo programming, schools in T1 received teacher training and pedagogical materials, and schools in T2 received teacher training and materials in conjunction with reading camps.[4] The randomization was stratified by

---

[3]The relevant school district in Mozambique is the Zona de Influencia Pedagógica, or ZIP, whereby a cluster of schools meet regularly to discuss pedagogical topics, plan lessons, and receive guidance and training.

[4]These 160 schools were essentially the full sample of public primary schools in the two districts, though a small number of newly created schools were not yet served by WV. All schools received

Figure 1. Program timeline
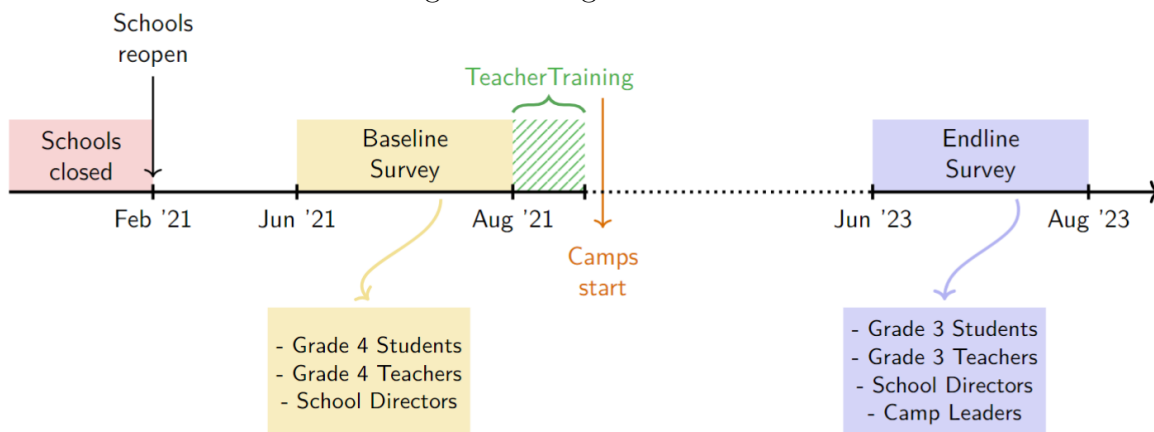
2019 student enrollment (above/below median) and number of teachers (above/below median), creating four strata. Ultimately, 54 schools were assigned to the control arm, 56 to T1, and 50 to T2.

Figure 1 describes the timeline. The government of Mozambique re-opened schools following a prolonged school closure linked to the COVID-19 pandemic in February 2021, and a baseline survey of the sample schools was conducted between June and August 2021. Randomization was conducted contemporaneously with the baseline, and the intervention (beginning with a two-day teacher training) was launched immediately following the survey. We collected the follow-up survey data between June and August 2023, following two years of implementation.

2.3 Data and sampling

This evaluation uses a repeated cross-section design drawing on two surveys of school staff and students conducted in 2021 and 2023, as noted above. The target student cohort was grade three students; however, given that the baseline data was collected following a nearly year-long interruption in schooling due to COVID-19, the baseline survey was conducted with grade four students. Grade four students graduated from grade two, but received only minimal instruction in grade three in the 2020 school year due to COVID-related closures; accordingly, their level of literacy should be an appropriate proxy for incoming grade three students' literacy in future cohorts. The follow-up survey round was conducted with grade three students.

Both survey rounds included surveys administered to three key staff members:

_____

school meals.

the deputy school director, the teacher of the target student cohort (grade four at baseline, grade three at follow-up), and the reading camp leader (follow-up only). The deputy school director survey collected data on characteristics of the school, characteristics of the deputy director, overall enrollment, and teacher attendance. The teacher survey included questions on characteristics of the teacher and class, and (at follow-up) participation in training and receipt of classroom materials. At follow-up, the teacher survey also included a classroom observation tool in which enumerators observed the teacher's classroom practices (for roughly 30–60 minutes) and recorded the pedagogical practices employed, as well as a short knowledge module (18 questions) directly administered to teachers to assess recall of material addressed in the training. At follow-up only, surveys were conducted with the reading camp leader to collect information around the characteristics of the camp leader, their methods and experiences with reading camps, and student participation.[5]

The student survey was administered to a random subsample of 10 grade four (three) students at baseline (endline) per school who were present on the day of the survey and for whom either a parent or the principal had provided consent (students were also requested to provide verbal assent to participate).[6] The student survey included information about child and household characteristics and experiences in the classroom; in addition, the EGRA was administered one on one to each sample child in Portuguese to measure reading proficiency. The version of EGRA employed is a 15-minute assessment that measures a student's aptitude in the five crucial reading subdomains previously described (Dubeck and Gove, 2015). More details about the sub-tasks are provided in Appendix A2. We also assessed Cronbach's alpha as a measure of internal reliability for the EGRA, which was at an intermediate level (0.67), though three out of five sub-tasks showed internal reliability above 0.9.

All data collection was conducted digitally using Tangerine, an application that also has modules available for recording the EGRA. Ethical approval was provided by the Institutional Review Board at IFPRI and the Comité Nacional de Bioética para a Saúde de Moçambique. More details about field procedures, including consent, are provided in Appendix Section A3.

---

[5]At baseline, reading camp leaders had not yet been recruited and thus could not be surveyed.

[6]In a small number of cases, fewer than 10 children were surveyed due to low attendance on the survey day.

2.4 Primary and secondary outcomes

All variables of interest were pre-specified in a detailed analysis plan registered on the AEA trial registry (AEARCTR-0007978). The primary outcome of interest is student performance on the EGRA on the five sub-tasks separately described above. The EGRA outcomes are defined using two methods. The first is the raw score on a given sub-task (number of questions answered correctly) and the second is an indicator variable for whether a student scored positively on the sub-task (i.e., did not score zero). Other primary outcomes of interest include teacher pedagogical knowledge (as measured in a knowledge module administered to teachers) and teacher pedagogical practices (as assessed in the classroom observation tool). Pedagogical knowledge is scored as the number of questions answered correctly, out of a total of 18 questions, listed in Appendix A4. An index of pedagogical practices is constructed using principal-component analysis on a series of 22 binary variables capturing features of the teacher's pedagogical practice (both recommended and discouraged practices, the latter being reverse-coded); the practices observed are listed in Panel C of Appendix Table A2.[7]

The secondary outcomes of interest include an index of school management practices; a binary variable for ZIP leader supervision (whether they visited so far this school year); an index of parental support for schooling; student dropout (for baseline students); and both teacher and student attendance.[8] The index of school management practices is constructed using principal component analysis analyzing a set of four binary variables capturing recommended school management practices.[9] The

---

[7]Positive practices are denoted with (+) and negative ones with (-). While 26 measures were collected, we felt that four measures were ambiguous in terms of whether they are good or bad practices and thus they are omitted from the index. These were: the teacher using Portuguese as the language of instruction (that is the official policy but studies have shown that using mother-tongue to teach until grade 3 facilitates learning), using books in Portuguese/Emakhuwa for the same reason, having students read silently on their own, and the teacher working on paperwork (while this reduces active time with students, they paperwork may have been important for student learning – for example, records needed by the school). These ambiguous practices are denoted with ($\sim$). We also run a specification with these four practices included and the results are very similar. Results are available on request.

[8]In the pre-analysis plan, both ZIP leader supervision and the parental support index were specified as variables capturing program exposure, rather than secondary outcome variables. Teacher and student attendance were formulated as variables capturing absenteeism, but were restructured (reverse-coded) to correspond to attendance.

[9]These include binary variables equal to one if: the school director was present at the start of the day of the school visit (reported by enumerators); if the school director conducted an observation of teachers at least once over the past month (reported by teachers); if the school council met at least once over the past month (reported by school director or deputy school director); and if parents contact the school to ask about schoolwork (reported by teachers).

parent support index is constructed using principal component analysis on a set of four binary variables capturing parental engagement with education and specifically literacy.[10] Summary statistics of the variables comprising these indices are contained in Appendix Table A2, by experimental arm.

School dropout is the only variable measured longitudinally, and is constructed by tracking the 10 primary sample students from the baseline cohort of grade four students over time (verifying their enrollment two years later, when they would be in grade six). Student attendance is measured using a student self-report of the number of days they were present in the past five school days. Teacher attendance over the same recall period (last five school days) is reported by deputy school directors.

3 Empirical Strategy

Our primary specification uses ordinary least squares to capture intent-to-treat effects given the randomized design. We estimate the following equation:

$$Y_{ist} = \alpha + \beta_1 T_{1s} + \beta_2 T_{2s} + \gamma y_{s,t-1} + \delta + \epsilon_{ist} \tag{1}$$

Here, y_{ist} is a school-level (s) or student-level (i) outcome of interest measured at time t (i.e., endline survey), and T_1 and T_2 are indicators for each treatment arm at the school level. Given that we do not observe the same students at baseline, we instead control for the baseline school-level mean $y_{s,t-1}$ for student-level outcomes; the baseline value of the same outcome is included as a control when available in the specifications estimated for some school and teacher-related outcomes. The specification also includes strata fixed effects δ and standard errors clustered at the school level (equivalent to simple heteroskedasticity-robust standard errors for school-level outcomes); we also report the p-value corresponding to the test $\beta_1 = \beta_2$.

We also estimate a simpler regression to capture the pooled effect of any treatment on outcomes of interest.

$$Y_{ist} = \alpha + \beta_{any} T_s + \gamma y_{s,t-1} + \delta + \epsilon_{it} \tag{2}$$

where T_s is an indicator variable for any treatment (T1 or T2) and other variables are as previously defined. For the EGRA outcomes (each of the five sub-tasks), we

---

[10]These include binary variables for reading books with students, incorporating learning into everyday activities, assisting with homework, and encouraging their attendance at reading camps (reported by students).

additionally test whether there is any treatment effect across all of the sub-tasks, and report this test separately for the raw scores and the binary variables for non-zero scores.

We assess balance of both school and student-level characteristics at baseline, as reported in Appendix Table A1. The first three columns present raw means and standard deviations for each of the experimental groups. Column 4 presents p-values of a test from whether the coefficients from a regression of the treatment dummies on the outcome variable are jointly statistically insignificant.[11] In general, the hypothesis that school characteristics are consistent across treatment and control arms cannot be rejected, but there is some limited evidence of imbalance in student characteristics. However, we also use the method in Kerwin et al. (2024), which recommends using randomization inference in balance tests since conventional balance tests tend to over-reject balance. The randomization inference p-value of a test that all coefficients are zero is 0.42.[12]

The sample schools are characterized by moderately sized classes, with about 30 students on average, and employ eight teachers.[13] Schools are relatively poorly resourced with almost no schools having electricity, but all have functioning latrines. Time spent teaching is short due to students only attending one of two shifts per day (only about three hours per day for each group of students), but attendance is high at 4 or 4.5 days within the previous five days for the students and teachers, respectively. The sample students are 11 years old on average; two thirds report having books at home that they can read, but about one third also work, primarily helping their parents on the family farm. Notably, only 3% of students and 30% of teachers report speaking Portuguese at home.

4 Results

This section reports the key empirical results. We begin by discussing program implementation, followed by the experimental effects for our primary and secondary outcomes.

---

[11]The specification is exactly as in Equation 2.

[12]Using the conventional method, there are two characteristics that show some evidence of imbalance. The first is children's age, with those in T1 being slightly older. Second, children in the control group are more likely to have books at home compared to children in T1 or T2. The magnitudes, however, are trivial.

[13]Primary school encompasses grades one to six in Mozambique.

4.1 Implementation

We first present evidence around program exposure in Table 1, where the findings suggest that the interventions were implemented with only moderate fidelity. The first column of Panel A shows that 13% of teachers in the control arm report receiving UL training in the past two years; suggestive of some moderate contamination or misrecall. In T1 and T2, 55% and 60% of teachers report attending UL training, respectively; the two treatment effects are statistically significant, but are not different from each other. Teachers in the treatment arm who attended training report attending for 2.3 days on average; many did not attend the follow-up session. The majority of schools do report receipt of teaching materials (70%), and all schools receiving materials report receipt of Portuguese materials, but only 23% report receipt of materials in Emakhuwa.

Panels B and C of the same table report outcomes linked to implementation fidelity of the reading camps, where we observe a somewhat higher quality of implementation vis-a-vis the teacher training. Again, some reading camps previously existed in many communities (these were also part of the Save the Children program), as almost 30% of schools in the control arm have a reading camp.[14] Schools in both T1 and T2 were significantly more likely to have a functional reading camp, and in T2, virtually all communities had such a camp, consistent with the evaluation design. The number of students enrolled in a camp is accordingly significantly higher in the treatment arms (inclusive of enrollment of zero in communities without a camp), as shown in Column (2) of Panel B. Conditional on reporting a camp, the probability of weekly meetings was around 86% and consistent across arms (Column 3). These latter two outcomes are reported by reading camp leaders.

Reports from students are, however, somewhat different (Panel C): in schools in the control and T1 arms, only 10% of students report membership in a reading camp, and even in T2, fewer than half of students report membership. Conditional on membership, around 50% of students in T2 report attending at least weekly. The reading camp quality index, calculated based on six variables describing activities conducted during the camps, shows positive but insignificant treatment effects in

---

[14]A community is identified as having a reading camp if the enumerators can identify a reading camp leader through reports of the deputy school director, teachers, or students, which were triangulated to ensure no "phantom" groups were included.

both arms,[15] and similarly for the teacher support index;[16] in both cases, the index is coded as missing for communities that do not report a camp.

---

[15]The variables are: the number of positive reading activities implemented, the camp meets once or more per week, the camp received books/learning materials as well as lesson plans, the number of visits received by the support teacher, and whether the community supports the reading camps, all reported by the reading camp leader.

[16]Variables include whether the support teacher: helps recruit students, liaises with parents and the school council, attends the reading camps, provides advice regarding pedagogy, and contributes materials.

Table 1. Implementation

Panel A: Teacher training

	(1) Attended UL training (last 2 years)	(2) Days of training attended	(3) UL provided materials	(4)
T1: Training (1)	0.421***	0.490**	0.529***	
	(0.081)	(0.248)	(0.082)	
T2: Training + RCs (2)	0.469***	0.575**	0.497***	
	(0.084)	(0.258)	(0.086)	
Control Mean	0.130	1.185	0.204	
p-value (T1) vs (T2)	0.617	0.700	0.720	
Any treatment	0.443***	0.530**	0.514***	
	(0.067)	(0.227)	(0.071)	
N	160	160	160	

Panel B: Reading camps – establishment and enrollment

	Functioning camp	Students enrolled	Camp meets at least weekly	
T1: Training (1)	0.223**	21.754*	-0.017	
	(0.090)	(11.744)	(0.120)	
T2: Training + RCs (2)	0.721***	36.295***	-0.173	
	(0.063)	(8.705)	(0.121)	
Control Mean	0.278	18.759	0.857	
p-value (T1) vs (T2)	0.000	0.230	0.107	
Any treatment	0.457***	28.612***	-0.12	
	(0.075)	(8.513)	(0.111)	
N	160	160	91	

Panel C: Reading camps – attendance and quality

	Student is camp member	Attending camp at least weekly	Reading camp quality index	Teacher support index
T1: Training (1)	-0.026	-0.058	0.189	0.227
	(0.026)	(0.100)	(0.376)	(0.474)
T2: Training + RCs (2)	0.305***	0.171*	0.235	0.509
	(0.043)	(0.096)	(0.324)	(0.443)
Control Mean	0.082	0.357	-0.747	-0.409
p-value (T1) vs (T2)	0.000	0.000	0.869	0.354
Any treatment	0.130***	0.093	0.219	0.412
	(0.034)	(0.094)	(0.315)	(0.429)
N	1,596	911	91	91

Notes: $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$. This table shows the fidelity of implementation of the two treatment arms. UL denotes the Unlock Literacy Program, RC denotes reading camp. The reading camp quality index is reported by RC leaders and is constructed as the first principal component of variables reported in Panel A of A2. The teacher support index is reported by the RC Leaders and is constructed as the first principal component of the variables reported in Panel B of Table A2. Indices are coded as missing for communities that do not report a camp. The rows "Any Treatment" report the pooled effect of receiving either of the two treatments. All regressions include strata fixed effects and all student-level regressions have standard errors clustered at the school level. School-level regressions include heteroskedasticity-robust standard errors.

4.2 Treatment effects

The primary goal of the UL program was to enhance children's reading abilities, and thus, we first report findings for the EGRA as a primary outcome. Table 2 reports the results of the two interventions on each of the five sub-tasks contained in the assessment; Panel A analyzes raw scores, and Panel B analyzes binary variables for a non-zero score.

In general, we observe weakly positive and often insignificant treatment effects in a context of very low levels of reading ability in the control arm, driven by a large proportion of zero scores. In Panel A, we see that in T1 there is a positive and significant treatment effect for listening comprehension, and in T2 there are positive and significant treatment effects for letter identification and listening comprehension (and a small and marginally significant effect for reading comprehension). The two treatment coefficients are not statistically different from one another for any sub-task. The p-value for the joint hypothesis that the impacts are statistically significant across all five sub-tasks is 0.071.

Panel B of Table 2 shows more positive effects on students scoring above zero on particular sub-tasks: in T1, there are positive and significant effects on all sub-tasks except reading comprehension, and in T2, there are positive and statistically significant effects on all sub-tasks except reading fluency. The effects generally suggest a 10 percentage point increase in the probability of a non-zero score (reduced for reading comprehension), relative to means in the control arm suggesting around 40–60% of students attain a non-zero score (except for reading comprehension, where the mean in the control arm is only 6%). The pooled treatment effects are uniformly positive and significant, and the p-value from a test of significance across all five sub-tasks is 0.052. These findings suggest that, while the treatment effects are relatively small, the effects are concentrated among students in the lowest part of the ability distribution.

Table 2. Early Grade Reading Assessment (EGRA) results

Panel A: Raw scores

	(1) Letter identification (out of 100)	(2) Familiar words (out of 30)	(3) Listening comprehension (out of 4)	(4) Reading fluency (out of 161)	(5) Reading comprehension (out of 4)
T1: Training	1.648	0.176	0.222**	0.667	0.037
	(1.368)	(0.473)	(0.101)	(0.597)	(0.027)
T2: Training + RCs	2.709**	0.675	0.237**	0.942	0.053*
	(1.312)	(0.485)	(0.101)	(0.599)	(0.029)
Control Mean	9.757	2.679	0.645	3.29	0.073
Control SD	(14.83)	(5.07)	(0.93)	(5.70)	(0.29)
p-value (T1) vs (T2)	0.394	0.262	0.893	0.629	0.628
p-value joint impact (T1)	0.107				
p-value joint impact (T2)	0.207				
Any treatment	2.154*	0.417	0.229***	0.8	0.045*
	(1.192)	(0.425)	(0.085)	(0.526)	(0.023)
p-value joint impact (T)	0.071				

Panel B: Scored greater than zero

	Letter identification	Words	Listening comprehension	Reading fluency	Reading comprehension
T1: Training	0.111**	0.117**	0.083*	0.139***	0.035
	(0.046)	(0.05)	(0.046)	(0.048)	(0.024)
T2: Training + RCs	0.085*	0.110**	0.107**	0.07	0.051**
	(0.047)	(0.05)	(0.044)	(0.049)	(0.026)
Control Mean	0.578	0.529	0.421	0.656	0.064
Control SD	(0.49)	(0.50)	(0.49)	(0.48)	(0.24)
p-value (T1) vs (T2)	0.568	0.885	0.583	0.133	0.568
p-value joint impact (T1)	0.075				
p-value joint impact (T2)	0.089				
Any treatment	0.099**	0.114**	0.095**	0.106**	0.043**
	(0.041)	(0.044)	(0.039)	(0.043)	(0.02)
p-value joint impact (T)	0.052				
N	1,596	1,596	1,596	1,596	1,594

Notes: *$p < 0.10$,** $p < 0.05$,*** $p < 0.01$. This table reports treatment effects of the two treatment arms on reading scores. Panel A reports continuous scores and Panel B reports an indicator equal to one if the student scored above zero. The rows "Any Treatment" report the pooled effect of receiving either of the two treatments. The p-value of the joint impact is calculated as the impact of any treatment across all outcomes in the panel. All regressions include strata fixed effects and baseline average scores at the school level as controls. Standard errors clustered at the school level.

Table 3 reports treatment effects for the two other primary outcome variables (teacher pedagogical knowledge and practice) and a range of secondary variables. There are no significant treatment effects for either pedagogical knowledge or practices.[17]; the constituent variables for these indices are enumerated in Panel E of Table A2.[18]

---

[17]We assessed Cronbach's alpha as a measure of internal reliability for the index of classroom observation measures; α was found to be 0.729, above the threshold for internal reliability.

[18]We conduct a robustness check dropping variables that are considered ambiguous in terms of being positive or negative practices (as described above), namely, whether the lesson was conducted in Portuguese, whether teachers used books in either Portuguese or Emakhuwa, and whether the teacher had the students read silently on their own. These practices may be considered ambiguous

This pattern of results is consistent with the results presented earlier, suggestive of generally weak intervention implementation, and that teachers' ability to retain or implement knowledge imparted in the training was low. There is also no evidence of any effects on the school management index, a variable capturing the prevalence of recommended practices in school management such as regular teacher observation, the provision of feedback to teachers, and regular school council meetings, and no effect on supervision provided by ZIP leaders. Parental support for education or literacy is similarly unchanged. Panel B of the same table reports findings for dropout and attendance for students and teachers, and again, the we find consistent null effects.

Table 3. Teacher, Student, and School Outcomes

Panel A: Teacher pedagogy and school management					
	(1)	(2)	(3)	(4)	(5)
	Teacher pedagogical knowledge	Teacher pedagogical practice	School management index	ZIP leader supervision	Parental support index
T1: Training	0.071	0.358	-0.124	0.079	0.026
	(0.282)	(0.350)	(0.209)	(0.078)	(0.108)
T2: Training + RCs	0.273	0.142	0.140	0.020	0.114
	(0.303)	(0.362)	(0.210)	(0.084)	(0.117)
Control Mean	11.574	-0.164	0.000	0.759	-0.041
p-value (T1) vs (T2)	0.423	0.532	0.205	0.442	0.437
Any treatment	0.166	0.256	0.001	0.051	0.068
	(0.263)	(0.311)	(0.182)	(0.071)	(0.097)
N	160	160	160	160	1,596
Panel B: Attendance					
	Student dropout	Student attendance (last 5 days)	Teacher attendance (last 5 days)		
T1: Training	-0.025	-0.059	0.203		
	(0.024)	(0.112)	(0.180)		
T2: Training + RCs	-0.020	0.124	-0.312		
	(0.027)	(0.102)	(0.246)		
Control Mean	0.117	4.243	4.519		
p-value (T1) vs (T2)	0.860	0.063	0.016		
Any treatment	-0.023	0.027	-0.040		
	(0.017)	(0.096)	(0.187)		
N	1,516	1,596	160		

Notes: $^*p < 0.10,^{**}p < 0.05,^{***}p < 0.01$ The teacher pedagogical knowledge measured by the teacher quiz score is the total score on an 18 question quiz, listed in Appendix Table A4, on concepts taught during teacher training. The teacher pedagogical practice score is the first principal component from the PCA of indicator variables from the classroom observation tool comprising of 22 questions reported in reported in Panel C of A2. The school management index is the first principal component of an index generated from five indicator variables listed in Panel D of A2. The parental support index is the first principal component of four indicator variables listed in Panel E of A2. Data for student drop-outs comes from the follow-up survey with the DSD based on the students surveyed at baseline. Student attendance is self-reported for the past five school days. Teacher attendance is reported by the DSD for the five school days preceding the interview. All regressions include strata fixed effects and robust standard errors. Student-level outcomes have standard errors clustered at the school level and school-level outcomes have heteroskedasticity-robust standard errors. Column 1 of Panel B includes weights for class size.

in terms of whether they are positive or negative. The results are consistent and are available upon request.

We then examine heterogeneity across several dimensions specified in the pre-analysis plan: child gender, the baseline school management index, a baseline school asset index (variables are defined in Panel F of TableA2), a baseline child-level wealth index (variables described in Panel G), and baseline performance on the EGRA (separately for each sub-task). The variables are constructed as baseline school-level averages and are divided into schools that lie at or above the median value in the distribution and those below the median. We find no evidence of heterogeneous impacts on the sub-tasks of the EGRA scores across any of these baseline variables. Appendix Table A3 reports the results.

5   Conclusion

This paper evaluates the effects of two early grade reading interventions on the reading scores of third grade students in rural Mozambique. The first intervention (T1) involves training teachers in good pedagogical practices for early grade literacy and the provision of learning materials. The second intervention (T2) additionally introduces reading camps that are led by a volunteer from the community and held outside of school hours to further promote reading.

Our findings generally suggest that the interventions were implemented with relatively low compliance, and show weak effects. Teachers received only two days of training on average, and only about half of students report attendance in a reading camp. Given this low engagement, somewhat unsurprisingly, the intervention effects are not large: there are a few, though small, positive effects on early grade literacy scores. Effects are stronger for the reading camp intervention, but are still small. We document larger impacts, generally stronger in T2, on the probability of a non-zero score, suggesting that the programs were more effective for lower-ability learners. There are no effects on teacher pedagogical knowledge or practices, or student or teacher attendance. Some underlying reasons for these weak effects likely include low teacher participation in the training, the provision of materials primarily in Portuguese when almost no students report speaking Portuguese, and the relatively low levels of education of the reading camp leaders coupled with high rates of illiteracy among this population. Overall, our findings are consistent with evidence from the literature that more intensive pedagogical interventions that are generally effective in shifting teacher behavior and student learning.

References

Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja, "Report cards: The impact of providing school and child test scores on educational markets," *American Economic Review*, 2017, *107* (6), 1535–1563.

Arancibia, Violeta, Anna Popova, and David Evans, "Training teachers on the job: What works and how to measure it," *World Bank Policy Research working paper*, 2016, (7834).

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland, and Michael Walton, "From proof of concept to scalable policies: Challenges and solutions, with an application," *Journal of Economic Perspectives*, 2017, *31* (4), 73–102.

_ , _ , _ , _ , _ , **Shobhini Mukherji, Marc Shotland, and Michael Walton**, "Mainstreaming an effective intervention: Evidence from randomized evaluations of "Teaching at the Right Level" in India," Technical Report, National Bureau of Economic Research 2016.

Banerjee, Abhijit V., Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani, "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India," *American Economic Journal: Economic Policy*, February 2010, *2* (1), 1–30.

Bank, World, 2024. https://genderdata.worldbank.org/en/economies/mozambique , retrieved September 6, 2024.

Barrera-Osorio, Felipe, Paul Gertler, Nozomi Nakajima, and Harry Patrinos, "Promoting parental involvement in schools: Evidence from two randomized experiments," 2020. National Bureau of Economic Research Working Paper 28040.

Behrman, Jere R., C. Simon Fan, Naijia Guo, Xiangdong Wei, Hongliang Zhang, and Junsen Zhang, "Tutoring efficacy, household substitution, and student achievement: Experimental evidence from an after-school tutoring program in rural China," *International Economic Review*, 2024, *65* (1), 149–189.

Bold, Tessa, Deon Filmer, Gayle Martin, Ezequiel Molina, Brian Stacy, Christophe Rockmore, Jakob Svensson, and Waly Wane, "Enrollment without Learning: Teacher Effort, Knowledge, and Skill in Primary Schools in Africa," *Journal of Economic Perspectives*, November 2017, *31* (4), 185–204.

Buhl-Wiggers, Julie, Jason T. Kerwin, Juan Muñoz-Morales, Jeffrey Smith, and Rebecca Thornton, "Some children left behind: Variation in the effects of an educational intervention," *Journal of Econometrics*, 2022.

Conn, Katharine M, "Identifying effective education interventions in sub-Saharan Africa: A meta-analysis of impact evaluations," *Review of Educational Research*, 2017, *87* (5), 863–898.

Das, Jishnu, Stefan Dercon, James Habyarimana, Pramila Krishnan, Karthik Muralidharan, and Venkatesh Sundararaman, "School inputs, household substitution, and test scores," *American Economic Journal: Applied Economics*, 2013, *5* (2), 29–57.

Dubeck, Margaret M. and Amber Gove, "The Early Grade Reading Assessment (EGRA): Its Theoretical Foundation, Purpose, and Limitations," *International Journal of Educational Development*, January 2015, *40*, 315–322.

Duflo, Annie, Jessica Kiessel, and Adrienne Lucas, "Experimental evidence on alternative policies to increase learning at scale," 2020. National Bureau of Economic Research WP 27298.

Duflo, Esther, Pascaline Dupas, and Michael Kremer, "Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya," *American Economic Review*, 2011, *101* (5), 1739–1774.

Eble, Alex, Chris Frost, Alpha Camara, Baboucarr Bouy, Momodou Bah, Maitri Sivaraman, Pei-Tseng Jenny Hsieh, Chitra Jayanty, Tony Brady, Piotr Gawron, Stijn Vansteelandt, Peter Boone, and Diana Elbourne, "How much can we remedy very low learning levels in rural parts of low-income countries? Impact and generalizability of a multi-pronged para-teacher intervention from a cluster-randomized trial in the Gambia," *Journal of Development Economics*, 2021, *148*, 102539.

Evans, David K and Amina Mendez Acosta, "Education in Africa: What are we learning?," *Journal of African Economies*, 2021, *30* (1), 13–54.

Evans, David K. and Anna Popova, "What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews," *The World Bank Research Observer*, 10 2016, *31* (2), 242–270.

Fazzio, Ila, Alex Eble, Robin L Lumsdaine, Peter Boone, Baboucarr Bouy, Pei-Tseng Jenny Hsieh, Chitra Jayanty, Simon Johnson, and Ana Filipa Silva, "Large learning gains in pockets of extreme poverty: Experimental evidence from Guinea Bissau," *Journal of Public Economics*, 2021, *199*, 104385.

Gilligan, Daniel O, Naureen Karachiwalla, Ibrahim Kasirye, Adrienne M Lucas, and Derek Neal, "Educator incentives and educational triage in rural primary schools," *Journal of Human Resources*, 2022, *57* (1), 79–111.

Glewwe, Paul, Michael Kremer, and Sylvie Moulin, "Many children left behind? Textbooks and test scores in Kenya," *American Economic Journal: Applied Economics*, 2009, *1* (1), 112–135.

_ , _ , _ , **and Eric Zitzewitz**, "Retrospective vs. prospective analyses of school inputs: the case of flip charts in Kenya," *Journal of Development Economics*, 2004, *74* (1), 251–268.

Glewwe, Paul W, Eric A Hanushek, Sarah D Humpage, and Renato Ravina, "School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010," 2011.

INDE/MINEDH, "Avaliação Nacional da 3a Classe," 2017.

Kerwin, Jason, Nada Rostom, and Olivier Sterck, "Striking the Right Balance: Why Standard Balance Tests Over-Reject the Null, and How to Fix It," Technical Report, IZA Discussion Papers 2024.

Kerwin, Jason T and Rebecca L Thornton, "Making the grade: The sensitivity of education program effectiveness to input choices and outcome measures," *Review of Economics and Statistics*, 2021, *103* (2), 251–264.

Kremer, Michael, Conner Brannen, and Rachel Glennerster, "The challenge of education and learning in the developing world," *Science*, 2013, *340* (6130), 297–300.

Krishnaratne, Shari and Howard White, "Quality education for all children? What works in education in developing countries," Technical Report, International Initiative for Impact Evaluation (3ie) 2013.

Mambo, Félix, Maren Basso, Vincenzo Salvucci, and Ricardo Santos, "An analysis of school dropout in Mozambique," 2019.

Maro, Vincenzo Di, Stefan Leeffers, Danila Serra, and Pedro C Vicente, "Mobilizing parents at home and at school: an experiment on primary education in Angola," *Economic Development and Cultural Change*, 2024, *72* (3), 1069–1116.

Mbiti, Isaac, Karthik Muralidharan, Mauricio Romero, Youdi Schipper, Constantine Manda, and Rakesh Rajani, "Inputs, incentives, and complementarities in education: Experimental evidence from Tanzania," *The Quarterly Journal of Economics*, 2019, *134* (3), 1627–1673.

Muralidharan, Karthik, "Field experiments in education in developing countries," in "Handbook of economic field experiments," Vol. 2, Elsevier, 2017, pp. 323–385.

_ , **Abhijeet Singh, and Alejandro J Ganimian**, "Disrupting education? Experimental evidence on technology-aided instruction in India," *American Economic Review*, 2019, *109* (4), 1426–1460.

Murnane, Richard J and Alejandro Ganimian, "Improving educational outcomes in developing countries: Lessons from rigorous impact evaluations," 2014. National Bureau of EConomic Research Working Paper w20284.

Nyqvist, Martina Björkman and Andrea Guariso, "Supporting learning in and out of school: Experimental evidence from India," 2021. Manuscript, Stockholm School of Economics.

Popova, Anna, David K Evans, Mary E Breeding, and Violeta Arancibia, "Teacher professional development around the world: The gap between evidence and practice," *The World Bank Research Observer*, 2022, *37* (1), 107–136.

Romero, Mauricio, Lisa Chen, and Noriko Magari, "Cross-age tutoring: Experimental evidence from Kenya," *Economic Development and Cultural Change*, 2022, *70* (3), 1133–1157.

United Nations, "Sustainable Development Goals No. 4: Quality Education," 2019. https://unstats.un.org/sdgs/report/2019/goal-04/; Retrieved July 5, 2024.

World Bank, *World Development Report 2018: Learning to realize education's promise*, The World Bank, 2017.

Appendix

A1 Additional Tables

Appendix Table A1. Baseline Balance

	Control (C)	Training (T1)	Training + RCs (T2)	p-value: $\beta_1 = \beta_2 = 0$	Overall Sample
Panel A: School-level variables					
Number of students in Grade 4	28.630	35.696	29.180	0.284	31.275
	(4.02)	(33.99)	(6.82)		(20.74)
DSD High school+	0.889	0.893	0.900	0.976	0.894
	(0.32)	(0.31)	(0.30)		(0.31)
DSD Male	0.852	0.875	0.880	0.918	0.869
	(0.36)	(0.33)	(0.33)		(0.34)
Director's Age	39.370	38.679	36.900	0.237	38.356
	(7.71)	(7.14)	(7.29)		(7.41)
DSD Experience (years)	15.056	14.946	12.580	0.081	14.244
	(7.20)	(7.04)	(5.73)		(6.77)
School Asset Index	0.014	-0.149	0.152	0.381	-0.000
	(1.15)	(1.43)	(0.81)		(1.17)
Teacher High school+	0.759	0.804	0.800	0.669	0.787
	(0.43)	(0.40)	(0.40)		(0.41)
Teacher's Age	31.407	32.786	31.180	0.507	31.819
	(6.57)	(8.58)	(5.87)		(7.14)
Teacher Male	0.315	0.179	0.260	0.255	0.250
	(0.47)	(0.39)	(0.44)		(0.43)
Years as a Teacher	7.444	9.125	6.720	0.193	7.806
	(4.73)	(8.20)	(5.31)		(6.36)
Panel B: Student-level variables					
Child Male	0.439	0.500	0.469	0.242	0.470
	(0.20)	(0.18)	(0.19)		(0.19)
Child's Age	11.270	11.830	11.334	0.058	11.485
	(1.16)	(1.48)	(0.77)		(1.21)
Number of Siblings	3.309	3.105	3.470	0.171	3.288
	(1.15)	(0.85)	(1.16)		(1.06)
Speak Portuguese at home	0.017	0.043	0.032	0.140	0.031
	(0.04)	(0.10)	(0.06)		(0.08)
Child has books at home	0.726	0.618	0.670	0.037	0.671
	(0.20)	(0.24)	(0.23)		(0.22)
Child's Asset Index	0.010	-0.083	0.058	0.297	-0.008
	(0.48)	(0.49)	(0.43)		(0.47)
p ($p = c/n$ joint test for balance)					0.42
Observations	54	56	50		160

DSD denotes deputy school director. Columns 1,2,3 and 5 report raw means with standard deviations reported in parentheses. Column 4 reports the p-values from the joint hypothesis test on coefficients for T1 and T2 both being zero. The school facilities index is the score for the first principal component of a PCA analysis for the school having the following: latrine, piped/tubewell water source, electricity, reading material in Portuguese and/or Emakhuwa, and soap/detergent for staff. The assets index for pupils reports the score for the first principal component of a PCA analysis for variables indicating their household's ownership of: radio, livestock, mobile phone, and bicycle. Estimated models include strata fixed effects and heteroskedasticity-robust standard errors for school-level outcomes, and standard errors clustered at the school level for student outcomes.

Appendix Table A2. Summary Statistics for Indices by Treatment Group

	Full sample	Control	Teacher Training (T1)	Reading Camps (T2)
Panel A: Reading Camp Quality Index Variables				
# of activities implemented	1.69	1.71	1.63	1.72
	(1.12)	(1.33)	(1.21)	(1.03)
Camp meets 1+ times a week	0.76	0.86	0.85	0.68
	(0.43)	(0.36)	(0.36)	(0.47)
Received books/learning material	0.92	0.79	0.93	0.96
	(0.27)	(0.43)	(0.27)	(0.20)
Received lesson plan	0.77	0.71	0.78	0.78
	(0.42)	(0.47)	(0.42)	(0.42)
# visits by support teacher	4.38	2.93	4.19	4.90
	(5.12)	(3.02)	(5.31)	(5.47)
Community supports RCs	0.89	0.93	0.93	0.86
	(0.31)	(0.27)	(0.27)	(0.35)
Observations	91	14	27	50
Panel B: RC Teacher Support Index Variables				
Helps recruit students	0.66	0.57	0.67	0.68
	(0.48)	(0.51)	(0.48)	(0.47)
Liaises with parents	0.33	0.21	0.33	0.36
	(0.47)	(0.43)	(0.48)	(0.48)
Liaises with the school council	0.23	0.36	0.30	0.16
	(0.42)	(0.50)	(0.47)	(0.37)
Attends the reading camps	0.47	0.29	0.41	0.56
	(0.50)	(0.47)	(0.50)	(0.50)
Provided advice regarding pedagogy	0.55	0.64	0.63	0.48
	(0.50)	(0.50)	(0.49)	(0.50)
				Continued

Appendix Table A2. Summary Statistics for Indices by Treatment Group (continued)

	Full sample	Control	Teacher Training (T1)	Reading Camps (T2)
Teacher contributed materials	0.36	0.50	0.33	0.34
	(0.48)	(0.52)	(0.48)	(0.48)
Other	0.05	0.00	0.11	0.04
	(0.23)	(0.00)	(0.32)	(0.20)
Observations	91	14	27	50
Panel C: Classroom Observation				
Teacher has a lesson plan (+)	0.88	0.85	0.89	0.88
	(0.33)	(0.36)	(0.31)	(0.33)
Lesson was in Portuguese ($\sim$)	0.96	0.96	0.98	0.94
	(0.19)	(0.19)	(0.13)	(0.24)
Teacher recapped previous lesson (+)	0.61	0.61	0.62	0.60
	(0.49)	(0.49)	(0.49)	(0.49)
Teacher outlined the day's lesson (+)	0.97	0.98	1.00	0.94
	(0.16)	(0.14)	(0.00)	(0.24)
Teacher used any books (if available) ($\sim$)	0.86	0.80	0.91	0.88
	(0.35)	(0.41)	(0.29)	(0.33)
Teacher used printed learning materials (+)	0.28	0.26	0.32	0.26
	(0.45)	(0.44)	(0.47)	(0.44)
The teacher used visual aids (+)	0.35	0.28	0.43	0.34
	(0.48)	(0.45)	(0.50)	(0.48)
Teacher used call & response for sounds (+)	0.37	0.41	0.32	0.38
	(0.48)	(0.50)	(0.47)	(0.49)

Appendix Table A2. Summary Statistics for Indices by Treatment Group (continued)

	Full sample	Control	Teacher Training (T1)	Reading Camps (T2)
The teacher used call and response for letters (+)	0.53	0.65	0.46	0.48
	(0.50)	(0.48)	(0.50)	(0.50)
The teacher used call and response for words or sentences (+)	0.76	0.80	0.73	0.76
	(0.43)	(0.41)	(0.45)	(0.43)
Teacher used songs/rhymes/alliteration (+)	0.28	0.20	0.36	0.28
	(0.45)	(0.41)	(0.48)	(0.45)
Teacher asked students to create a story/song/picture/dance (+)	0.11	0.09	0.14	0.08
	(0.31)	(0.29)	(0.35)	(0.27)
Teacher told/read a story (+)	0.11	0.09	0.18	0.06
	(0.32)	(0.29)	(0.39)	(0.24)
The teacher paused to ask questions about a topic they were teaching (+)	0.70	0.65	0.79	0.66
	(0.46)	(0.48)	(0.41)	(0.48)
The teacher paused to check whether students were comprehending (+)	0.94	0.93	0.93	0.98
	(0.23)	(0.26)	(0.26)	(0.14)
Teacher asked questions to the whole class (+)	0.92	0.94	0.89	0.92
	(0.27)	(0.23)	(0.31)	(0.27)
Teacher called on students individually (+)	0.64	0.65	0.68	0.58

Appendix Table A2. Summary Statistics for Indices by Treatment Group (continued)

	Full sample	Control	Teacher Training (T1)	Reading Camps (T2)
	(0.48)	(0.48)	(0.47)	(0.50)
Teacher allowed students to ask questions (+)	0.84	0.83	0.86	0.84
	(0.36)	(0.38)	(0.35)	(0.37)
Teacher split students into groups (+)	0.10	0.06	0.11	0.14
	(0.30)	(0.23)	(0.31)	(0.35)
The teacher had students read silently on their own (∼)	0.13	0.09	0.12	0.18
	(0.34)	(0.29)	(0.33)	(0.39)
There was an assessment during the class (+)	0.86	0.87	0.84	0.86
	(0.35)	(0.34)	(0.37)	(0.35)
Teacher was on their phone (-)	0.05	0.06	0.04	0.06
	(0.22)	(0.23)	(0.19)	(0.24)
Teacher left the class unattended (-)	0.05	0.07	0.04	0.04
	(0.22)	(0.26)	(0.19)	(0.20)
Teacher assigned homework (+)	0.61	0.57	0.62	0.64
	(0.49)	(0.50)	(0.49)	(0.48)
Teacher checked homework (+)	0.25	0.20	0.25	0.30
	(0.43)	(0.41)	(0.44)	(0.46)
The teacher worked on paperwork (∼)	0.01	0.02	0.00	0.00
	(0.08)	(0.14)	(0.00)	(0.00)
Observations	160	54	56	50
Panel D: School Management Index Variables				
DSD was present	0.62	0.57	0.66	0.62
				Continued

Appendix Table A2. Summary Statistics for Indices by Treatment Group (continued)

	Full sample	Control	Teacher Training (T1)	Reading Camps (T2)
	(0.49)	(0.50)	(0.48)	(0.49)
DSD visits 1+ times a month	0.66	0.63	0.68	0.66
	(0.48)	(0.49)	(0.47)	(0.48)
Council met 1+ times this month	0.36	0.31	0.48	0.26
	(0.48)	(0.47)	(0.50)	(0.44)
Parents contact about schoolwork	0.91	0.93	0.89	0.90
	(0.29)	(0.26)	(0.31)	(0.30)
Observations	160	54	56	50
Panel E: Parental Support Index Variables				
Parents read to pupil	0.23	0.22	0.22	0.26
	(0.42)	(0.42)	(0.41)	(0.44)
Parents incorporate learning into everyday activities	0.06	0.06	0.06	0.05
	(0.23)	(0.23)	(0.24)	(0.21)
Parents assist with homework	0.44	0.44	0.45	0.44
	(0.50)	(0.50)	(0.50)	(0.50)
Parents encourage to attend RC	0.03	0.01	0.02	0.05
	(0.16)	(0.12)	(0.13)	(0.21)
Observations	1596	535	561	500
Panel F: School Facilities Index				
Functioning latrine	1.00	1.00	1.00	1.00
	(0.00)	(0.00)	(0.00)	(0.00)
Use piped/tubewell water	0.72	0.72	0.73	0.70
	(0.45)	(0.45)	(0.45)	(0.46)
School has electricity	0.03	0.04	0.05	0.00
	(0.17)	(0.19)	(0.23)	(0.00)

Continued

29

Appendix Table A2. Summary Statistics for Indices by Treatment Group (continued)

	Full sample	Control	Teacher Training (T1)	Reading Camps (T2)
Books in Portuguese	0.94	0.94	0.91	0.98
	(0.23)	(0.23)	(0.29)	(0.14)
Books in local language	0.06	0.07	0.04	0.06
	(0.23)	(0.26)	(0.19)	(0.24)
Soap available for staff	0.84	0.85	0.84	0.82
	(0.37)	(0.36)	(0.37)	(0.39)
Observations	160	54	56	50
Panel G: Student Assets Index				
Family Has Radio (1=Yes)	0.32	0.33	0.29	0.35
	(0.47)	(0.47)	(0.45)	(0.48)
Family Has Livestock (1=Yes)	0.88	0.84	0.88	0.91
	(0.33)	(0.36)	(0.32)	(0.29)
Family Has Mobile Phone (1=Yes)	0.62	0.62	0.61	0.63
	(0.49)	(0.49)	(0.49)	(0.48)
Family Has Bicycle (1=Yes)	0.41	0.39	0.40	0.47
	(0.49)	(0.49)	(0.49)	(0.50)
Observations	1521	515	529	477

This table reports the means within each treatment group for variables used in the parents' support, school management, teacher support to reading camp (RC), RC quality, and classroom observation indices. Standard deviations are reported in parentheses. In Panel C for pedagogical practices, all observations marked (+) were considered a "good" practice and scored positively. Observations marked (-) counted as a "bad" practice and scored negatively. Those marked ($\sim$) were considered ambiguous - these were excluded from the index used in the main results. For robustness, a version of the index with these included was also tested - there was no significant difference between the results. The index is the first principal component from a PCA on these variables.

Appendix Table A3. Heterogeneous Effects on EGRA scores

	(1) Letter identification (out of 100)	(2) Familiar words (out of 30)	(3) Listening comprehen- sion (out of 4)	(4) Reading fluency (out of 161)	(5) Reading com- prehension (out of 4)
Baseline Student Gender					
Male * T1	-0.447	-0.455	0.064	0.422	0.059
	(2.000)	(0.647)	(0.109)	(0.717)	(0.039)
Male * T2	0.114	-0.253	0.085	-0.076	0.041
	(2.054)	(0.642)	(0.114)	(0.845)	(0.044)
Baseline School Management Index					
Above med. score * T1	-2.281	0.216	-0.195	-0.593	0.054
	(3.754)	(1.222)	(0.258)	(1.576)	(0.073)
Above med. score * T2	-2.954	-0.472	-0.054	-1.003	0.052
	(3.576)	(1.276)	(0.269)	(1.568)	(0.078)
Baseline School Facilities Index					
Above med. score * T1	-4.427	-1.309	-0.223	-1.348	-0.015
	(2.787)	(0.910)	(0.227)	(1.114)	(0.056)
Above med. score * T2	-1.019	-0.426	0.083	-0.386	0.037
	(2.679)	(1.009)	(0.214)	(1.152)	(0.060)
Baseline Pupil Asset Index					
Above med. score * T1	3.362	1.042	-0.044	1.567	0.077
	(2.698)	(0.920)	(0.206)	(1.112)	(0.055)
Above med. score * T2	0.397	-0.077	-0.166	0.604	-0.001
	(2.609)	(0.983)	(0.208)	(1.176)	(0.059)
Baseline Average EGRA Scores for School					
Above med. score * T1	1.022	-0.583	0.058	-0.895	-0.065
	(2.781)	(0.977)	(0.216)	(1.128)	(0.055)
Above med. score * T2	-3.807	-0.611	-0.049	-0.072	-0.072
	(2.476)	(1.002)	(0.210)	(1.202)	(0.059)
Control Mean	9.757	2.679	0.645	3.290	0.073
N	1,596	1,596	1,596	1,596	1,594

A2   Detailed overview of EGRA

The EGRA exam measures student skills in five domains or sub-tasks. These include the following:

1. Letter name identification and reading: This sub-task is used to measure whether children can identify and read letters presented in a random order, both lower-case and uppercase. In this sub-task, children were presented with a chart of 100 letters and asked to read as many of these letters as they could within one minute. The score is the number of letters correctly identified and read in one minute.

2. Familiar words reading: This sub-task assesses the ability of children to decode printed words and read them correctly. The task reflects both the accuracy and fluency of reading, which are fundamental skills for developing the ability to read and comprehend what is read. The children were presented with a card containing 30 words common to their daily life, including their school life, and asked to read as many words as possible in one minute. The score is the number of words correctly read in one minute.

3. Listening comprehension: This sub-task is used to measure whether children have basic knowledge of the language in question and whether they can process what they hear in that language. In this sub-task, the enumerators read aloud a short text comprising 56 words for the children and then asked them four questions to check their comprehension. The children were not given a copy of the text to refer to when answering the questions. The score is the number of correct answers given by the children to the four questions asked.

4. Oral reading fluency: This sub-task assesses the speed, accuracy, and expressivity at which children read texts. The task reflects the ability to translate letters into sounds, recognize familiar words, decode unfamiliar words, and make sense of the text's meaning. The children were given a card with a narrative text of

120 words and asked to read as many words as possible in one minute. The metric of oral reading fluency was the number of correct words per minute (cwpm) read by the student.

5. Reading comprehension: This sub-task assesses the ability of the children to extract and construct meaning out of the texts they read. After reading the narrative text used as the stimulus for the oral reading fluency sub-task, the children were asked up to five questions based on how much text they had read. For example, the enumerator only asked the first question to those children who managed to read at least nine words of the text given. The score is the proportion of comprehension questions correctly answered.

A3 Overview of field procedures

Careful quality assurance protocols were used to ensure fidelity to high-quality data collection principles. Supervisors monitored enumerators' work and directly observed interviews as appropriate. Regular data checks were conducted by the survey firm as well as IFPRI.

During the baseline survey, it was deemed appropriate to seek informed consent from the parents of all students as the sample included a supplementary questionnaire for grade seven students collecting data on sensitive questions linked to mental health. Since the endline survey did not involve such questions, in addition to student assent, informed consent was sought from the school principal, who acts on behalf of the parents as they are the ones who have been entrusted with the children during school hours.

The consent seeking procedure was as follows: Before starting fieldwork, a verification exercise took place by mobilizers whereby the survey firm visited schools to introduce themselves and collect information on school and class timings. Upon arrival at the school, the enumerators spoke with the school principal and informed them of the objectives of the study, read them the comprehensive informed consent form, and asked whether they consent to having the school participate in the study and having students interviewed. When students were subsequently interviewed, they were also read a comprehensive consent statement and provided verbal assent. Children were informed more than once that they do not have to agree to participate. A letter was sent home with each student to provide to their parents stating what activities occurred at school during the day and contact information of a PI was provided

in case any parent wanted to remove their child from the data.

For adults who were surveyed, written consent was sought on the day of the survey and they were provided with a consent statement that included: 1) objectives of the study; 2) study procedures; 3) risks and benefits of participating in the study; 4) strategies used by researchers to minimize risks; 5) costs/compensation associated with participating in the study; 6) the duration of the interviews; 7) the voluntary nature of the study and the participant's right to refuse to answer questions or leave the study; 8) that all information would be confidential, that nobody would be able to identify any particular individual's responses, and that their data would be kept securely; and 9) contact information for study staff.

A4   Teacher Knowledge Quiz

Appendix Table A4. Teacher Quiz Questions

Q.	Question	Response Code	Answer
1	What are the five components of effective reading instruction? (not prompted)	1 - Letter knowledge 2 - Phonemic awareness 3 - Reading fluency 4 - Vocabulary 5 - Reading comprehension	Any of 1-5 mentioned
2	What are some examples of students learning vocabulary indirectly? (not prompted)	1 - Being read to at home 2 - Being read to at school 3 - Being read to in any/other contexts 4 - Seeing words in various contexts 5 - Conversations with adults 6 - Reading on their own	Any of 1-6 mentioned
3	List items that may contribute to a "print-rich environment" in the classroom (not prompted)	1 - Posters/artwork with pictures and writing by teachers 2 - Posters/artwork with pictures and writing by students 3 - Posters with lists 4 - Number tables/other math 5 - Flashcards/ anagrams/letter lines 6 - Books on display/reading corners/reading baskets 7 - Objects in the classroom being labeled 8 - Texts from students' homes - phone books/flyers/grocery lists/letters etc. 9 - Class-specific material (e.g. class rules, schedules, lesson plans etc.) 99 - Other, specify	Any of 1-9 mentioned (or another appropriate response)

No.	Question	Response Code	Answer
4	What are some goals of start of year and regular reading assessments? (not prompted)	1 - Identify a child's knowledge and skills in reading 2 - Help teachers identify specific instructional practices to help that child progress in reading 3 - Reflect existing achievement goals to gauge the child's progress and inform the establishment of new goals 4 - Provide tools for communicating about the child's progress in reading to the child, his or her parents or caregivers and other teachers 5 - To track student progress 98 – Don't know 99 – Other, specify	Any of 1-5 mentioned (or another appropriate response)
5	Teachers, and not students, should always be the ones reading text aloud in class as students may make mistakes	True/False	F
6	The teacher should always ask the same students to read aloud in class every day to help them form a habit	True/False	F
7	Only the inclusion of text written by the teacher counts as an example of "print-rich environment"	True/False	F
8	Classrooms with desks organized in rows can be helpful for silent, independent reading or writing	True/False	T
9	Reading Comprehension refers to the ability to read a text accurately and quickly	True/False	F

No.	Question	Response Code	Answer
10	Letter Knowledge is the ability to notice, think about, and work with the individual sounds in spoken words	True/False	F
11	To teach vocabulary, the teacher should expose students to the same words repeatedly	True/False	T
12	A marker of reading fluency is when a student's voice tones rise and fall at appropriate points when reading	True/False	T
13	Upper and lower case letters should be taught simultaneously rather than sequentially	True/False	F
14	Acting out stories can be a useful exercise to promote reading fluency	True/False	T
15	Breaking words into syllables helps children learn phonemic awareness	True/False	T
16	The ability to identify words that rhyme is part of phonemic awareness	True/False	T
17	Singing songs with high-frequency vocabulary words can help improve students' reading comprehension skills	True/False	F
18	The ability to derive meaning from a text indicates the literal comprehension part of reading comprehension	True/False	F