Journal of Public Economics 199 (2021) 104385

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

Large learning gains in pockets of extreme poverty: Experimental evidence from Guinea Bissau

Ila Fazzio [a,1], Alex Eble [b,1,*], Robin L. Lumsdaine [c,d,e,f], Peter Boone [a], Baboucarr Bouy [g], Pei-Tseng Jenny Hsieh [h], Chitra Jayanty [i], Simon Johnson [e,j], Ana Filipa Silva [k]

[a] Effective Intervention, UK
[b] Teachers College, Columbia University, USA
[c] Kogod School of Business, American University, USA
[d] Erasmus University Rotterdam, The Netherlands
[e] National Bureau of Economic Research, USA
[f] Tinbergen Institute, The Netherlands
[g] Effective Intervention, The Gambia
[h] University of Oxford, UK
[i] Independent Consultant, India
[j] Massachusetts Institute of Technology, USA
[k] Effective Intervention, Guinea Bissau

ARTICLE INFO

Article history:
Received 27 November 2019
Revised 4 February 2021
Accepted 8 February 2021
Available online 9 June 2021

Keywords:
Education
Bundled intervention
Randomized controlled trial (RCT)
State capacity
Literacy
Numeracy

ABSTRACT

Children in many extremely poor, remote regions are growing up illiterate and innumerate despite high reported school enrollment ratios. Possible explanations for such poor outcomes include demand – for example, low perceived returns to education compared to opportunity cost; and supply – poor state provision and inability of parents to coordinate and finance better schooling. We conducted a cluster-randomized trial in rural Guinea Bissau to understand the effectiveness and cost of concerted supply-based interventions in such contexts. Our intervention created simple schools offering four years of education to primary-school aged children in lieu of the government. At endline, children receiving the intervention scored 58.1 percentage points better than controls on early grade reading and math tests, demonstrating that the intervention taught children to read and perform basic arithmetic, from a counterfactual condition of very high illiteracy. Our results provide evidence that particularly needy areas may require more concerted, dramatic interventions in education than those usually considered, but that such interventions hold great potential for increasing education levels among the world's poorest people.

1. Introduction

Children in many extremely poor, remote regions are growing up illiterate and innumerate despite high reported school enrollment ratios (Glewwe and Muralidharan, 2016). This "schooling without learning" has many alleged sources, such as insufficient demand for schooling, inadequate schooling materials, and lack of qualified, motivated teachers (Kremer et al., 2013; Pritchett, 2013). This leads to at least three important social phenomena: one, a substantial part of the population being illiterate and innumerate; two, for these children, lower lifetime incomes as a result, and less opportunity to succeed in the growing worlds around them; and three, potentially greater socioeconomic inequality between these children and children in areas which receive better schooling.

In this article, we report the results of a cluster-randomized controlled trial (RCT) evaluating a supply-based intervention which aims to dramatically increase learning levels in particularly poor, rural areas of the developing world. The intervention provides the early years of primary school in lieu of the government; this entails hiring, training, and monitoring teachers tasked with delivering schooling, from the pre-primary level on to grade 3, to primary-aged children. The intervention uses a bespoke curriculum which includes teacher training materials and teaching and learning materials for both teachers and students. It also employs frequent monitoring and assessment of teachers and children and

* Corresponding author.
E-mail address: eble@tc.columbia.edu (A. Eble).
[1] Fazzio and Eble share "co-first author" status, as they contributed equally to the work (order of co-first author names was randomized according to Ray and Robson 2018).

https://doi.org/10.1016/j.jpubeco.2021.104385
0047-2727/

regular community outreach / involvement. We conducted this RCT in rural areas of Guinea Bissau, one of the poorest and most troubled countries on the planet (Silva and Oliveira, 2017).

The intervention yielded transformative learning gains among children who would otherwise be unlikely to ever achieve literacy and numeracy. After four years of receiving the intervention, children in the intervention group scored 58 percentage points better than children in the control group on a composite score of tests of mathematics and reading ability. This difference comprises large gains in both math and reading ability across the difficulty spectrum, from letter and number recognition to reading comprehension and two-digit subtraction with borrowing. A very high proportion of control children had zero scores on these tests. Using the test's definition of literacy, 63 percent of intervention children demonstrate literacy at endline, compared to less than 0.1 percent of control children. Unfortunately, there is no consensus on how to measure numeracy in these tests. Instead, we report two results. Using a benchmark from Ghana[3], 21.3 percent of intervention children demonstrate numeracy at endline while no control children do. Using a measure of more basic numeracy, 73.2 percent of intervention children display basic numerical skills, while less than 0.1 percent of control children do.

These gains are dramatic in absolute as well as relative terms, with intervention children from rural Guinea Bissau exhibiting literacy and numeracy skills similar to children in much wealthier countries with functioning school systems. A commonly-used metric for measuring reading skill among early grade children is oral reading fluency (ORF), measured by the correct number of words read per minute from a set passage. Endline ORF of children randomized to receive the intervention was 75 correct words per minute. This compares favorably to the ORF measured in a 2014 national assessment of third grade students in the Philippines and is similar to that of the (much wealthier) Latin American countries who have used similar tests.[4]

Our approach has important common traits with the influential studies of ambitious, highly-resourced interventions in the US designed to address inequality and raise outcomes for the less fortunate. The most famous of these are the Perry Pre-school and Abecedarian programs (Campbell and Ramey, 1994; 1995; Heckman et al., 2013). There are three main similarities: first, these programs targeted needy or at-risk children. Second, they provided a suite of services, including a comprehensive educational intervention which comprised well-trained and well-supervised teachers, a structured curriculum, and family outreach. Finally, similar to our program, those programs were also relatively expensive, but demonstrated a positive return on investment above that of equity (Heckman et al., 2010). Overall, we argue that our study provides proof of concept that a resource-intensive intervention can generate large gains in a challenging setting, but perhaps with a model that might be difficult to scale or replicate. This is reflected in other work which documents that achieving scalable impacts in education is difficult, especially among highly effective interventions (Banerjee et al., 2017; Bold et al., 2018).

Our approach also parallels research on the efficacy of charter schools and "model schools" in the US (Angrist et al., 2013; Dobbie and Fryer Jr, 2013). These studies show that new, non-governmental schools which combine a suite of teaching practices and other components known to be effective can substantially improve learning, relative to traditional public schools. Furthermore, gains are largest in contexts, similar to ours, where the status quo option is of particularly low quality (Chabrier et al., 2016).

Our findings contribute to ongoing efforts to identify effective means to increase learning levels, and welfare more generally, in the poorest parts of the world (McEwan, 2015; Glewwe and Muralidharan, 2016). A growing set of studies shows the potential for targeted interventions to achieve large gains in settings with low learning levels (c.f., Burde et al., 2013; Muralidharan et al., 2019). We advance this work by showing the success of a concerted supply-based intervention –which delivered all aspects of early primary education instead of the government – in achieving these goals in a particularly challenging setting. Our approach mirrors the use of "bundled" interventions to tackle otherwise intractable problems, such as extreme poverty (Banerjee et al., 2015).

The rest of our paper proceeds as follows. Section 2 describes the context we work in, the challenges we encountered in initial implementation, and the final intervention design. Section 3 describes our research design. Section 4 presents our main results. Section 5 discusses our results in the context of other studies of education in disadvantaged areas and Section 6 concludes.

2. Background and intervention details

In this section, we describe the context in which the study took place, the initial challenges faced in early attempts to implement the intervention, and the final intervention we study.

2.1. Context

Guinea Bissau is a Lusophone country in West Africa with a population of approximately 1.8 million people. Once a Portuguese colony, it attained independence in 1974. Since then, it has been beset by political and economic troubles. There have been four coups d'etat since its founding. Until 2018, there had been no elected president who had completed a full five-year term. It is one of the poorest countries in the world both on per-capita GDP terms and according to the UN's Human Development Index (Silva and Oliveira, 2017).[5] Aside from some parts of the capital, there is no national power or water grid. The official language of the country is Portuguese but the dominant language is Crioulo – a hybrid of Portuguese and several local tongues – which is spoken as a first or second language by the majority of the population.

In Fig. A.1 we show a map of the country and our study areas. Our study took place in villages in the Quinara and Tombali regions in the southwest of the country. These regions were selected for two reasons: first, the government requested that we work in the two regions as they were less well-served by existing NGO work; second, Boone et al. (2014) identified them as the regions with the lowest learning levels in the country.

2.2. Education, literacy, and numeracy in Guinea Bissau

Guinea Bissau's official education system comprises three levels: nine years of compulsory, basic education (four years of lower primary, called the "first cycle"; two years of upper primary, or second cycle; and three years of middle school, or third cycle), followed by three years of elective secondary school and then higher education. The official ages for primary school are currently 6–12.[6] As in many developing countries, the age at which children actually enter school varies widely.

Boone et al. (2014) report the results of a nationally representative survey of schools, families, and children across Guinea Bissau

[3] Described in Section 4.

[4] Philippines: https://earlygradereadingbarometer.org/overview, accessed on October 28, 2019. Latin America: the average grade three ORF is 73 words per minute in English, and 79 in Spanish according to USAID (2019).

[5] The economy is largely dependent on agriculture, primarily cashews. Because of its geographic location and low state capacity, Guinea Bissau has been used as a way station for the transportation of cocaine to Europe, adding to corruption and governance issues (Silva and Oliveira, 2017).

[6] They were 7–13 at the start of our trial.

in 2010.[7] They found very low education levels among parents: among fathers, approximately 40 percent had ever been to school, and 24 percent were able to read a printed paragraph. Among mothers, only nine percent had ever been to school, and 2.8 percent were able to read the same paragraph. Among children, however, the survey found substantial enrollment in school: approximately 85 percent of interviewed children between the ages of 7 and 17 had been to school, and 70 percent were currently enrolled. Unfortunately, these high enrollments did not translate to learning. Fewer than one third of these children could recognize a single digit number or read a single, simple Portuguese word.

Parents recognized the low quality of the education their children were getting, and expressed demand for higher quality schools. Of the over 8,500 parents and caregivers interviewed, more than 98 percent asserted that they would be willing to pay, on average, approximately 20 percent of household income per school-aged child, for better schooling for the child. The authors of that study conclude that there is probably substantial demand in rural Guinea Bissau for quality schooling, but some combination of income, credit market failures, capacity, and collective action constraints impede its provision. Even so, the extremely poor educational outcomes in these regions – regardless of the type of schools – suggests that either demand or supply could be the key reason that children grow up mostly illiterate and innumerate. These findings motivated the current study.

2.3. Status quo provision of education in study area

Guinea Bissau is often considered a "failed state" because of its frequent coups, highly irregular payment of its civil servants, and the absence of many basic government services. Education is one such service, and the reach of government schools in most areas, including our study area, is uneven and erratic. At baseline, only half of the schools in our trial area were run by the government, with the rest run by either the local community (35%) or an NGO or other private organization (15%). Ostensibly, children are meant to attend school for four hours per day, five days per week, nine months out of the year. In practice, government schools were open less frequently in our study area because of teacher strikes in these schools; according to official data, strikes disrupted roughly 25% of school days for government schools during our study. Not all official strikes made their way to our rural areas, however, and roughly half of the schools in the control area were not run by the government and so were not affected.

While statistical data from the government and other sources is sparse, Boone et al. (2014) also provide a thorough description of the "status quo" of education provision in rural Guinea Bissau. The study visited schools to collect data on teachers (presence and demographic data), as well as infrastructure data from a representative sample of 351 schools and 781 teachers. The authors found that 86 percent of visited schools were open, with teachers present and teaching, and 72 percent of enrolled children were present when the schools were visited.[8] These schools all had chalkboards and roughly one textbook for every 30 children. The average pupil:teacher ratio (for combined grades 1–4, as many schools have combined classrooms) was 63.4, with a high standard deviation (24.4). Boone et al. (2014) found very low correlation between either teacher qualifications or school resources and child learning levels, corroborating prior research (Lepri, 1988; Daun, 1997).

Overall, these areas are characterized by extremely low learning levels despite the fact that, barring strikes, schools are usually open and teacher and student absenteeism is relatively low. Although Boone et al. (2014) set out to find examples of success in these areas, it found no such examples. A main conclusion of their paper, which also motivated this study, is that in Guinea Bissau "the public sector cannot be relied on to provide regular services due to political instability, institutional capacity, and a political system that does not serve the very poor."

2.4. Intervention design

Initially, we recruited a group of nearly 50 prospective "untrained" teachers to deliver the intervention and trained them for one year.[9] At the end of this year of training, the trainees reneged on their commitments to us, demanding a dramatic change in the agreed-upon employment conditions – including a salary increase to a level equivalent to that of the education ministry's director-general – and sued us in the country's courts. While the government sided with us and these individuals' suit was determined to be without merit, we were forced to postpone the study until the court case was resolved. The case was ultimately resolved in our favor, but resulted in our loss of all 48 selected candidates. In Appendix A, we explain this experience in greater detail.

We then had to begin the search for – and training of – candidates anew, and we decided to hire certified teachers instead of untrained ones. The logic behind this decision was twofold: one, these teachers required less training and so the extra training we gave them would be less likely to cause them to demand dramatically higher compensation; two, it would allow us to start the intervention more promptly. Using this strategy, we were able to identify fewer willing and suitable candidates. We describe the impact of this on our study design in the next section.

In villages randomly selected to receive the intervention, we provided four years of school – first, a year of pre-primary school focusing on Portuguese language acquisition, then grades 1–3 of the national primary education curriculum. This schooling was meant to take the place of official instruction in these years usually delivered by Guinea Bissau's government educational system. We included the year of pre-primary because the national curriculum is in Portuguese. To the best of our knowledge, only a trivially small number of children in our study area had any knowledge of the language at the time of school entry.

We aimed to have 25–30 students per class, resulting in a total of 24 academic classes across the 16 intervention villages in our study. Classes were held in spaces provided and furnished by each community.[10] The curriculum of these classes was designed to maximize child participation throughout the day. The overall intervention strategy was inspired by the experience, design, tools, and teaching methods of an early primary school intervention designed by the Naandi Foundation and evaluated in a prior RCT in India (Lakshminarayana et al., 2013).[11] Final instructional tools were developed in consultation with, and with review by, the ministry

[7] Excluding the islands of Bolama and Bilagós.

[8] This level of teacher absenteeism is less severe than found in Uganda in Chaudhury et al. (2006) and at the lower end of the range of what Blimpo et al. (2011) observe in Gambia.

[9] Originally this study was part of a larger effort to study the generalizability of a para teacher intervention in India (Lakshminarayana et al., 2013), run in tandem with a similar effort in The Gambia (Eble et al., 2021).

[10] This request for support from the community was intended to promote community backing of the intervention and to increase parent involvement in the formal education of their children and the management of the academic classes.

[11] This study, along with the study reported in Eble et al. (2021), were a part of larger efforts to attempt to replicate the success of Lakshminarayana et al. (2013) in newer, more challenging contexts. In Eble et al. (2021), which took place in The Gambia, the authors used the after-school supplementary lesson design of the intervention studied in Lakshminarayana et al (2013). In Guinea Bissau, we shifted our strategy to providing regular schooling, instead of the state, in light of the history of frequent, prolonged disruptions to state-provided education.

of education in Guinea Bissau, covering the content in the official Guinea Bissau primary curriculum. These tools included daily lesson plans, a teacher handbook, child workbooks, and other grade-specific didactic materials.

Teachers were recruited with the requirement that they be able to speak and teach in the local language spoken in the community in which they were assigned to work. Once hired, they received two types of training: first, 10 weeks of initial pre-service training in how to implement the intervention; second, four weeks of in-service training conducted annually before the beginning of each new academic year to prepare teachers to teach the next year's content.[12] In each village, the intervention also hired a local adult for the first four months who spoke the most prominent local language. This person assisted the teacher with classroom management and the children's transition from use of their mother tongue to Portuguese.

Teachers conducted classes for five hours per day, five days a week, plus additional hours when required by the curriculum plan or teachers' assessments of child learning needs, for nine months each year. The duration of the intervention spanned February 2014 to December 2017, comprising 730 school days in total. Teachers were paid salaries of 200,000 Central African Francs (or CFA; roughly, US $345) per month, with an additional per-diem to compensate them for the difficulty of living in the villages in which they worked (1,500 CFA, or US $2.59, per day).[13]

The intervention team monitored both teachers' work and children's learning in order to track progress and ensure that learning was progressing as planned. Monitors – a separate cadre of staff recruited by the intervention arm – visited each academic class for two days each month. The team conducted monthly, two-day review meetings for teachers and monitors. In these meetings, teachers received feedback and training based on the evidence collected during that month's classroom observations/monitoring. These meetings were also used to reinforce the intervention's main methodology and teaching strategies, focusing on concrete examples of what to do, how to do it, and what not to do. Each month, the intervention team assessed some children on the curriculum in their current grade, and conducted larger-scale evaluations of child learning every six months.

Implementing this intervention was intensely challenging. We chose to work in small, isolated villages; the rugged terrain, long distances between villages, and poor state of the roads between them made frequent, spontaneous monitoring difficult, particularly during the rainy season when some villages become inaccessible. These villages lacked internet connections and reading materials, and had few or no literate residents who might reinforce child learning. This also made it difficult to recruit qualified teachers, who were required to reside in the village.[14] Further complicating literacy efforts, multiple languages are spoken in these regions, none of which have their own script. Finally, none of the parents enumerated were native speakers of Portuguese, the official language of the curriculum and of the intervention; this also restricted children's ability to practice and apply the lessons from class outside of school.

3. Research design

This section describes our research design, including the study population, our sample size/power calculations, the nature of the data collected, and the pre-specified (relative to unblinding of the data) analysis plan.

3.1. Study design

In the first screening of villages for eligibility, we began with all four hundred and thirty-nine villages in the Quinara and Tombali regions with between 50 and 400 households according to the Guinea Bissau National Institute of the Census.[15] We used existing map information and Quantum GIS (version 1.7.2) to select villages that were at least nine kilometers apart from each other to avoid risks of spillover from one village to another. With this method we preselected 49 villages for enumeration, along with a set of backups should there be need for replacement.

We then conducted field visits to record the GPS points of these villages and confirm whether they met the following three eligibility criteria for inclusion in our study: i) the village had between 50 and 400 households; ii) the village was reachable by land during the country's dry season; and iii) the village had no other NGO-administered education program taking place. Within these villages, our eligibility criteria for enrolling children in the study were that: i) the child was born between January 2007 and September 2008; ii) the child was resident in an eligible village; iii) the child did not have any serious physical or mental conditions that may have impaired learning, i.e., severe developmental handicaps; and iv) the child's parents gave consent to participate in the study.

We further restricted eligibility to villages which had at least 20 eligible children. After the initial village visits to confirm eligibility, four of the 49 pre-selected villages had fewer than 20 eligible children and therefore were not included; these villages were replaced with other villages from the list of backups. We then enrolled these final 49 villages, containing a total of 2,112 eligible children, for participation in our study.[16] Given the teacher recruitment challenges noted in the previous section, we switched from a 1:1 control:intervention cluster ratio to a 2:1 ratio to ensure that we only worked in as many villages as we could find qualified teachers for. Our final sample comprised 16 intervention villages and 33 control villages.

We conducted randomization by computer, stratifying at the village level based on a composite variable comprising a weighted average of several indicators: the village's distance to the nearest road, the highest grade taught by the local school (in the one case where the village did not have a school, we set this to zero), the number of households in the village, the proportion of mothers speaking Crioulo in the village, and the third quartile of mothers' educational attainment in the village. We selected these variables on the assumption that they would be correlated with the primary outcome, as shown in Boone et al. (2014). The results of our cluster analysis suggested that randomizing within two strata was sufficient.[17] This led to the generation of one stratum with 32 villages, in which villages were randomized 2:1 to control and intervention status, and another stratum with 17 villages and the same randomization profile.

From December 2012 to April 2013, we conducted our baseline enumeration for the purposes of enrolling children into the study. The mean number of enumerated children per village was 43. To conduct our sample size calculation, we took attrition figures from a study of child health in the country, which suggested roughly 17%

[12] These trainings emphasized the use of relevant, grade-appropriate teaching strategies as well as use of the intervention's bespoke teaching and learning materials.

[13] This was raised midway through the trial to be a 250,000 salary and 2,500 per diem, respectively.

[14] Although recruitment of teachers was difficult, once recruited, all teachers remained in the project until its completion.

[15] In this initial screening we also included villages for which information on the number of households was missing.

[16] While the sample size is smaller than we initially planned, it is consistent with or somewhat larger than the sample size of studies of other hard-to-reach populations, e.g., Burde et al.'s 2013 study of community schools in Afghanistan.

[17] The cluster analysis was conducted in SAS Software version 9.3, using the command "PROC CLUSTER."

Table 1
Baseline cluster characteristics.

Variable	(1) Intervention	(2) Control	(3) Difference
Overall distance to a main road* in km (distance = 0 if village has a road)	7.88	8.52	−0.64
Randomized children: mean (SD)	40.56 (19.12)	44.33 (23.59)	3.77
Predominant ethnic group			
Balanta	25% (4)	51.5% (17)	−21.5%
Fula	25% (4)	15.2% (5)	9.8%
Beafada	25% (4)	24.2% (8)	0.8%
Other	25% (4)	9.1% (3)	15.9%
Cluster size (number of households): mean (SD)	117.31 (47.36)	128.85 (74.59)	11.54
Number of villages	16	33	–
F-statistic for test of joint significance (p-value)	–	–	1.51 (0.199)

Notes: this table shows baseline characteristics for the villages in our trial, separately by treatment group and the raw difference between these values.*: Main road is defined as a road that is connected to at least one peri-urban or urban area via regular public transport.

loss to follow-up over the course of the study (Mann et al., 2009). Using this, we expected an average of 35 children per village to be present for the endline test, and thus contribute to the primary outcome.

This led to the following power calculation, conducted before commencing randomization: a study population of 49 villages, with an average of 35 eligible children per village and a 2:1 control:intervention randomization ratio, provides 92% power to detect a difference in test scores of at least 0.25 SD in a two-sided test with a five percent significance level, assuming an intra-cluster coefficient of 0.03. In Appendix Table A.1 we show similar calculations for different scenarios (greater loss to follow-up and a 1:1 control:intervention ratio). We registered our statistical pre-analysis plan (also known as an SAP or PAP) at www.socialscienceregistry.com prior to unblinding of the data.[18]

While the study was unblinded to participants – it was impossible to prevent parents from knowing whether or not they were in a village that was receiving materials and teaching support – the research team that conducted the surveys and tested the children were not given information on which villages were in each arm. Furthermore, these staff were closely monitored to ensure that data collection procedures were consistent across all villages.

In Tables 1 and 2, we provide summary statistics at the village and child level, respectively, showing characteristics separately by whether the village/child is in the intervention or control group. Relative to intervention villages, control villages tended to be slightly more remote and larger in population. For the most part, children in the intervention and control arms were quite similar. At the bottom of each table, we conduct a test for the joint significance of these characteristics in predicting randomization status, as in Bruhn and McKenzie (2009).

3.2. Primary outcome and analysis methods

The pre-specified primary outcome of our study is the child's "composite score." This is the arithmetic mean of the child's scores on EGRA and EGMA tests, administered sequentially, to each enrolled child present in the village at time of testing in November and December of 2017.[19] EGRA and EGMA tests assess early grade reading and math ability, respectively (Platas et al., 2014; Dubeck and Gove 2015). They are administered orally, one-on-one between instructor and child. We chose them to serve as our primary outcome because they are particularly sensitive in measuring small differences in ability among children who have very low levels of learning, such as those in many parts of our trial area. Each test paper has several different subtasks, evaluating a different skill or competency. In Table A.2, we describe the nature of each subtask (the full test papers we used are given in Appendix B). In line with other work using EGRA and EGMA tests, we also present individual test scores, subtask scores, zero scores, and fluency measures (Platas et al., 2014; Dubeck and Gove, 2015).

For our primary analysis, we use a linear regression to estimate the child-level difference between intervention and control groups in the primary outcome, controlling for the stratification factor used in the randomization and nothing else. In all analyses we report robust standard errors, clustering at the village level. Secondary analyses extend this model to (separately) investigate interactions by a series of prespecified subgroups. For secondary outcomes that are continuous, we also use a linear model. For those that are dichotomous (such as whether the child was enrolled in school), we show both "adjusted" differences from a linear probability model (i.e., the estimated coefficient for the intervention variable from the regression) and odds ratios from our (pre-specified) logit model. To account for bias from potential differential attrition between groups, we calculate Lee bounds (Lee, 2009) for our primary outcome and the individual EGRA and EGMA scores.

3.3. Attrition and adherence

We next describe the flow of participants through the trial. Table 3 presents data on whether enrolled children were present in their village at the trial's midline survey and again at the endline survey. We observe roughly 13 percent attrition at midline (in the 2014/15 school year), and roughly 20 percent attrition at endline, with greater attrition from the control arm than from the intervention arm. We show the broader flow via a CONSORT-style diagram, in Fig. A.2 (M. K. Campbell et al., 2012). We also present data on how frequently children assigned to the intervention attended the intervention classes in Table A.3. The average of all intervention children's attendance in intervention classes is above 80%, and about nine percent of intervention children attended no intervention classes.

[18] RCT ID: AEARCTR-0003670.

[19] Our aggregation of EGRA and EGMA tests into a composite score was chosen for simplicity as a single primary outcome, and for consistency with related work on delivering educational interventions to other deprived areas (Lakshminarayana et al., 2013; McEwan, 2015; Evans and Popova, 2016; Eble et al., 2021). We note that this method of aggregation is a departure from conventional use of EGRA and EGMA scores.

Table 2
Baseline child characteristics.

Variable	(1) Intervention	(2) Control	(3) Difference
Child is female	49.15% (319)	48.60% (711)	0.55%
Identity of the interviewed caregiver for the child			
Mother	49.77% (323)	51.26% (750)	−1.49%
Father	16.02% (104)	18.87% (276)	−2.85%
Grandmother	10.32% (67)	10.39% (152)	−0.07%
Grandfather	2.00% (13)	0.96% (14)	1.04%
Aunt	11.71% (76)	7.52% (110)	4.19%
Uncle	3.39% (22)	4.03% (59)	−0.64%
Other	6.78% (44)	6.97% (102)	−0.19%
Mother's education			
No education	66.10% (429)	71.16% (1,041)	−5.06%
Grades 1 to 4	22.96% (149)	18.80% (275)	4.16%
Grades 5 to 10	7.86% (51)	4.99% (73)	2.87%
Grades 11+	0.31% (2)	0.48% (7)	−0.17%
Don't know	2.62% (17)	4.31% (63)	−1.69%
Father's education			
No education	28.35% (184)	30.69% (449)	−2.34%
Grades 1 to 4	16.18% (105)	19.62% (287)	−3.34%
Grades 5 to 10	18.95% (123)	17.02% (249)	1.93%
Grades 11+	4.01% (26)	2.12% (31)	1.89%
Don't know	29.28% (190)	29.12% (426)	0.17%
Child's age at baseline (SD)*	4.81 (0.58)	4.76 (0.58)	0.05
Number of observations	649	1463	–
F-statistic for test of joint significance (p-value)	–	–	1.15 (0.334)

Notes: this table shows baseline characteristics (percent, with corresponding number in parentheses) for the children in the villages in our trial, separately by treatment group, and the raw difference between these values. For age at baseline, mean age is reported (within treatment group standard deviation in parentheses). For mother's education, one observation is missing from the intervention and four from the controls. *: Due to the paucity of official birth or health records, we only have precise child age for 200 intervention children and 332 control children. To calculate the F-statistic, we replace missing age values with an arbitrary number not equal to any observed value and add a dummy for missing age.

Table 3
Children resident in study village (migration).

Year residence measured	(1) Intervention	(2) Control	(3) Adjusted difference	(4) *p*-value
Midline (late 2014/early 2015) (N: I = 648; C = 1,462)	89.04% (577)	84.95% (1,242)	4.51% (1.94)	0.025
Endline (early 2017) (N: I = 646, C = 1,455)	87.77% (567)	75.19% (1,094)	12.53% (2.24)	<0.001

Notes: columns 1 and 2 show the group-specific proportion of children whom we observed at the time of a midline survey in late 2014/early 2015, and at the endline survey in early 2017, respectively (number of observations shown in parentheses below). Column 3 shows the "adjusted" difference estimated using our main estimating equation (i.e., the coefficient on the intervention variable in the linear regression described in the previous section), with standard errors, clustered at the village level, below in parentheses. Column 4 shows the p-value of a test of the null that the adjusted difference is zero.

4. Main results

In this section, we present empirical analyses describing the main results of our study. We begin with the primary outcome – the composite test score – and then present comparisons by test (reading or math) and subtasks within each test. We then analyze heterogeneity in these results, the intervention's impact on enrollment in school and attendance, and spillover effects to the child's siblings.

4.1. Primary outcome

We show our primary outcome, alongside the secondary outcomes for overall math and reading scores, in Table 4, Panel A. We observe a very large difference in composite test scores between children in the control and intervention arms at the end of our study. The control child mean score was 11.2%; for intervention children, this mean is 70.5%, or a 58.1 percentage point adjusted difference.[20] A common learning metric in similar studies is to use the standard deviation of the control group as a scale factor. In our setting, this is uninformative given the extremely low learning levels of the control group.[21] We show the distribution of test scores of the two groups in Fig. 1. Decomposing the composite score into its reading and math components, we observe large differences in both tests, although they are larger in reading (6.8% correct vs. 72.5%) than in math (15.6% correct vs. 68.5%). All differences are statistically significant ($p < 0.001$). To bound the potential impact of differential

[20] The adjusted difference is the intervention-control difference for a given variable after controlling for the stratum variable as pre-specified for our main analysis; equivalently, this is the regression coefficient on the intervention variable using our main regression specification.

[21] Were we to use the control SD as a scale factor, the 58.1 percent difference in scores would correspond to a 5.31 SD difference in test scores between the two groups.

Table 4
EGRA and EGMA total scores.

Variable	(1) Intervention (SD)	(2) Control (SD)	(3) Adjusted difference (SE)	(4) Conventional *p*-value	(5) Lee bounds (SE)	(6) RI finite sample *p*-value
Panel A: Pre-specified outcomes						
Composite test score	70.48 (15.35)	11.21 (10.93)	58.14 (1.28)	$p < 0.001$	L: 55.04 (1.27) U: 63.20 (1.25)	$p < 0.001$
Reading score	72.48 (17.07)	6.84 (8.85)	64.44 (0.98)	$p < 0.001$	L: 60.57 (1.09) U: 69.73 (1.17)	$p < 0.001$
Math score	68.48 (16.55)	15.58 (14.82)	51.85 (1.83)	$p < 0.001$	L: 48.87 (1.71) U: 57.67 (1.69)	$p < 0.001$
Panel B: Summary measures						
Composite test score is zero	0% (0)	5.18% (22.17)	−4.02% (1.16)	$p = 0.001$	–* –*	$p = 0.032$
Child is literate	63.94% (48.06)	0.09% (3.04)	62.91% (2.01)	$p < 0.001$	L: 54.15 (2.91) U: 72.04 (3.70)	$p < 0.001$
Child is numerate**	21.31% (40.99)	0% (0)	20.49% (2.73)	$p < 0.001$	L: 4.96 (4.49) U: 22.14 (2.78)	$p < 0.001$
Observations	563	1,081	–	–	–	–

Notes: columns 1 and 2 show the group-specific mean test scores (group-specific SD in parentheses below). Column 3 shows the adjusted difference between the two groups (i.e., the coefficient on the intervention variable in a linear regression, estimated with the inclusion of a control for the stratum variable) with standard errors, clustered at the village level, below in parentheses. Column 4 shows the *p*-value of a test that this difference is equal to zero. Column 5 shows Lee bounds on the estimate in column 3. Column 6 shows exact randomization inference *p*-values of the adjusted difference. *: Estimation of Lee bounds are degenerate for this variable due to there being zero observations with composite score equal to zero in the intervention group and a small number of observations with composite score equal to zero in the control group. Because of this, we do not report them. **: As discussed in the text, the measure of numeracy used here is less well-established and relatively stringent. Using a measure of more basic numeracy (consistently recognizing which of two distinct two- or three-digit numbers is larger and successfully performing at least half of simple addition tasks), we estimate a treatment effect of the intervention on basic numeracy of 71.5 percentage points.

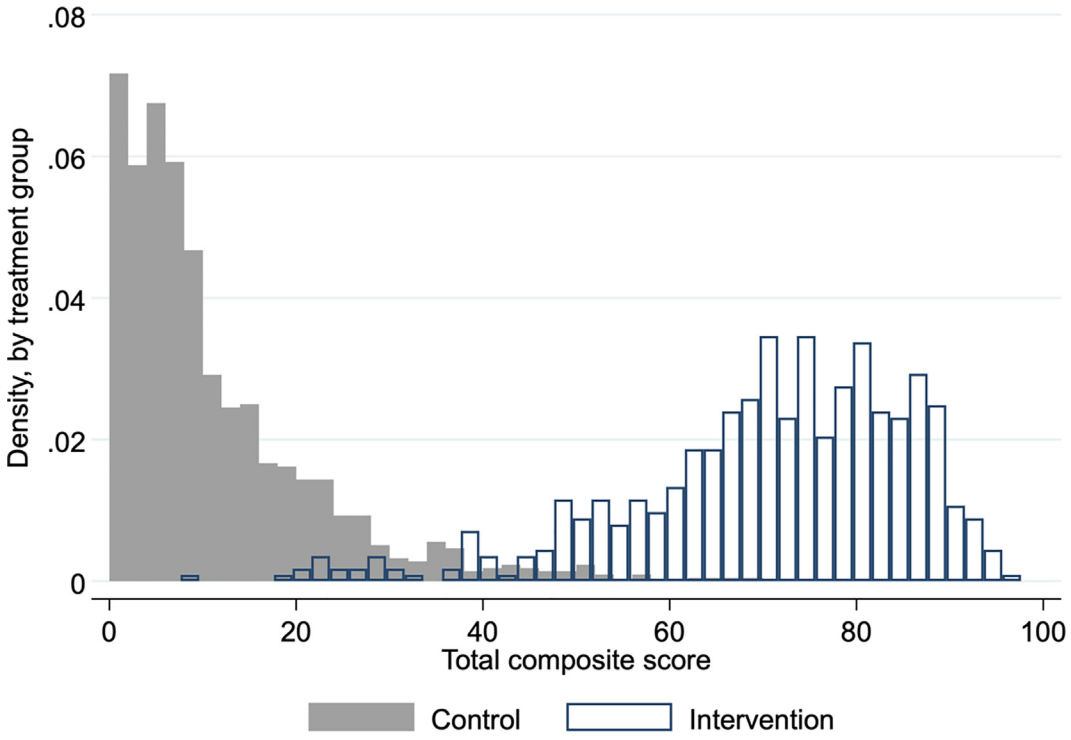

Fig. 1. Distribution of test scores, by treatment group. *Notes:* this figure shows the distribution of the composite test score for the control and intervention groups, separately, for all children who took the endline test.

attrition on our primary outcome estimates, we calculate Lee bounds and show them in column 5 (Lee, 2009). Because our randomization was conducted with a small number of clusters, we also present finite sample randomization inference *p*-values in column 6. These yield strong evidence that the control-intervention test score differences we estimate are not likely to be the result of differential attrition or chance.

We include three additional transformations of the primary outcome in Table 4, Panel B. First, we show the effect of the intervention on the proportion of children with a score of exactly zero on the composite exam. Roughly five percent of control children score exactly zero, while no intervention children register this score. This suggests that while learning levels are very low, the EGRA and EGMA tests we used were successful in avoiding floor effects.

Second, we estimate the impact of the intervention on literacy and numeracy, rather than just the reading and math test scores. It is generally accepted that children are considered proficient readers when they read "with good fluency" (at least 45 words per minute) and can correctly answer 80% or more of the reading

Table 5
EGRA subtasks.

	Percent correct			*Fluency scores*			*Percent with zero score*		
Subtask	**Interv.**	**Control**	***p*-value**	**Interv.**	**Control**	***p*-value**	**Interv.**	**Control**	***p*-value**
Letter recognition (1)	68.3%	11.5%	$p < 0.001$	68.7	11.4	$p < 0.001$	0.0%	35.2%	$p < 0.001$
Initial sound recognition (2)	63.1%	20.9%	$p < 0.001$	–	–	–	3.7%	43.0%	$p < 0.001$
Invented word reading (3)	58.0%	2.3%	$p < 0.001$	29.3	1.2	$p < 0.001$	1.6%	90.6%	$p < 0.001$
Familiar word reading (4)	79.1%	2.7%	$p < 0.001$	45.5	1.3	$p < 0.001$	1.2%	88.8%	$p < 0.001$
Oral reading fluency* (5a)	86.9%	4.3%	$p < 0.001$	75.1	2.9	$p < 0.001$	0.2%	59.1%	$p < 0.001$
Reading comprehension (5b)	72.3%	1.1%	$p < 0.001$	–	–	–	2.8%	95.9%	$p < 0.001$
Listening comprehension (6)	79.7%	5.1%	$p < 0.001$	–	–	–	6.2%	89.2%	$p < 0.001$
Observations	563	1,081	–	563	1,081	–	563	1,081	–

Notes: this table shows the mean percent of correct answers, fluency scores, and zero scores on the individual components of the reading test by treatment group. The number in parentheses next to each subtask label corresponds to the subtask number given in Table A.2. *: The lower proportion of control group zero scores on subtask 5a is a result of the fact that the first question in this subtask happened to be substantially less difficult than the questions asking children to read familiar or made-up words in subtasks 3 and 4. For each type of score (percent correct, fluency, zero score) we also include the *p*-value from a test that the difference between intervention and control values of a given subtask is zero.

Table 6
EGMA subtasks.

	Percent correct			*Fluency scores*			*Percent with zero score*		
Subtask	**Interv.**	**Control**	***p*-value**	**Interv.**	**Control**	***p*-value**	**Interv.**	**Control**	***p*-value**
Number identification (1)	96.7%	30.6%	$p < 0.001$	47.7	7.3	$p < 0.001$	0.0%	15.5%	$p < 0.001$
Quantitative comparisons (2)	89.7%	19.9%	$p < 0.001$	–	–	–	0.2%	41.4%	$p < 0.001$
Missing number (3)	64.7%	11.0%	$p < 0.001$	–	–	–	0.5%	41.6%	$p < 0.001$
Addition level 1 (4a)	67.0%	10.7%	$p < 0.001$	14.6	2.7	$p < 0.001$	1.6%	52.8%	$p < 0.001$
Addition level 2* (4b)	54.8%	3.5%	$p < 0.001$	–	–	–	9.2%	88.5%	$p < 0.001$
Subtraction level 1 (5a)	45.6%	4.5%	$p < 0.001$	9.6	1.3	$p < 0.001$	4.3%	72.5%	$p < 0.001$
Subtraction level 2* (5b)	33.0%	1.0%	$p < 0.001$	–	–	–	28.2%	95.8%	$p < 0.001$
Word problems (6)	52.0%	18.8%	$p < 0.001$	–	–	–	5.7%	37.7%	$p < 0.001$
Observations	563	1,081	–	563	1,081	–	563	1,081	–

Notes: this table shows the mean percent of correct answers, fluency scores, and zero scores on the individual components of the math test by treatment group. The number in parentheses next to the subtask label corresponds to the subtask number given in Table A.2. There are 6–40 missing values in some timed subtasks; adjusting for these missing values changes the fluency score estimates by 0.01–0.35. Given the large intervention-control differences in fluency scores, we do not report these sensitivity analyses here. *: Level 2 subtasks were only administered to children with non-zero scores in addition level 1 and subtraction level 1, respectively. For each type of score (percent correct, fluency, zero score) we also include the *p*-value from a test that the difference between intervention and control values of a given subtask is zero.

comprehension questions associated with the text read (Dubeck and Gove, 2015). Using this classification to generate a binary variable for literacy, we find that the intervention raises literacy rates by 62.9 percentage points, from a baseline of less than a tenth of a percent of control children reaching literacy. Unfortunately, there is no similar consensus on the definition of numeracy using these tests. Using benchmarks from USAID work in Ghana[22], we can create a binary numeracy variable equal to one if the child completes at least 70% of the missing number sequence questions correctly (subtask 3) and at least 80% of the word problem questions correctly (subtask 6). Under this definition, the intervention raises numeracy by 20.5 percentage points, compared to precisely zero control children reaching this level, as reported in Table 4. This is a stringent definition of numeracy; for reference, in 2013 less than 4% of Ghanaian schoolchildren achieved this level of performance on these two subtasks. We also create a variable capturing more basic numeracy skills: the child's ability to compare the magnitude of pairs of two- or three- digit numbers (subtask 2) and compute simple sums (subtask 4a). Using this measure, we estimate a treatment effect of the intervention on basic child numeracy of 71.5 percentage points (not reported in the table).

4.2. Reading

In this section, we describe the results of the EGRA test in greater detail. These are shown in Table 5. In this table, we show three separate scores for each subtask: i) the average percent correct, ii), for timed subtasks, the fluency scores, and iii) the percent of children with a zero score. Intervention children substantially outperformed control children in reading: in all subtasks, the control-intervention difference in the percent of correct answers is at least 42 percentage points (out of 100). Children in the intervention group demonstrated reading skill mastery across subtasks of all difficulty levels. They were able to correctly read more than two thirds of the letters presented to them (under a one minute time limit). For familiar word reading, the mean intervention child read 79 percent of the 50 words presented correctly in one minute. For connected text reading, the intervention children achieve a mean reading fluency of 75 words per minute, which is higher than the defined reading proficiency benchmark for Grade 3 in most of the EGRA countries (RTI International, 2017). It is also much higher than oral reading fluency measures from other African countries who have used EGRA: average grade 3 oral reading fluency in English-speaking African countries is 9.2, and in Francophone African countries it is 32.4 (USAID, 2019). This level of performance is comparable to EGRA results from wealthier Latin American countries, such as Guatemala, Jamaica, and Peru; average oral reading fluency in Latin America is 73 words per minute in English, and 79 in Spanish. For the untimed tasks, the pattern was roughly the same. In the subtask measuring children's comprehension of a connected text, the mean score for intervention children was 72% of questions answered correctly. For the control group, it was one percent.

Another meaningful comparison in EGRA- and EGMA-style tests is the proportion of children with zero correct answers (i.e., a "zero

[22] Source: https://pdf.usaid.gov/pdf_docs/PA00KS7N.pdf, accessed January 20th, 2021.

Table 7
Composite test scores by subgroup, with interaction tests.

Group	(1) Intervention (SD)	(2) Control (SD)	(3) Adjusted difference (SE)	(4) *p*-value
Child gender				
Male	72.57	12.58	58.89	0.188
(*N*: *I* = 297, *C* = 586)	(14.07)	(11.49)	(1.40)	
Female	68.14	9.59	57.41	
(*N*: *I* = 266, *C* = 495)	(16.37)	(9.99)	(1.38)	
*Household wealth**				
Low wealth index	70.47	10.73	58.59	0.835
(*N*: *I* = 227, *C* = 489)	(15.60)	(10.05)	(1.78)	
High wealth index	71.03	12.08	58.15	
(*N*: *I* = 320, *C* = 475)	(14.37)	(11.73)	(1.52)	
Mother's education				
No education	69.41	10.23	58.21	0.900
(*N*: *I* = 366, *C* = 765)	(15.80)	(10.19)	(1.45)	
At least grade 1 education	72.46	13.59	57.98	
(*N*: *I* = 197, *C* = 316)	(14.30)	(12.23)	(1.63)	
Father's education				
No education	70.83	9.72	60.24	0.025
(*N*: *I* = 157, *C* = 335)	(15.45)	(10.29)	(1.41)	
At least grade 1 education	70.35	11.88	57.24	
(*N*: *I* = 406, *C* = 746)	(15.32)	(11.15)	(1.41)	

Notes: this table follows the format of columns 1–4 in Table 4. It shows group-specific means and standard deviation in parentheses below) in columns 1 and 2, and adjusted control/intervention differences in children's scores on the composite test by subgroup in column 3 (with standard errors, clustered at the village level, in parentheses below). *P*-values are for tests of the null of an equal effect of the intervention across subgroups, estimated by calculating the *p*-value on an interaction term between the treatment variable and the subgroup indicator variable. *: The wealth index is high if the caregiver reports 1) that they could find money to pay a sudden medical bill of 42,000 CFA (roughly US $72), and 2) that in the last year their family went no longer than one month without income; it is low otherwise.

score") in each subtask. We show these results in the three rightmost columns of Table 5. These data highlight the exceptionally low learning levels among the control group. In four of the five most difficult reading subtasks, 88 percent or more of the control group earned zero scores. For example, more than 88 percent of the control children tested at endline were unable to read even one of the 50 familiar words presented, compared to only 1.2 percent of children in the intervention group (subtask 4). Similar patterns appear across all subtasks involving reading or oral comprehension, corroborating the very low levels of literacy found in Boone et al. (2014).

4.3. Math

Next, we discuss children's performance, by intervention arm, on math subtasks. We present these results in Table 6, mirroring the format of Table 5. Children in intervention villages also dramatically outperformed children in control villages in terms of math ability, as seen in scores for all subtasks. Intervention children could solve around 15 simple addition problems and around 10 simple subtraction problems per minute, compared with around three addition problems and one subtraction problem for control children, respectively. This suggests intervention children were at least five times more "fluent" in these core arithmetic skills, fundamental and important predictors for subsequent mathematical development (Jordan et al., 2009). For two-digit problems, some with borrowing/carrying, intervention children answered 55% of addition problems and 33% of subtraction problems correctly, compared with 3.5% and 1%, respectively, for control children. For the subtask that evaluates children's ability to discern and complete number patterns – EGMA subtask 3, identifying the missing number in a sequence such as [2, 4, 6, __] – more than half of the intervention group correctly answered 60% or more of the questions. This would be classified as reaching a desired level of performance in this skill for third grade students in several other countries which use the EGMA test to assess child learning (RTI International, 2009). Only 0.2% of the control group score this well on subtask 3. As with reading, far fewer intervention children had zero scores on math subtasks than did control children, with larger control/intervention gaps for more difficult subtasks.

4.4. Heterogeneity in effect size for the primary outcome

In this section, we present a series of pre-specified and exploratory tests for heterogeneity in the effect of the intervention. First, we present our pre-specified tests across a series of demographic characteristics, shown in Table 7. We investigate differential effects of the treatment by child gender, a proxy for the wealth of the family, and the level of education of the child's mother and, separately, father. We see large control-intervention test score differences across all subgroups, but the only statistically significant dimension of heterogeneity is for father's education, and this result is not robust to standard adjustments for multiple hypothesis testing, such as a Bonferroni adjustment (List et al., 2019).[23]

We next report results of exploratory heterogeneity analysis by characteristics of the school in the village. All but one village had some sort of school in it at baseline. We conduct our analyses based on the number of teachers in the village, the type of school in the village, the highest grade taught in the school, and the quality of the school infrastructure, proxied by the material of its roof. We show these results in Table A.4. We find no evidence of mean-

[23] We also pre-specified heterogeneity tests by the village's distance to the main road, whether the child most commonly speaks Crioulo, as opposed to other languages, and whether there was an economic shock to the main breadwinner of the child's family during the course of the trial. We found no evidence of heterogeneity on these dimensions and do not present these analyses here for the sake of brevity.

Table 8
Enrollment and progression in school.

Panel A: Child is enrolled in school

Date of measurement	(1) Intervention (N)	(2) Control (N)	(3) Adjusted difference (SE)	(4) Odds ratio (95% CI)	(5) *p*-value
At midline (2015)	96.82%	63.96%	31.68%	15.27	$p < 0.001$
(*N*: *I* = 629, *C* = 1,379)	(609)	(882)	(3.84)	(9.16, 25.46)	
At endline (2017)	97.05%	84.72%	10.90%	5.00	$p < 0.001$
(*N*: *I* = 611, *C* = 1,354)	(593)	(1,148)	(2.54)	(2.48, 10.07)	

Panel B: Child's grade in school at endline

			Effect of intervention on probability child is in grade 2 or higher at endline	
Grade in school	(1) Number of intervention children	(2) Number of control children	(3) Estimated effect (SE)	(4) *p*-value
Not enrolled	18	71	65.5% (3.79)	$p < 0.001$
Pre-school	2	29		
Grade 1	11	743		
Grade 2	20	254		
Grade 3	527	43		
Grade 4 or 5	15	8		
Number of observations	593	1,148		

Notes: in Panel A, we show the proportion of students enrolled in school, in each group, at the time of midline and endline surveys. Column 3 shows the adjusted difference as in earlier tables, column 4 shows the odds ratio, and column 5 shows the *p*-value for a test of the null hypothesis of equal enrollment across treatment groups, as was pre-specified. In Panel B, we show the grade in which children were enrolled in school at the time of the endline survey. In the right of the table, we show our exploratory (not pre-specified) estimate of the effect of the intervention on the probability a child is enrolled in at least grade 2 at endline using our main specification and the *p*-value for a test of the null hypothesis that there was no effect.

ingful heterogeneity in the effect of the intervention along any of these dimensions, consistent with the consensus from prior work showing that, in rural areas like those we study, existing variation in school type, school resources, and even teacher credentials generate very little variation in student learning levels (Daun, 1997; Boone et al., 2014; Silva and Oliveira, 2017).

4.5. Other effects

In this section we discuss the impact of the intervention on children's enrollment in school and their grade progression. In Table 8, we report a pre-specified analysis of enrollment in school and an exploratory analysis of grade progression. We first estimate the impact of the intervention on the proportion of children in each randomization group enrolled in school at the midline and endline of the study. At midline in 2014, approximately 97% of intervention children were enrolled in school, while only 64% of control children were. This gap narrows at endline in 2016, driven largely by an increase in enrollment among the control group: 97% of intervention children were enrolled in school at the end of the trial, while 85% of control children were.

These differences are both statistically significant. We see the intervention also has a large impact on grade progression. In Panel B we show that, at endline, intervention children are 65.5 percentage points more likely to be enrolled in at least the second grade, relative to control children.

We also collected parents' report of whether or not the child missed any school in the past two weeks at the midline and endline surveys. In Fig. A.3, we show these results, which suggest that intervention children are much less likely than control children to miss school in both AY 2014–15 and AY 2016–17. Because we are missing attendance data for many of these children, particularly for controls, we have put these particular results in the appendix and urge caution in their interpretation.

At endline, we collected information from the child's nearest older sibling and nearest younger sibling about their enrollment in school up to that point. We also administered simple ASER-style reading and math tests (Pratham, 2010). We were only able to locate siblings in between 25 and 40 percent of cases. Of the siblings we did find, we found little difference in enrollment in school (see Table A.5). Nonetheless, among these children we found significantly higher literacy and numeracy among the intervention group for both older and younger siblings. We show these differences in Fig. A.4. This suggests potential spillovers of learning to siblings, with two important caveats. First, the magnitudes of the differences are very small compared to the differences we find for study children. Second, because roughly 70 percent of siblings were not found, we are hesitant to draw strong conclusions from these analyses.

4.6. Benefit–cost analysis

We estimate that this intervention would cost approximately US $1,700 per child to run for four years; equivalently, the per-child, per-year cost is roughly $425.[24] While this is a very highly-resourced intervention relative to others in this literature, such as those described in Kremer et al. (2013), it achieves learning gains of unprecedented magnitude in an exceedingly challenging environment.

We provide a rough estimate of a lower bound for the benefit–cost ratio of this intervention (Levin et al., 2017). To generate our assumption about the per-person benefit, we need an approximation of the income premium that achieving literacy and numeracy might yield later in life. To generate this, we use the following assumptions. One, using estimates from Table 4, we assume that the intervention generates a 62.9 percentage point increase in the likelihood a child will be literate. Two, we assume that, as a result, the child's future employment is characterized by the following probability set: they continue subsistence farming (30%

[24] To calculate the cost of the intervention, we use the projected costs for the ongoing (at time of writing) expansion of the project. We chose this instead of the actual costs incurred during the implementation of this study because of the costs incurred during the previously described challenges with early implementation. Without dramatic assumptions, it is not clear how to extract the "true" costs of the final project from those data (e.g., the "right-sizing" of administration, procurement, and other costs for this smaller scale).

chance), they work in their village for a local NGO (30% chance) they become a community teacher (30% chance), or they progress in school until the 12th grade, at which point they gain employment in a national NGO (10% chance).[25] We estimate the lifetime gain in income, over a baseline of subsistence farming with certainty, given current salaries for these positions[26], and assuming a 5% annual GDP growth rate (The World Bank, 2019) and a 5% annual discount rate (Duflo, 2001). Finally, we assume that affected individuals work from age 17 to age 55, during which time they earn the income benefit assumed above.

Using these assumptions, the intervention has a benefit–cost ratio of at least 3.12. We expect this to be a lower bound on the true ratio, given the various, harder-to-estimate returns to literacy and numeracy that accrue in health, longevity, and welfare more broadly (Dickson and Harmon, 2011). This ratio suggests the intervention is highly cost-efficient, and compares favorably with many other studies in similar contexts (Evans and Popova, 2016).

An increasingly common approach to this type of analysis is to calculate the "marginal value of public funds" or MVPF (Hendren and Sprung-Keyser, 2020). This calculates the after-tax benefit to participants, accounting for changes in tax revenue because of the program. These changes can be negative (e.g., distorting behavior away from productive activity in order to qualify for the program) or positive (e.g., generating externalities). In Guinea Bissau, the effective tax rate is zero for most people, as most government revenue comes from two sources: cashew nut exports and foreign aid. We assume, therefore, that there are no negative externality-type changes in revenue that would accrue from implementing this policy. The likely positive externalities of the policy – greater economic, health, and political benefits from a higher literacy rate – make our benefit–cost calculation a lower bound on the true benefit–cost ratio.

5. Features, uniqueness, and scalability of the intervention

In this section, we discuss potential explanations for the large magnitude of the results we find, describing what features of the intervention are unique and its potential for scalability.

We think there are two core reasons for the large impacts we observe. First, the intervention's focus was on child learning, as opposed to test score improvement or child or teacher attendance. All implementers, from teachers to monitors to senior staff, understood that learning was the main objective. This focus informed the design of all teaching and learning materials, from textbooks to teacher handbooks and lesson plans. These materials also incorporated scripted lessons, which have been shown to work in numerous settings (Piper et al., 2018; Romero et al., 2020; Eble et al., 2021) and are alleged to be particularly helpful for teachers with less training and suboptimal supervision, potentially raising the level of the "floor" of teaching quality in challenging contexts. The absence of heterogeneity in the treatment effect shows the intervention worked similarly for all children. This is a common feature of scripted lessons (Muralidharan et al., 2019) and suggests the important contribution of scripting in generating the effects we estimate.

Second, we conducted regular, in-depth, and responsive monitoring of both student learning and, separately, teaching. This is in stark contrast to the control condition, where there is little monitoring of teaching or student learning. Monitoring focused on improving teaching skill, not just teacher attendance. The intervention invested heavily in teachers, including three months of preservice training in how to use the intervention's pedagogical model and materials and ongoing training in how to teach new content using new lesson plans. The intervention employed two tiers of monitoring staff who observed teachers, provided feedback, and used these lessons to guide subsequent training. This pairing of monitoring and training with the goal of improving teacher practice has previously yielded large improvements in learning across diverse settings (Piper et al., 2018; Eble et al., 2021), and likely made an important contribution to the intervention's success. In addition, the intervention team measured child learning regularly through in-class tests and periodic external testing,[27] monitoring results and giving teachers of lagging students extra attention or assistance. The team also worked with parents to ensure child attendance in classes and provided additional after-school support to struggling students. Other clear contributors include increased instructional time and resources provided, though extra time and money are no guarantee of a large effect (Woessmann, 2016; De De Ree et al., 2018).[28]

We believe that efficient implementation of these core components could lead to quality education in many other contexts, even in the absence of a large influx of resources. As described in Banerjee et al. (2017) and Bold et al. (2018), however, an important challenge is stakeholder buy-in. Implementing such a system would constitute a large change in focus and responsibilities from teaching and support staff, which may meet resistance. Nonetheless, we think that our results provide important guidance on how to proceed in poor, remote areas such as the one we study. Furthermore, our ongoing work shows that this model is scalable.[29] Aside from buy-in, the main barrier to scalability, as we see it, is resources. Implemented outside of the government, this is a highly expensive intervention. Implemented within the government, we anticipate both political and logistical challenges to widespread adoption (c.f., Bold et al., 2018).

There are two other important potential explanations for our results: teaching to the test and test floor effects. We use EGRA and EGMA tests precisely because they focus on the skills necessary to read, make sense of written content, to do arithmetic, and to make sense of simple arithmetic expressions. These skills are aligned with the goals for almost all education systems at this level of learning and, in many other contexts, EGRA and EGMA tests are used by government itself to measure learning (Sprenger-Charolles, 2008; USAID, 2019). The second potential contributor is floor effects, i.e., that the tests were not sensitive enough to pick up very basic skills. EGRA and EGMA tests are designed to be particularly sensitive at measuring low levels of learning (Platas et al., 2014; Dubeck and Gove, 2015). Comparing subtasks where the control group has a substantial amount of nonzero scores provides little evidence of floor effects.[30] The small proportion of absolute

[25] We generated these probabilities based on our understanding of the local labor markets and discussion with project staff.

[26] NGO community worker salary: 15,000 CFA per month for 12 months per year. Community teacher salary: 25,000 per month for 9 months per year. National NGO salary: 100,000 CFA per month for 12 months per year.

[27] These tests were designed in-house and deliberately diverged from EGRA- and EGMA-style tests to ensure that teachers were not "teaching to the test."

[28] Intervention students had an additional hour per day in school, and our schools did not suffer from the teacher strikes that occurred in government schools over the period of our study. The absence of a difference in effect size between villages with and without a government school suggests that more instructional time does not necessarily translate into learning in this context. The long literature on credit constraints in education shows both theoretically and empirically that, in such areas, private provision of education is also particularly likely to under-supply quality (c.f., Becker, 1994; Lochner and Monge-Naranjo, 2012; Heckman and Mosso, 2014).

[29] In response to the preliminary results of this study, we are on track to scale up the intervention in Guinea Bissau to an additional 2,000 children. In Telangana, India, and The Gambia, we have scaled up a para-teacher intervention with similar foci to 15,000 and 4,000 children, respectively.

[30] We generate this "alternate composite score" by calculating the arithmetic mean of average performance on EGMA subtasks 1-4a and on EGRA subtasks 1–2. Using this yields a treatment effect estimate of 54.08 percentage points, as compared to 58.14 percentage points using the original composite score.

zero composite scores shown in Table 4 further suggests that our tests were sufficiently sensitive for measuring learning in this population.

Our study design did not attempt to identify individual mechanisms behind the intervention's effects. Instead, we targeted areas with great need and evaluated a comprehensive intervention to dramatically increase learning levels in them. This "bundled" approach is in the spirit of the multifaceted poverty alleviation program studied by Banerjee et al. (2015). This does not allow us to isolate mechanisms driving the results we observe, though we speculate that there are complementarities between the individual components, as in Mbiti et al., (2019).

At the outset, we were unsure whether such an intervention would work. If demand factors explained most of the lack of schooling – i.e., parents and their children do not believe education merits the opportunity cost – then the poor outcomes of children might not be impacted by changes to the provision of schooling. Furthermore, as we experienced, implementation challenges could have derailed our efforts entirely and it is important to document this. We also show the costs of implementing such a program in an exceedingly deprived and difficult environment. Due to the fragility of the state in Guinea Bissau, public institutions such as schools, customs, and the courts often function poorly or not at all (Sangreman et al., 2018). Working in hard-to-reach, extremely poor regions within Guinea Bissau made provision even more expensive, and logistics more difficult, than in the country's urban or peri-urban areas.

The other main contributor to the large difference between children in control and intervention villages is, sadly, the failure of the state and other actors to deliver education in these areas. Education levels in Guinea Bissau have remained consistently low over the last fifty years, and there is little evidence that, in the absence of external intervention such as the one we study, this is likely to change (Daun, 1997; Boone et al., 2014; Silva and Oliveira, 2017). During the course of our study, government provision of education in the control villages, as in the rest of the country, was of low quality and sometimes erratic. The counterfactual case, therefore, is one in which many children reach adulthood without achieving meaningful levels of literacy and numeracy. We expect that it is easier to raise learning from such a low baseline than it would be in contexts with higher learning levels.

6. Conclusion

In the least fortunate parts of the developing world, many children receive schooling which is unable to teach them even basic literacy and numeracy. We ran an RCT in rural Guinea Bissau to evaluate an intervention that provided schooling in lieu of the state and other status quo providers for four years. We find the intervention yielded dramatic increases in learning among recipient children, leading them to be functionally numerate and literate in a way that the vast majority of them would not have been in the absence of the intervention.

Our findings contain a few core messages. First, we show that offering this kind of an intervention at a near-free price to parents and children in two regions with extremely low learning and economic outcomes leads to a very high proportion of take-up. This suggests that supply constraints may be more important than demand constraints in understanding low educational outcomes in these and similar areas. Second, our results suggest there may be similarly large learning gains that can be realized by motivated donors or agencies through implementing a similar type of intervention in contexts where the status quo provider of education is either irregular or of extremely low quality.

This intervention achieved learning gains of unprecedented magnitude. While the intervention is much more highly-resourced than other interventions in this literature, a rough benefit–cost calculation suggests that, even using conservative assumptions, it is highly cost-efficient. This work, in conjunction with Eble et al. (2021), shows that the upper bound on the magnitude of intervention-driven learning gains in such deprived areas is much larger than previously thought. Finally, our study provides an opportunity to follow these children later in life, and learn about the longer-term economic and social returns to education, and literacy and numeracy more specifically, in a particularly poor region. This, we hope, will advance our understanding of two important phenomena: one, how best to help similar regions; and two, to quantify where, when, and how these basic skills can transform lives in the developing world.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We are grateful to our administrative, research, and implementation teams in Bissau for their tireless work. Ana Forjaz oversaw the fieldwork at endline. Mark Fisher designed and maintained the database. Gilda Piaggio conducted the power calculation and randomization at the beginning of the study. Yixun Zeng provided research assistance in the preliminary stages of the data analysis. Clive Belfield and Sharon Wolf gave helpful comments on the manuscript. We are grateful to the National Bureau of Economic Research (NBER) for funding the exploratory research on learning levels in Guinea Bissau that motivated this study. We thank the editor and two anonymous referees for generous input which greatly improved the work. This study was funded by Effective Intervention, a UK registered charity. The study passed ethical/institutional review by the Ministry of Education of Guinea Bissau on 30 August, 2012, and the NBER ethics committee on 1 July, 2014 (ref: IRB Ref#14_06).

Appendix A

See Figs. A1–A4 and Tables A1–A5.

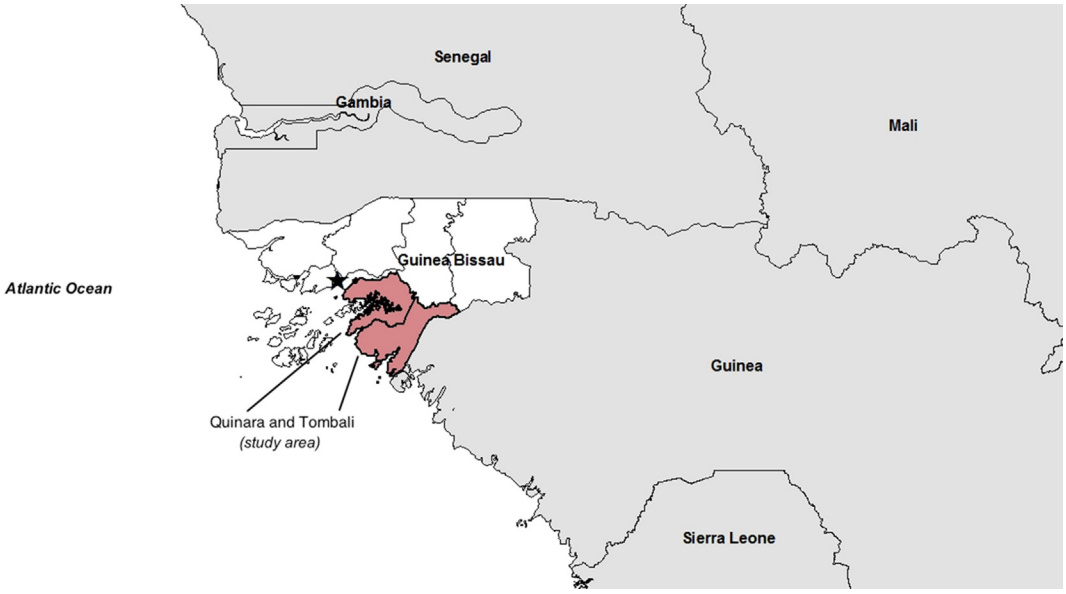

Fig. A1. Map of Guinea Bissau and study area. *Notes:* this figure shows a map of Guinea Bissau and surrounding (not studied) countries, with the regions of Guinea Bissau in white, and the two study regions shaded in red and labeled.

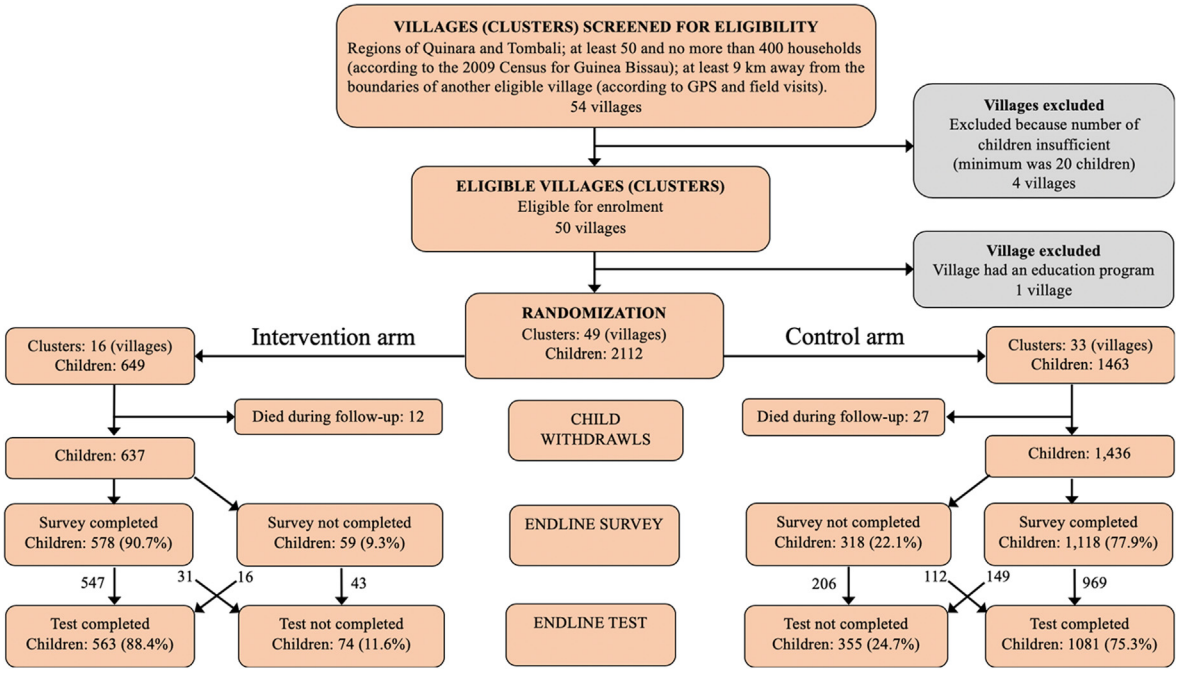

Fig. A2. CONSORT-style diagram of flow of participants through the study. *Notes:* this figure shows how participants (villages and children) flowed through the trial, from screening for eligibility to participation in the endline survey and test.

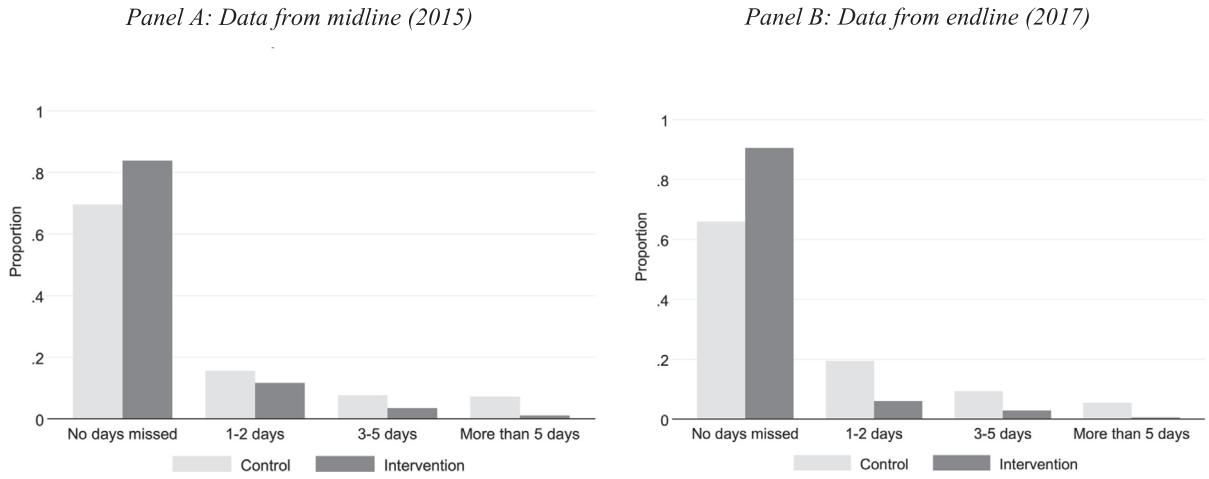

Fig. A3. Attendance in school: number of days missed in last two weeks. *Notes:* This figure shows parents' report of how many days their child missed school in the two weeks prior to being interviewed, separately at the midline and at the endline surveys (in Panels A and B, respectively), and separately by randomization group. We present results only for those children who were enrolled in school at the time of survey. A simple chi-square test rejects the null of no relationship between attendance and intervention status, with p<0.001 in both panels.

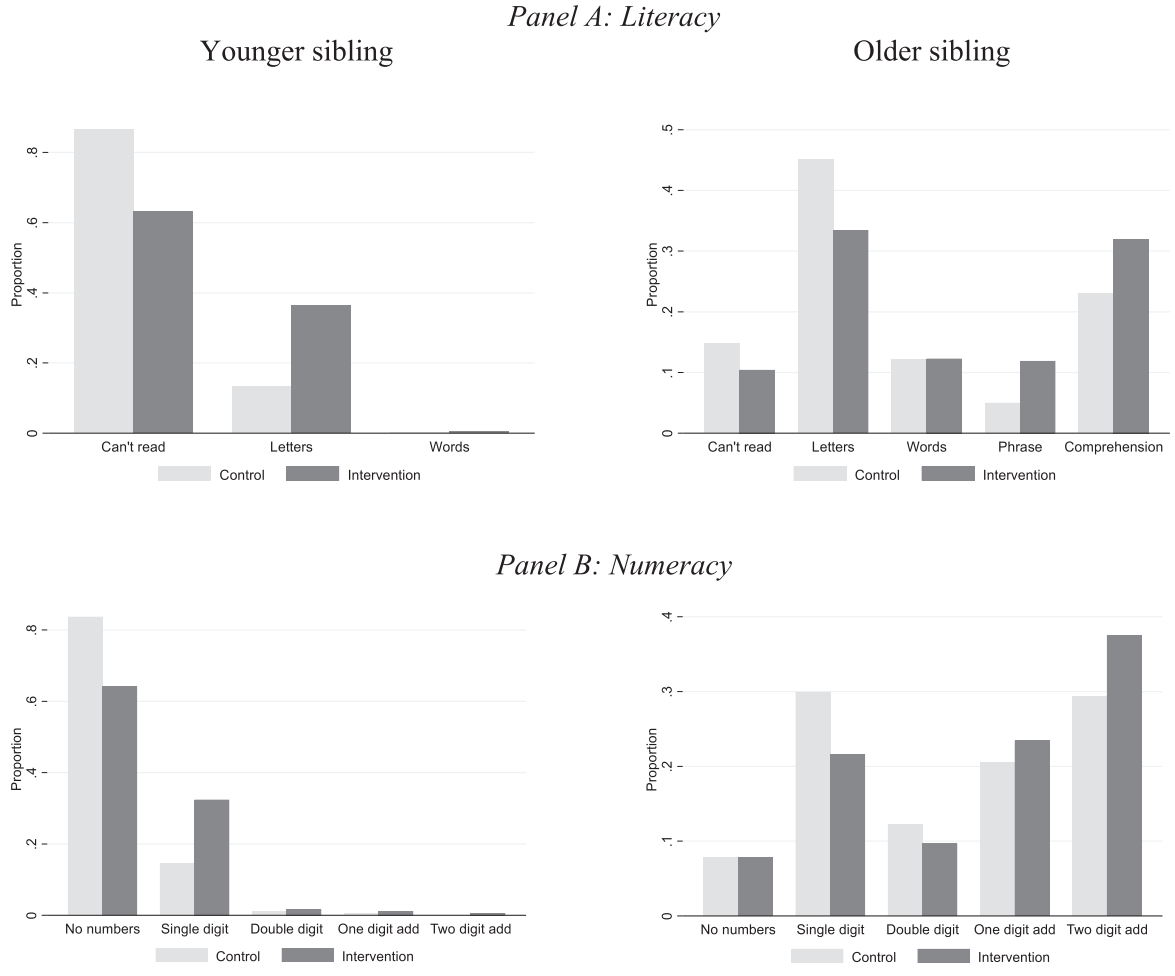

Fig. A4. Sibling literacy and numeracy tests. *Notes:* This figure shows the results of the sibling literacy and numeracy tests administered at the endline for students present in the survey who also had siblings present in the village at the time of the endline survey. There were 362 younger siblings and 521 older siblings found in the control villages, and 176 younger siblings and 269 older siblings in the intervention villages. A simple chi-square test rejects the null of no relationship between literacy and intervention status, with p < 0.001 for both older and younger siblings. It rejects the null of no relationship between numeracy and intervention status with p=0.040 for older siblings, and p < 0.001 for younger siblings.

Table A1
Power calculation

Allocation ratio intervention: control	Loss to follow-up	Average no. of children per cluster after loss to follow-up*	Minimum difference to detect (%)	
			20	**25**
1:1	17%	35	80	94
	25%	32	82	95
1:2	17%	35	75	91
	25%	32	77	92

Notes: Power obtained with a two-sided 5% level test with 49 clusters total, assuming an Intra-cluster Coefficient =0.03. *: the assumed average number of children per cluster before loss to follow up is 43.

Table A2
Description of subtasks.

EGRA	EGMA
1: Read a letter's sound (e.g., "oh" for o)	1: Read a number (e.g., 2, 9, 45)
2: Differentiate sounds (e.g., which word starts with a different sound: casa, livro, or cama)	2: Choose the larger number (e.g., 7 or 5)
3: Read a made-up word (e.g., tila)	3: Complete a sequence (e.g., 14 15 16 __)
4: Read a familiar (Portuguese) word (e.g., sol)	4a: Simple addition (e.g., 1+3)
	4b: Two-digit addition (e.g., 14+25)
5a: Read a short passage	5a: Simple subtraction (e.g., 5–2)
5b: Answer questions on the passage's content	5b: Two-digit subtraction (e.g., 26–14)
6: Listen to a different short passage, answer questions on the passage's content	6: Solve a simple word problem read aloud

Notes: this table provides descriptions of the different types of questions asked on the reading (EGRA) and math (EGMA) tests, respectively. These are referred to as "tasks" or "subtasks", by the number given in this table.

Table A3
Attendance of enrolled children in intervention classes.

	(1) **Attendance (N)**
Mean	85.72%
SD	30.80%
Distribution of attendance	
0% of classes	9.27% (60)
>0 to 25% of classes	1.24% (8)
>25% to 50% of classes	2.32% (15)
>50% to 75% of classes	2.01% (13)
>75% to 100% of classes	85.16% (551)
Missing data	0.31% (2)
Number of non-missing observations	647

Notes: this table shows the average attendance of children in the intervention arm at intervention classes, as a proportion of total classes held. The number of observations corresponding to these proportions are given in parentheses.

Table A4
Heterogeneity of effect by village school traits.

Group	(1) Intervention (SD)	(2) Control (SD)	(3) Adjusted difference (SE)	(4) *p*-value
Highest grade taught in village				
Third or fourth grade	71.34	12.60	57.71	0.77
(*N*: *I* = 337, C = 459)	(15.24)	(12.06)	(0.92)	
Fifth grade or higher	69.20	10.32	58.10	
(*N*: *I* = 226, C = 607)	(15.45)	(9.95)	(0.98)	
Total number of teachers in village				
One or two teachers	69.38	10.89	58.08	0.19
(*N*: *I* = 393, C = 932)	(14.73)	(10.35)	(0.75)	
Three or four teachers	73.03	14.16	55.86	
(*N*: *I* = 170, C = 134)	(16.44)	(14.27)	(1.53)	
Lowest quality material of school roof				
Roof is natural	72.51	13.60	57.73	0.80
(*N*: *I* = 122, C = 48)	(13.53)	(12.01)	(0.73)	
Roof is synthetic	69.92	11.20	58.29	
(*N*: *I* = 441, C = 1018)	(15.78)	(10.90)	(2.13)	
Presence of public school in village				
No public school in village	67.94	12.04	56.77	0.21
(*N*: *I* = 179, C = 284)	(15.89)	(11.31)	(1.21)	
Public school in village	71.66	11.04	58.67	
(*N*: *I* = 384, C = 782)	(14.96)	(10.83)	(0.87)	
Presence of community school in village				
No community school in village	69.62	11.05	57.86	0.91
(*N*: *I* = 416, C = 892)	(14.62)	(10.47)	(0.75)	
Community school in village	72.91	12.61	57.69	
(*N*: *I* = 147, C = 174)	(17.05)	(13.16)	(1.46)	

Notes: this table shows exploratory estimates of heterogeneity in the effect of the intervention on composite test scores by the characteristics of the schools in the village, following the convention of Table 7. There is only one village in our study which does not have a school in the village, and we exclude it from this analysis.

Table A5
Sibling enrollment in school.

Group	(1) Intervention (SD)	(2) Control (SD)	(3) Adjusted difference (SE)	(4) *p*-value
Older sibling enrolled in school	0.892	0.923	−0.050	0.023
(*N*: *I* = 269, C = 521)	(0.311)	(0.266)	(0.022)	
Younger sibling enrolled in school	0.636	0.556	0.013	0.777
(*N*: *I* = 176, C = 363)	(0.482)	(0.497)	(0.046)	

Notes: this table shows the levels of enrollment of the child's next-younger and next-older siblings in school, and tests for differences across treatment group, following the convention of Table 4.

Appendix B. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jpubeco.2021.104385.

References

Angrist, Joshua D., Pathak, Parag A., Walters, Christopher R., 2013. Explaining charter school effectiveness. Am. Econ. J.: Appl. Econ. 5 (4), 1–27.

Banerjee, Abhijit, Banerji, Rukmini, Berry, James, Duflo, Esther, Kannan, Harini, Mukerji, Shobhini, Shotland, Marc, Walton, Michael, 2017. From proof of concept to scalable policies: challenges and solutions, with an application. J. Econ. Perspect. 31 (4), 73–102.

Banerjee, Abhijit, Duflo, Esther, Goldberg, Nathanael, Karlan, Dean, Osei, Robert, Parienté, William, Shapiro, Jeremy, Thuysbaert, Bram, Udry, Christopher, 2015. A multifaceted program causes lasting progress for the very poor: evidence from six countries. Science 348 (6236).

Becker, Gary S., 1994. Human Capital. University of Chicago press, Chicago, IL.

Blimpo, Moussa P., Evans, David K., Lahire, Nathalie, 2011. School-based management and educational outcomes: lessons from a randomized field experiment. Unpublished Manuscript.

Bold, Tessa, Kimenyi, Mwangi, Mwabu, Germano, Ng'ang'a, Alice, Sandefur, Justin. 2018. Experimental evidence on scaling up education reforms in Kenya. J. Public Econ. 168 (December): 1–20.

Boone, Peter, Fazzio, Ila, Jandhyala, Kameshwari, Jayanty, Chitra, Jayanty, Ganghadar, Johnson, Simon, Ramachandran, Vimala, Silva, Ana Filipa, Zhan, Zhaoguo, 2014. The surprisingly dire situation of children's education in rural West Africa: results from the CREO study in Guinea-Bissau (Comprehensive Review of Education Outcomes). In: African Successes, Volume II: Human Capital. University of Chicago Press, Chicago, IL, pp. 255–280.

Bruhn, Miriam, McKenzie, David, 2009. In pursuit of balance: randomization in practice in development field experiments. Am. Econ. J.: Appl. Econ., 200–232

Burde, Dana, Linden, Leigh, L., 2013. Bringing education to Afghan girls: a randomized controlled trial of village-based schools. Am. Econ. J.: Appl. Econ. 5 (3), 27–40.

Campbell, Frances A., Ramey, Craig T., 1994. Effects of early intervention on intellectual and academic achievement: a follow-up study of children from low-income families. Child Dev. 65 (2), 684–698.

Campbell, Frances A., Ramey, Craig T., 1995. Cognitive and school outcomes for high-risk African-American students at middle adolescence: positive effects of early intervention. Am. Educ. Res. J. 32 (4), 743–772.

Campbell, Marion K., Piaggio, Gilda, Elbourne, Diana R., Altman, Douglas G., 2012. Consort 2010 statement: extension to cluster randomised trials. BMJ 345, e5661.

Chabrier, Julia, Cohodes, Sarah, Oreopoulos, Philip, 2016. What can we learn from charter school lotteries? J. Econ. Perspect. 30 (3), 57–84.

Chaudhury, Nazmul, Hammer, Jeffrey, Kremer, Michael, Muralidharan, Karthik, Rogers, F. Halsey, 2006. Missing in action: teacher and health worker absence in developing countries. J. Econ. Perspect. 20 (1), 91–116.
Daun, Holger, 1997. Teachers' needs, culturally-significant teacher education and educational achievement in an African context—the case of Guinea-Bissau. Int. J. Educ. Dev. 17 (1), 59–71.
De Ree, Joppe, Muralidharan, Karthik, Pradhan, Menno, Rogers, Halsey, 2018. Double for nothing? Experimental evidence on an unconditional teacher salary increase in Indonesia. Q. J. Econ. 133 (2), 993–1039.
Dickson, Matt, Harmon, Colm, 2011. Economic returns to education: what we know, what we don't know, and where we are going—some brief pointers. Econ. Educ. Rev. 30 (6), 1118–1122.
Dobbie, Will, Fryer Jr, Roland G., 2013. Getting beneath the veil of effective schools: evidence from New York City. Am. Econ. J.: Appl. Econ. 5 (4), 28–60.
Dubeck, Margaret M., Gove, Amber, 2015. The early grade reading assessment (EGRA): its theoretical foundation, purpose, and limitations. Int. J. Educ. Dev. 40, 315–322.
Duflo, Esther, 2001. Schooling and labor market consequences of school construction in Indonesia: evidence from an unusual policy experiment. Am. Econ. Rev. 91 (4), 795–813.
Eble, Alex, Frost, Chris, Camara, Alpha, Bouy, Baboucarr, Bah, Momodou, Sivaraman, Maitri, Hsieh, Jenny, et al., 2021. 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. J. Dev. Econ. 148, 102539.
Evans, David K., Popova, Anna, 2016. What really works to improve learning in developing countries? An analysis of divergent findings in systematic reviews. The World Bank Res. Observer 31 (2), 242–270.
Glewwe, Paul, Muralidharan, Karthik, 2016. Improving education outcomes in developing countries: evidence, knowledge gaps, and policy implications. In: Handbook of the Economics of Education, 5. Elsevier, Amsterdam, Holland, pp. 653–743.
Heckman, James J., Moon, Seong Hyeok, Pinto, Rodrigo, Savelyev, Peter A., Yavitz, Adam, 2010. The rate of return to the Highscope Perry preschool program. J. Public Econ. 94 (1), 114–128.
Heckman, James J., Mosso, Stefano, 2014. The economics of human development and social mobility. Annu. Rev. Econ. 6 (1), 689–733.
Heckman, James, Pinto, Rodrigo, Savelyev, Peter, 2013. Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. Am. Econ. Rev. 103 (6), 2052–2086. https://doi.org/10.1257/aer.103.6.2052.
Hendren, Nathaniel, Sprung-Keyser, Ben, 2020. A unified welfare analysis of government policies. Q. J. Econ. 135 (3), 1209–1318.
Jordan, Nancy C., Kaplan, David, Ramineni, Chaitanya, Locuniak, Maria N., 2009. Early math matters: kindergarten number competence and later mathematics outcomes. Dev. Psychol. 45 (3), 850–867.
Kremer, Michael, Brannen, Conner, Glennerster, Rachel, 2013. The challenge of education and learning in the developing world. Science 340 (6130), 297–300.
Lakshminarayana, Rashmi, Eble, Alex, Bhakta, Preetha, Frost, Chris, Boone, Peter, Elbourne, Diana, Mann, Vera, 2013. The support to rural India's public education system (STRIPES) trial: a cluster randomised controlled trial of supplementary teaching, learning material and material support. PLoS ONE 8 (7). e65775.
Lee, David S., 2009. Training, wages, and sample selection: estimating sharp bounds on treatment effects. Rev. Econ. Stud. 76 (3), 1071–1102.
Lepri, Jean-Pierre, 1988. Formação de Professores, Locais, Materiais Escolares e Insucesso Escolar. Soronda: Revista de Estudos Guineenses 5, 83–102.
Levin, Henry M., McEwan, Patrick J., Belfield, Clive, Bowden, A. Brooks, Shand, Robert, 2017. Economic Evaluation in Education: Cost-Effectiveness and Benefit-Cost Analysis. SAGE Publications, Thousand Oaks, CA.
List, John A., Shaikh, Azeem M., Yang, Xu, 2019. Multiple hypothesis testing in experimental economics. Exp. Econ. 22 (4), 773–793.
Lochner, Lance, Monge-Naranjo, Alexander, 2012. Credit constraints in education. Ann. Rev. Econ. 4 (1), 225–256.
Mann, Vera, Fazzio, Ila, King, Rebecca, Walker, Polly, dos Santos, Albino, Carlos de Sa, Jose, Jayanty, Chitra, Frost, Chris, Elbourne, Diana, Boone, Peter, 2009. The EPICS Trial: enabling parents to increase child survival through the introduction of community-based health interventions in rural Guinea Bissau. BMC Public Health 9 (1), 1–12.
Mbiti, Isaac, Muralidharan, Karthik, Romero, Mauricio, Schipper, Youdi, Manda, Constantine, Rajani, Rakesh, 2019. Inputs, incentives, and complementarities in education: experimental evidence from Tanzania. Q. J. Econ. 134 (3), 1627–1673. https://doi.org/10.1093/qje/qjz010.
McEwan, Patrick J., 2015. Improving learning in primary schools of developing countries: a meta-analysis of randomized experiments. Rev. Edu. Res. 85 (3), 353–394.
Muralidharan, Karthik, Singh, Abhijeet, Ganimian, Alejandro, 2019. Disrupting education? Experimental evidence on technology-led education in India. Am. Econ. Rev. 109 (4), 1426–1460.
Piper, Benjamin, Sitabkhan, Yasmin, Mejía, Jessica, Betts, Kellie, 2018. Effectiveness of teachers' guides in the global south: scripting, learning outcomes, and classroom utilization.. Occasional Paper. RTI Press Publication OP-0053-1805. RTI International, Research Triangle Park, NC.
Platas, L.M., Ketterlin-Gellar, L., Brombacher, A., Sitabkhan, Y., 2014. Early Grade Mathematics Assessment (EGMA) Toolkit. RTI International, Research Triangle Park, NC.
Pratham, 2010. Annual Status of Education Report (Rural) 2010. http://www.pratham.org/aser08/ASER_2010_Report.pdf.
Pritchett, Lant, 2013. The Rebirth of Education: Schooling Ain't Learning. CGD Books. http://booksgoogle.com/books?hl=en&lr=&id=PQ72AAAAQBAJ&oi=fnd&pg=PR1&dq=pritchett+schooling+aint+learning&ots=uvSg4RtJhA&sig=1jSzmH3E1acmSrT3eRBDQCyjXwA.
Ray, Debraj, Robson, Arthur, 2018. Certified random: a new order for coauthorship. Am. Econ. Rev. 108 (2), 489–520.
RTI International, 2009. Early Grade Reading Assessment Toolkit.
RTI International, 2017. All children reading-Asia: EGRA benchmarks and standards research report. 2017. https://shared.rti.org/content/all-children-reading-asia-egra-benchmarks-and-standards-research-report.
Romero, Mauricio, Sandefur, Justin, Sandholtz, Wayne Aaron, 2020. Outsourcing education: experimental evidence from Liberia. Am. Econ. Rev. 110 (2), 364–400.
Sangreman, Carlos, Delgado, Fátima, Vaz Martins, Luis, 2018. Guinea-Bissau (2014-2016). An empirical study of economic and social human rights in a fragile state. Adv. Social Sci. Res. J. 5 (2), 697–711.
Silva, Rui da, Oliveira, Joana, 2017. 40 Years of educational research in Guinea-Bissau: mapping the terrain. Int. J. Educ. Dev. 57, 21–29.
Sprenger-Charolles, Liliane, 2008. "The Gambia : Early Grade Reading Assessment." World Bank Policy Report. https://openknowledge.worldbank.org/handle/10986/12972.
The World Bank, 2019. "Economic Data on Guinea Bissau" Website accessed October 28, 2019. URL: https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?locations=GW, .
USAID, 2019. About EGRA: Early grade reading assessment. 2019. https://www.earlygradereadingbarometer.org/pages/about_egra.
Woessmann, Ludger, 2016. The importance of school systems: evidence from international differences in student achievement. J. Econ. Perspect. 30 (3), 3–32.