COLUMBIA | SIPA
Center for Development Economics and Policy

**CDEP-CGEG WORKING PAPER SERIES**

CDEP-CGEG WP No. 79

# How Much Can We Remedy Very Low Learning Levels in Rural Parts of Low-Income Countries?

Alex Eble, Chris Frost, Alpha Camara, Baboucarr Bouy, Momodou Bah, Maitri Sivaraman, Jenny Hsieh, Chitra Jayanty, Tony Brady, Piotr Gawron, Peter Boone, Diana Elbourne

COLUMBIA | SIPA
Center on Global Economic Governance

# 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

Alex Eble, Chris Frost, Alpha Camara, Baboucarr Bouy, Momodou Bah, Maitri Sivaraman, Jenny Hsieh, Chitra Jayanty, Tony Brady, Piotr Gawron, Peter Boone, Diana Elbourne*

November 2019

**Abstract**

Despite large schooling and learning gains in many developing countries, children in highly deprived areas are often unlikely to achieve even basic literacy and numeracy. We study how much of this problem can be resolved using a multi-pronged intervention combining several distinct interventions known to be effective in isolation. We conducted a cluster-randomized trial in The Gambia evaluating a literacy and numeracy intervention designed for primary-aged children in remote parts of poor countries. The intervention combines para teachers delivering after-school supplementary classes, scripted lesson plans, and frequent monitoring focusing on improving teacher practice (coaching). A similar intervention previously demonstrated large learning gains in a cluster-randomized trial in rural India. After three academic years, Gambian children receiving the intervention scored 46 percentage points (3.2 SD) better on a combined literacy and numeracy test than control children. This intervention holds great promise to address low learning levels in other poor, remote settings.

* Eble (corresponding author): Teachers College, Columbia University. Email: eble@tc.columbia.edu. Bah (newsbah@yahoo.com), Boone (pb@effint.org), Bouy (bbouy@effint.net), Camara (a.camara@effint.net), and Sivaraman (maitri.effectiveintervention@gmail.com): Effective Intervention, UK and The Gambia. Hsieh (ptjhsieh@gmail.com): Center for Comparative and International Education, Department of Education, University of Oxford. Elbourne (Diana.Elbourne@lshtm.ac.uk) and Frost (chris.frost@lshtm.ac.uk): London School of Hygiene and Tropical Medicine, UK. Jayanty (chitra.jayanty@gmail.com): independent consultant. Brady and Gawron: Sealed Envelope, Ltd. (support@sealedenvelope.com). We would like to acknowledge the contributions of our team in The Gambia, particularly Lenin Balan, Lamin Janneh, Alieu Jibba, Jenieri Sagnia, and Abhishek Thakur, and the database designed and managed by Piotr Gawron and Tony Brady at Sealed Envelope Ltd. This study was approved by the Institutional Review Board of the London School of Hygiene and Tropical Medicine, protocol number 8767. Funding for this study was provided by Effective Intervention, a UK-registered charity.

## 1. Introduction

While children in developing countries are much more likely than before to be in school, in dozens of countries, they are still highly unlikely to acquire the skills expected of them at each grade level (Bold et al. 2017; Pritchett 2013). Learning levels are often even lower in rural areas of these countries (Chaudhury et al. 2006; Glewwe 2002). Reviews show that hundreds of studies have evaluated a wide range of different interventions attempting to raise learning levels in these contexts (Evans and Popova 2016; Ganimian and Murnane 2016; McEwan 2015). Among the many studies yielding positive results, the majority find modest test score or ability changes, usually in the range of 0.1-0.5 test score standard deviations, or SDs (Kremer, Brannen, and Glennerster 2013). This suggests that, to date, we know very little about how to generate the type of large gains necessary to close the learning gap between developing and developed countries (Glewwe and Muralidharan 2016).

In this paper, we report results from a randomized controlled trial asking the following research question: if we deliver a multi-pronged, well-resourced intervention to children in a highly deprived setting, how much of this learning gap can we close? The intervention we study combines three well-known levers for improving learning: i) the use of para teachers, instead of civil servants or volunteers, to deliver after school lessons (Banerjee et al. 2007; Duflo, Dupas, and Kremer 2015; Lakshminarayana et al. 2013; Muralidharan and Sundararaman 2013); ii) an improved, scripted curriculum targeted at students' current learning levels (Banerjee et al. 2007; 2017; Lakshminarayana et al. 2013; Piper, Zuilkowski, and Mugenda 2014); and iii) extensive monitoring of these teachers, the aim of which was to provide regular feedback on teaching methods and practice (Kraft, Blazar, and Hogan 2018; Muralidharan et al. 2017). This intervention was originally designed by The Naandi Foundation, an Indian non-governmental

organization, and raised learning levels among primary-aged children by 0.75 SD in rural Telangana, India after two years of implementation (Lakshminarayana et al. 2013). In this study, we partly attempt to learn whether the gains in learning achieved by this intervention in India can, with appropriate adaptation, be realized in a new, more challenging context: rural areas of The Gambia.

We find that this approach – using a well-resourced, multi-pronged intervention to target literacy and numeracy of children in particularly deprived areas – generates extremely large learning gains in rural Gambia. Children receiving the intervention scored 46 percentage points (3.2 SD) better on a composite literacy and numeracy test than children in the control villages. For comparison, these rural Gambian children performed as well or better in all comparable English reading skill tests than children in a nationally representative assessment of these skills among third grade students in the Philippines, a country with a per-capita GDP several times greater than that of The Gambia.[1] Our results demonstrate that, in this type of area, the upper limit for learning gains from aggressive interventions such as ours is far higher than previously shown.

We worked in 169 small villages in the two central regions of The Gambia. We began with a census of 6 to 8 year-olds in these villages whose caregivers planned to enroll them in primary school the following year, and we followed these children for the next three years. Villages were randomized in clusters, with half receiving the intervention and half not. Villages assigned to receive the intervention received it from early January 2016 to early May 2018. Delivery of the intervention constituted recruiting, training, and deploying para teachers to deliver scripted, supplementary lessons in mathematics and reading for 12 hours per week. The

[1] EGRA data for the Philippines come from https://earlygradereadingbarometer.org/overview, accessed October 16, 2019. GDP data come from https://data.worldbank.org/indicator/NY.GDP.PCAP.CD, accessed on the same date.

intervention was adapted from that in Lakshminarayana et al (2013) to the Gambian setting, advancing with children through the first three years of the official primary school curriculum of The Gambia. The para teachers were given pre-service training in both pedagogical content knowledge and subject matter (Ding, He, and Leung 2014; Even 1993), and later given further training on how to teach concepts as children progressed through the curriculum. They were regularly monitored throughout the trial and provided with regular feedback to improve their teaching methods, a practice commonly referred to as "coaching" (Kraft, Blazar, and Hogan 2018).

The pre-specified primary outcome of the trial was the arithmetic mean of the child's score on Early Grade Reading and Mathematics Assessment-style tests (also known as EGRA and EGMA tests: Dubeck and Gove 2015, Platas et al. 2014) administered at the end of the trial. Attrition from enrollment in the trial to the endline test was less than 14%. The magnitude of the difference between control and intervention child scores is similar for the mathematics and reading tests, respectively.

Each of these two tests comprises subtasks of varying difficulty. Performance across subtasks shows that intervention children had substantially higher scores than control children on tasks that are relatively simple (e.g., letter recognition and number recognition), moderately difficult (e.g., familiar word reading and single-digit addition), and more difficult (e.g., reading comprehension and two- or three-digit subtraction with borrowing).

Performance by subtask also shows one possible reason for the large gains we measure: the typical third grade child in these areas – as represented by children in the control group – has extremely low levels of literacy and numeracy. This pattern also appears in regular national assessments of child learning using EGRA- and EGMA-style tests administered by the Gambian

government and by third parties (e.g., Sprenger-Charolles 2008). Among enrolled children in our control villages evaluated at the end of the study, more than two thirds of children could not read a single short (two- to seven-letter) word taken from the second-grade curriculum, and half could not successfully complete even one single-digit addition problem. These learning levels are substantially worse than those observed in similar assessments of children's learning in other developing countries, such as India and Tanzania (Pratham 2010; Rajani 2010) and similar to what some of us have observed in recently completed work in rural Guinea Bissau.

In the discussion, we address two questions: one, why are our effect estimates so large, and two, what can we learn from these results about policy options for raising learning levels in these areas? We argue that the dramatic increase in learning we observe in the intervention group is likely to be due to two main factors. The first factor is that the intervention combines several tools – para teachers delivering after school lessons; an improved, scripted, and targeted curriculum; and extensive monitoring of these para teachers with an emphasis on pedagogical improvement – known to be effective in isolation (e.g., Banerjee et al. 2017; Kraft et al. 2018; Muralidharan et al. 2017; Muralidharan and Sundararaman 2013; Piper et al. 2018).[2] The second factor is the very low learning levels of children in rural parts of The Gambia, even among those enrolled in school. This allows for the potential for exceptionally large relative improvements (Blimpo, Evans, and Lahire 2011). Burde and Linden (2013) find that a supply-side intervention yields learning gains of 0.4-0.65 SD in rural Afghanistan, a setting with comparably low baseline literacy and numeracy.

[2] This combination of multiple known best practices, offered through a contracted provider, parallels solutions from studies of the provision of healthcare in similarly remote and disadvantaged areas (Salehi et al. 2018) and of "graduation" programs helping individuals out of extreme poverty (Banerjee et al. 2015).

Our findings have two clear implications for policy. The first is that there exists a demonstrated way to reach the learning gains that many have called for in particularly disadvantaged areas (Pritchett 2013; Glewwe and Muralidharan 2016). Our study, in conjunction with Lakshminarayana et al. (2013), shows that this model of intervention can achieve dramatic learning gains in remote, rural parts of developing countries. More broadly, our results suggest that the upper bound for the magnitude of intervention-driven learning gains in similar settings is much higher than previously observed (McEwan 2015). The second implication is that the choice facing governments and donors who wish to reap such gains is whether to attempt to operationalize this within the government system, or contract it out.

We argue that these findings also provide evidence on the relative importance of supply of and demand for education in explaining the low levels of learning in such settings. The low outcomes we observe could reflect very poor schools, or reflect the lack of demand for education. We interpret our main empirical finding – the large learning gains we measure in response to an experimentally induced change in supply – as evidence that the poor outcomes in The Gambia are at least primarily supply-driven.[3] We also contribute to the growing set of studies looking at the replication, scalability, and generalizability of results from RCTs, particularly those that try to raise learning levels in developing countries (Banerjee et al. 2017; Bold et al. 2018).

The rest of the paper proceeds as follows: in Section 2 we describe the intervention we study and our setting. In Section 3 we provide an overview of our research design. In Section 4 we present our main results for our pre-specified primary and secondary outcomes. In Section 5, we discuss the reliability and implications of our results. Section 6 concludes. In the Appendices,

[3] This echoes other work from The Gambia showing that a major fee reduction in secondary schools increased girls' persistence in high school (Blimpo, Gajigo, and Pugatch 2016).

we include a further description of the intervention and its motivation, a sample size calculation, the full statistical analysis tables according to our pre-specified statistical analysis plan, and the final test papers we used to assess child reading and mathematics ability.

## 2. Intervention and setting

In this section we briefly describe the logic of the intervention we study, its implementation, and the setting we work in.

### *2.1 Intervention design*

The intervention we study uses locally hired and previously untrained individuals[4] to serve as para teachers, providing after-school, supplementary education to children in the early years of primary school. Its purpose is to ensure that recipient children master literacy and numeracy at the appropriate age and level of schooling. It is in the spirit of the "Balsakhi" intervention studied in Banerjee et al. (2007), wherein previously untrained individuals were trained by an NGO and then paid to teach remedial education to children.

Because of their contract structure, para teachers are potentially more motivated to perform well, more amenable to frequent monitoring (a key facet of our intervention), and less likely to suffer from the adverse selection problems that plague recruitment and management of civil-service teachers (Chaudhury et al. 2006; Muralidharan and Sundararaman 2013). These differences in incentives, monitoring, and teacher selection may partially address the teacher absenteeism, suboptimal effort, and several of the other problems documented in numerous

[4] That is, individuals not previously trained to serve as teachers in the government system.

studies of the delivery of education in such areas (e.g., Chaudhury et al. 2006, and Duflo, Dupas, and Kremer 2015).

We sought to hire individuals primarily from the village in which they served, though in approximately half of our villages, we were unable to find someone with sufficient education to do so, and so we had to recruit from nearby communities. The goal here was to exploit the "informational and motivational advantages" that come with hiring local individuals - also known as local delivery agents – to administer services (Bandiera et al., 2018). We required all teachers to live in the village in which they served, improving the ability of children and parents to address teacher attendance issues directly, i.e., by going to the teacher's home to find them should they be absent.

Our intervention was based on the intervention studied in Lakshminarayana et al. (2013). That study ran a cluster-randomized trial evaluating a similar para teacher intervention. That intervention was designed by The Naandi Foundation, which had been implementing it in multiple Indian states for several years prior to the start of the trial in 2008. The trial reported in Lakshminarayana et al. (2013) took place in 214 villages in rural Telangana (then Andhra Pradesh), India, over a period of two academic years. The study found that the intervention yielded a 0.75 SD increase in reading and mathematics test scores among children in intervention villages, relative to those in control villages.

Based on these results, we aimed to study whether the success of this model could generalize to a new, more challenging setting. We chose The Gambia as our new setting for the following reasons. One, much of The Gambia is rural, hard to reach, and has low levels of literacy and numeracy, making it an appropriate setting for our intervention. Two, its government was interested in hosting the study and, were the intervention found to be effective, in the

potential for subsequently implementing it themselves. Three, the official language and language for instruction is English, making it easier for us to maintain quality control in adapting the materials from Lakshminarayana et al. (2013) to the local context.

*2.2 Implementation*

In those trial villages randomly selected to receive the intervention, implementation proceeded according to the following steps: first, we held a meeting in each community, announcing the intervention and asking all community members for their support. Second, with the community, we attempted to identify an adult from the village with at least a 12$^{th}$ grade education, to serve as a para teacher. In the absence of such an individual, we relaxed either the locality or education requirement (or both) and found the most qualified individual who met a set of minimum qualifications, passed a proficiency test, and who was willing to reside in the village, with a preference for those from nearby communities. We paid these individuals a post-tax salary of 3,550 Gambian Dalasis (GMD) per month (US $81.12), roughly 1/7 more than the government teachers received (3,085 GMD per month, or US $70.55).[5] Due to other benefits from the government and greater monetary and in-kind transfers from parents to government teachers, overall compensation was roughly equivalent for the two groups.

To emphasize their role in the community, these para teachers were hired under the official title of "community educator", or "CE." At the start of the intervention, we provided an initial two-month pre-service training for the recruited CEs in pedagogical content knowledge related to our curriculum. This curriculum was based on the official Gambian national curriculum for these grades, but incorporated scripted daily lesson plans and a series of activities

[5] All GMD to USD conversions are made using the average exchange rate over the period 2015-2018 taken from https://www.exchangerates.org.uk/ on May 15, 2019.

for each lesson, using the core tools from Lakshminarayana et al. (2013). It was also designed to be easily implementable by our team of CEs who, when hired, had no teaching experience.

After their training, CEs returned to their villages to commence the intervention. Each CE administered 12 hours of after-school lessons per week using our daily scripted lesson plans. These lessons took place either in the local school or, in the absence of a nearby school, a structure which we supplied with mats for sitting and a chalkboard. In these cases, the community either furnished a suitable place to hold the lessons, such as a local madrassa (an Islamic religious school) or meeting hall, or constructed a one-room structure in which to hold the lessons. We monitored the CEs throughout the course of the intervention and provided regular training as required by the progression of the curriculum. We provide further details of the intervention's setup, implementation, and motivating ideas in Appendix A.

The intervention targeted all children in the village who were born between September 1, 2007 and August 31, 2009 (roughly, ages six to eight at time of enumeration in early 2015) and whose caregiver reported a) that the child was not currently enrolled in the first grade or higher, and b) that they intended to enroll the child in the first grade in the 2015-2016 school year.[6] We chose to focus on children entering the first grade to ensure clarity about what stage of education our enrolled children were in. In the absence of such a restriction, the varying ages in which children enter the first grade (or, conversely, secular variance in the proportion of children of this age group who enrolled in the first grade) would add further complexity to our analysis.

We further describe the recruitment procedure in the next section. We describe levels of enrollment in school over the course of the trial, by randomization assignment (intervention vs.

[6] Nonetheless, we were only able to capture parent intention to enroll their child. As a result, our sample includes many children who did not ultimately enroll in school the following school year and some children (a subset of this group) for whom, this intervention is not "supplementary," but rather the only formal education that they received. We present relevant summary statistics on enrolment over the course of the study in Table 8.

control), in Section 4. In Appendix B, we provide a series of tables, as specified in our Statistical Analysis Plan[7], which give greater detail about the socio-economic and demographic characteristics of participants in our study.

*2.3 Setting*

The Gambia is a small country in West Africa, roughly 475 km long and 25 to 50km wide. It comprises the area around the River Gambia to the Atlantic coast and bordered on the north, south, and east by Senegal. Its main exports are tourism and agricultural products. It was formerly a British colony and home to a large part of Britain's slave trade. Its per-capita gross domestic product was estimated to be US $483 in 2017 by the World Bank (market rate, not purchasing power parity), placing it among the 10 poorest countries in the world according to that ranking (The World Bank 2018a).

The education system in The Gambia comprises basic education – six years of primary schooling (lower basic) and three years of middle school (upper basic) – followed by three years of high school or vocational training, and then university. Its gross primary school enrollment rate was 100 percent in 2017, slightly higher than that for sub-Saharan Africa overall (97 percent).[8] Population-level data on the average years of education completed are sparse, particularly for older generations, but in a census conducted in 2013, literacy rates for individuals aged 15 or older were estimated to be approximately 42% overall, with higher levels for males (51%) than females (34%) (The World Bank 2018b). Since 2007, The Gambia has run regular assessments of the literacy of its children using EGRA-style tests, with the addition of EGMA-

[7] In addition to our published study protocol (Boone et al. 2015), we also wrote out a more detailed statistical analysis plan prior to analysis of the data. As stated above, the full results according to this analysis plan are given in the Appendix. The accompanying text of the analysis plan is available from the authors on request.
[8] Data taken from https://data.worldbank.org/indicator/se.prm.enrr on May 22, 2019.

style tests beginning in 2013. The first assessment yielded startling results – country-wide, 46% of third grade students could not read a single word in a sentence, and fewer than 20% were at the level expected of them by the national curriculum. In the regions in which our study takes place, these levels are much lower. Since then, The Gambia's Ministry of Basic and Secondary Education (MoBSE) has regularly measured learning levels using EGRA- and, later, EGMA-style assessments, and have put pressure on its staff to raise them.

The Gambia has six administrative regions – Banjul (the capital), West Coast, Lower River, North Bank, Central River, and Upper River. We followed the recommendation of MoBSE to work in the Lower River and North Bank regions, located in the center of the country. MoBSE justified this recommendation in light of the fact that these regions were needier than the Western regions (Banjul and West Coast) and had fewer ongoing NGO interventions than the Eastern Regions (Central River and Upper River).

## 3. Research Design

This section describes our research design. We published a protocol prior to executing the study which specified our primary and secondary outcomes and our analysis methods (Boone et al. 2015). The protocol also provides greater detail on the issues discussed in this section relevant to the design of the study.

### *3.1 Eligibility and enrollment of villages and children*

We began with a list of the 323 villages in the Lower River and North Bank regions with between 15 and 300 households according to the 2013 Gambian national census. In each village on this list, we conducted a census of all dwellings in order to enumerate the number of children

born between January 1, 2006 and December 31, 2010 currently resident in the village, counting only those children whom we could meet face-to-face.[9] For a village to be eligible for inclusion in the randomization, we required that it have at least 10 children born between September 1, 2007 and August 31, 2009, who had not yet enrolled in the first grade, and whose caregiver intended to enroll them in the first grade in the coming academic year.[10]

Among villages that were eligible, we drew a circle with a radius of 2.5 kilometers around each to serve as a buffer area. We then generated clusters of villages, the unit of randomization, from contiguous groups of village buffer zones. If there were three or more villages in a given cluster, we removed one or more of these villages from the trial to generate the maximum number of clusters such that there were at least five kilometers between the GPS coordinates of any village in the cluster to those of all villages in all other clusters. This left us with 169 villages grouped into 111 clusters. We enrolled all eligible children in these 169 villages into the trial, obtaining consent from village chiefs and each child's primary caregiver.

We conducted our randomization by cluster of villages, with clusters defined by geographic proximity as described above, to avoid the risk of spillover, e.g., children in control villages being able to walk into intervention villages and avail themselves of intervention classes there. We used a random number generator, stratifying on two criteria: whether a cluster was in the Lower River or North Bank region, and whether the cluster was above or below the median distance to the main road in its region. The sample size calculation we used to design this sampling frame and to confirm that we would have adequate statistical power is described in Appendix C.

[9] Birthdate data were confirmed with birth and health records in all cases where possible.
[10] We started with the wider range – 2006 to 2010 – to avoid parental misreporting of a child's birth date in order to satisfy the September 1, 2007 to August 31, 2009 eligibility rule. In the list of children we collected, we see no bunching around either the earlier or later eligibility cutoff.

*3.2 Data collection*

We collected information from families, villages, and schools at baseline as part of our census of potentially eligible villages. After finalizing which villages would be included in the trial, we conducted a second, in-depth survey of all eligible children in these villages to enroll them in the trial (we could not enroll children prior to determination of their village's eligibility) and in order to collect further data on the child and her or his caregiver. We collected end-of-school-year surveys at the end of the first and second academic years (i.e., May-June 2016 and 2017) to measure child migration, enrollment in school, and attendance. Between February and May 2018, we conducted an endline survey with caregivers of children to measure attitudes, time use, and other child- and family-level variables.

Finally, between May and June 2018, we conducted EGRA and EGMA-style tests among all enrolled children in our program. EGRA (Early Grade Reading Assessments) and EGMA (Early Grade Mathematics Assessments) are reading and mathematics tests, respectively. They were initially designed by RTI International and are intended to test a set of basic skills related to each subject (Dubeck and Gove 2015; Platas et al. 2014). They are meant to be adapted to a local context for each administration; specifically, the test questions are meant to differ from test to test to conform to local needs and standards while following standardized EGRA/EGMA protocol design guidelines. They are oral assessments conducted one-on-one with the child and do not require the child to complete a paper-based assessment.

In Table 1, we briefly describe the skills tested in each of the two tests. The full test papers are given in Appendix D. One of the authors (Hsieh) was contracted to design our tests to be consistent with prior EGRA and EGMA-style assessments in The Gambia and to ensure

consistency of training and implementation by assessors. This design, training, and implementation followed RTI's guidelines for creating reliable and accurate EGRA and EGMA-style assessments.

**Table 1: Description of EGRA and EGMA test content**

| EGRA | EGMA |
|---|---|
| 1: Read a letter's sound (e.g., "eh" for e) | 1: Read a number (e.g., 1, 5, 22) |
| 2: Differentiate sounds (e.g., which word starts with a different sound: book, dog, or boy) | 2: Choose the larger number (e.g., 7 or 5) |
| 3: Read a made-up word (e.g., tob) | 3: Complete a sequence (e.g., 2 4 6 __ ) |
| 4: Read a familiar word (e.g., but) | 4a: Simple addition (e.g., 3+2) |
| | 4b: Two- and three-digit addition (e.g., 38+26) |
| 5a: Read a short passage | 5a: Simple subtraction (e.g., 5-3) |
| 5b: Answer questions on the passage's content | 5b: Two- and three-digit subtraction (e.g., 59-37) |
| 6: Listen to a different short passage, answer questions on the passage's content | 6: Solve a simple word problem read aloud |

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

*3.3. Pre-specified primary and secondary endpoints and analysis method*

Our published study protocol (Boone et al. 2015) contains key information on our analysis plan, including primary and secondary outcomes and method of analysis, specified prior to conducting the trial. The primary investigators agreed upon and signed off on a detailed statistical analysis plan prior to the start of statistical analysis.[11] Our primary outcome is a

[11] The full tables of this are included in Appendix B. The accompanying text is available by request.

composite test score, calculated as the arithmetic mean of a child's scores (scaled 0-100) on consecutively administered reading (EGRA) and mathematics (EGMA) tests.[12]

MoBSE has used EGRA-style tests since 2007 to assess child learning, with the addition of EGMA-style tests in 2013. Since MoBSE regularly assesses student progress and teacher performance using data from its own administration of these tests in The Gambia, our choice of primary endpoint is aligned with the goals and standards of the Ministry. In cases such as India, where the curriculum advances rapidly, there is reason to believe that a study such as ours with one specific endpoint might distort teacher effort (Muralidharan and Sundararaman 2011). Another related concern is that arbitrarily choosing a test or set of skills for our primary outcome could lead to a spurious conclusion that we raised learning levels when, in fact, the status quo system and the intervention were targeting different goals. In this setting, however, the intervention, the Gambian primary education system, and the tests we use all target the same measures of literacy and numeracy. As a result, we argue, the risk of either such distortion of effort or of spurious conclusions of efficacy is minimal. Finally, due to the nature of the intervention, participants were aware of their randomization status. Because of the objective nature of the primary outcome data, we argue this knowledge is unlikely to bias our results (Wood et al. 2008).

We use a linear regression model to compare child-specific composite endline test scores between intervention and control groups. In our model, we control only for the two stratification factors included in the randomization (region: lower river or north bank; distance to main road in region: above or below median distance). Here, and in other analyses, we perform hypothesis

[12] We chose this composite score because we needed to select one outcome to serve as the primary outcome of our study; we acknowledge that this is a departure from the intended and conventional use of EGRA and EGMA tests. To align with more conventional use, we present subtask scores in Figures 3.1 and 3.2. Other commonly presented scores – e.g., fluency scores and zero scores – are available on request.

tests and calculate 95% confidence intervals using robust standard errors, allowing for correlated responses by cluster of villages. For the primary outcome, we divide the adjusted difference in means by the standard deviation (SD) of the test score in the control group to give a standardized difference in SD terms, using the total SD for the control group estimated by fitting a linear mixed model that allowed for between and within cluster variability. We present this standardized difference together with a nonparametric bootstrap confidence interval.[13]

In our secondary analyses, we extend the linear regression model described above to investigate interactions by ethnic group, gender, wealth, caregiver education, and geographic location, respectively. Test performance data by subtask are presented in bar charts. We do not test for statistical significance in the differences in these measures between intervention and control groups to avoid the risk of Type I error. Our pre-specified secondary outcomes include school attendance, enrollment, performance on nationally administered exams[14], parents' spending on education, spillover learning to siblings and family members, and school-related time use of parents and children. For dichotomous secondary outcomes – such as whether the child was enrolled in school – we present odds ratios with 95% confidence intervals obtained from a logistic regression model fitted within a generalized estimating equation (GEE) framework with a binary outcome, a logit link, and a 'working' assumption of independence, with robust standard errors to take account of clustering. For ordered categorial secondary outcomes – the spillover literacy and numeracy outcomes – we compare intervention and control groups using an ordered logistic regression model fitted within a GEE framework with a 'working' assumption of independence, with robust standard errors to take account of clustering

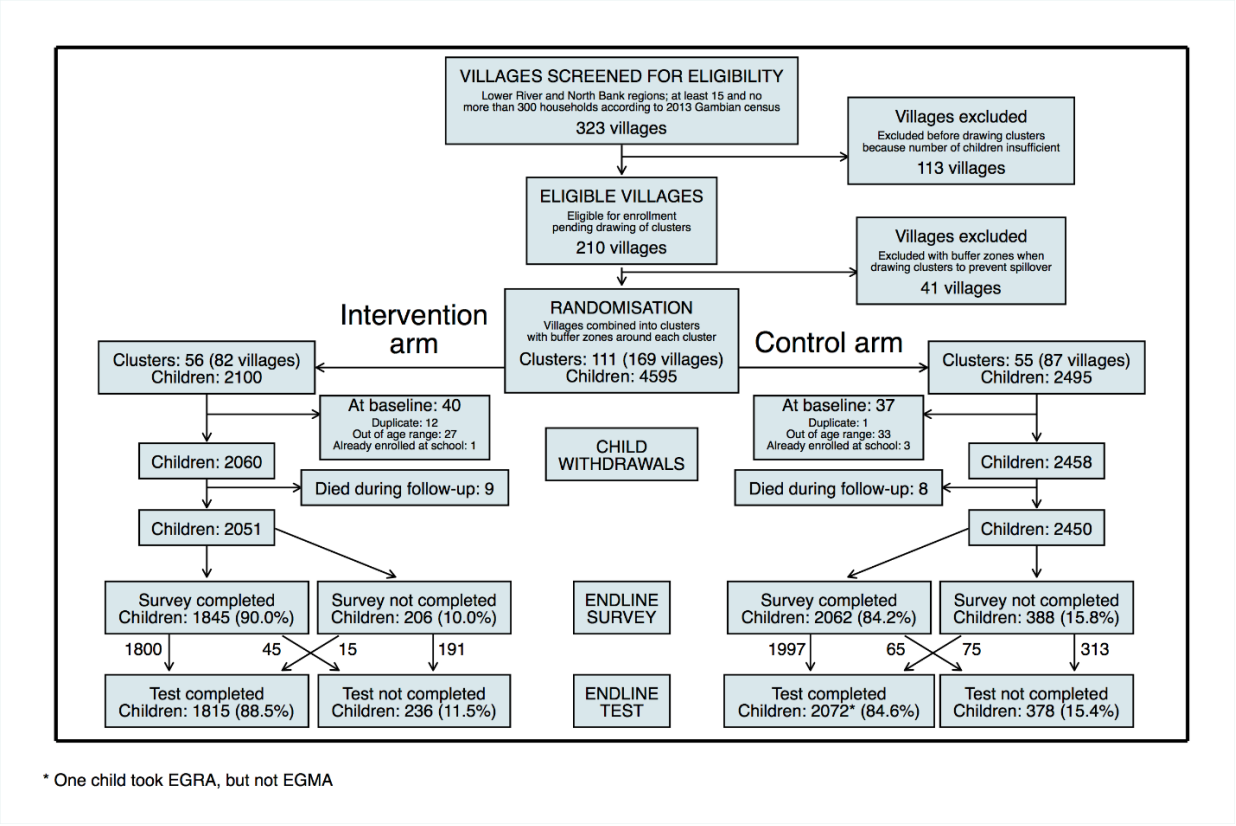
[13] Bias corrected and accelerated, 2000 replications, clustered by clusters of village, stratified by randomized group.
[14] Due to a policy reducing the frequency of nationally administered exams after we specified our analysis plan, our children were not administered the Gambian National Assessment Test and so we are not able to conduct this analysis.

(Liang and Zeger 1986). We conduct all primary analysis on an intent-to-treat basis, including children in the group that their village was randomized to irrespective of attendance at classes or school. In Appendix E, we describe and report pre-specified, secondary, per-protocol analyses of the primary outcome in those villages and children whose class schedule (villages) and child attendance (villages and children) met pre-specified attendance thresholds.

*3.4 Randomization and balance*

In Figure 1, we present a CONSORT-style diagram (Campbell et al. 2012) showing the numbers of villages, clusters, and children involved at various stages of the trial. This diagram shows how villages and children flowed through the trial from consideration for eligibility to the endline test.

**Figure 1: CONSORT-style flowchart of participants through trial**



In Table 2, we show the baseline characteristics of the clusters in our trial, the level at which we randomized. Fifty six clusters were randomized to the intervention group and fifty five clusters to the control group. By chance, the six largest clusters were all randomized to the control group; as a result, clusters receiving the intervention have a somewhat smaller mean number of eligible children per cluster than those in the control group. After excluding ineligible children (see Figure 1) there were 2,060 and 2,458 children enrolled in the trial at baseline in the intervention and control groups respectively.

**Table 2: Baseline characteristics of clusters**

| Variable | | (1) Intervention | (2) Control |
|---|---|---|---|
| Number of clusters | | 56 | 55 |
| *Number of clusters by stratum* | | | |
| Region: North Bank / Lower River | | 36:20 | 35:20 |
| Distance to road: above median / below median | | 28:28 | 27:28 |
| Mean cluster distance to road in km (SD) | | 2.00 (2.92) | 1.65 (2.80) |
| Number of randomized eligible children per cluster: mean (SD) | | 36.8 (20.8) | 44.7 (35.7) |
| *Villages per cluster:* | | | |
| | 1 | 35 | 32 |
| | 2 | 18 | 16 |
| | 3 | 1 | 5 |
| | 4 | 2 | 2 |
| Mean cluster population (SD) | | 1188 (556) | 1415 (1007) |

Note: this table presents baseline characteristics of control and intervention clusters.

In Table 3, we present characteristics of these enrolled children, their caregivers, and their mothers (mothers are usually, but not always, the primary caregiver for children) separately by intervention group. Following recommended best practice (Bruhn and McKenzie 2009; Campbell et al. 2012; Moher et al. 2010) we do not conduct statistical tests for differences in baseline characteristics, as any differences would necessarily have arisen by chance. We see similar gender ratios, identities of the child's caregiver, caregiver education, caregiver literacy, and child age across intervention and control groups. In both groups, roughly three-quarters of parents had never gone to school. By chance, the ethnicity of children varies somewhat by randomization group, with control children somewhat more likely to be Wolof than intervention children. Such a chance imbalance is not unexpected given that randomization is by village cluster, with ethnicity variations being pronounced across villages.

**Table 3: Baseline characteristics of children and their caregivers**

| Variable | (1) Intervention | (2) Control |
|---|---|---|
| Number of children enrolled in the trial at baseline | 2060 | 2458 |
| Gender: | | |
| *Male* | 51.9% (1070) | 50.4% (1239) |
| *Female* | 48.0% (989) | 49.5% (1217) |
| *Not Known* | 0.0% (1) | 0.1% (2) |
| Ethnic group: | | |
| *Mandinka* | 40.9% (842) | 42.3% (1040) |
| *Wolof* | 16.2% (334) | 24.7% (608) |
| *Fula* | 25.0% (516) | 19.7% (485) |
| *Other* | 15.1% (312) | 11.4% (279) |
| *Missing* | 2.7% (56) | 1.9% (46) |
| Child's main caregiver: | | |
| *Biological mother* | 73.3% (1511) | 74.6% (1833) |
| *Biological father* | 3.3% (69) | 5.2% (128) |
| *Grandmother* | 11.0% (227) | 10.1% (249) |
| *Grandfather* | 0.5% (11) | 0.7% (18) |
| *Step/foster mother* | 4.1% (84) | 4.0% (99) |

| | | |
|---|---|---|
| *Step/foster father* | 0.5% (11) | 0.5% (13) |
| *Other caregiver* | 4.6% (94) | 3.2% (78) |
| *Missing* | 2.6% (53) | 1.6% (40) |
| Caregiver's education: | | |
| *No education* | 73.8% (1520) | 75.7% (1861) |
| *Pre-K or primary* | 15.4% (318) | 14.3% (352) |
| *Junior Secondary* | 5.3% (110) | 5.8% (142) |
| *Senior secondary or higher* | 2.6% (54) | 2.3% (57) |
| *Don't know* | 0.1% (2) | 0.0% (0) |
| *Missing* | 2.7% (56) | 1.9% (46) |
| Child's age in September 2015 | 6.87 (SD=0.55) | 6.88 (SD=0.55) |
| Language spoken in home: | | |
| *Mandinka* | 42.2% (869) | 44.5% (1093) |
| *Wolof* | 18.4% (379) | 28.1% (691) |
| *Fula* | 24.6% (506) | 17.9% (440) |
| *Other* | 12.1% (250) | 7.6% (188) |
| *Missing* | 2.7% (56) | 1.9% (46) |
| Caregiver literacy at baseline: | | |
| *Can't read* | 75.6% (1557) | 77.9% (1915) |
| *Can read at least one letter, but not an entire word* | 9.9% (204) | 8.5% (209) |
| *Can read at least one word, but not entire card* | 3.9% (80) | 3.8% (93) |
| *Read entire card slowly* | 2.8% (57) | 3.6% (89) |
| *Read entire card fluently* | 5.1% (106) | 4.3% (105) |
| *Refused* | 0.0% (0) | 0.0% (1) |
| *Missing* | 2.7% (56) | 1.9% (46) |
| Wealth*: | | |
| *Category 1* | 7.3% (151) | 4.6% (113) |
| *Category 2* | 66.7% (1373) | 66.5% (1635) |
| *Category 3* | 23.3% (480) | 27.0% (664) |
| *Missing* | 2.7% (56) | 1.9% (46) |
| Has older sibling: | | |
| *Yes* | 84.6% (1742) | 85.4% (2098) |
| *No* | 15.4% (318) | 14.6% (360) |
| Has younger sibling: | | |
| *Yes* | 87.6% (1805) | 90.3% (2220) |
| *No* | 12.4% (255) | 9.7% (360) |

Note: Except when labeled otherwise, this table presents the group-specific proportion of children holding each characteristic with the number of observations in parentheses. *: We define wealth as a categorical measure defined by the materials of the roof, walls, and floor of the child's home at baseline. Category 1 is that all materials are

natural (e.g., a thatched roof, mud walls, and an earthen floor); category 2 is that some but not all materials are synthetic (e.g., a steel roof, but natural walls and floor); category 3 is that all materials are synthetic (e.g., a steel roof, brick walls, and a tile or concrete floor).

Given that our participants were children who had not yet enrolled in school, and that we were working with a population in which less than 8% of children's caregivers were literate[15], we did not administer a test of literacy and numeracy at baseline to children enrolled in the trial under the assumption that all but a trivially small number of children would register zero scores. The large number of zero scores among control group children at the endline test corroborates this assumption.

*3.5 Intervention*

In Table 4, we show key details on the nature of the intervention. In most of the 82 villages which received the intervention, it comprised two hours of teaching, given six times per week. In four villages, the schedule was adjusted slightly because a large proportion of students had to attend both traditional school and Qur'anic school during the week. This led us to reduce the number of total days of class in these villages and compensate with a longer class on some days. This is documented in the top panel of Table 4.

The levels of take-up of the intervention are shown in the bottom panel of Table 4. The first column shows that on a week-by-week basis there was no deviation from the schedule (this masks a substantial amount of on-the-ground rescheduling of classes, for instance, when teachers were sick). The second and third columns show that, on average, intervention children attended our after-school classes slightly more than 75 percent of the intended time.

[15] According to the literacy test administered to caregivers at baseline. Literacy defined as those who "read entire card slowly" or "read entire card fluently"; results shown in Table 3.

**Table 4. Class schedules and adherence, intervention group only**

*Panel A: Schedule*

| Schedule | Number of villages |
|---|---|
| Six two-hour classes per week | 78 |
| Five classes per week *(four weekday classes at two hours per day, and one weekend class at four hours per day)* | 2 |
| Four classes per week *(two weekday classes at two hours per day, and two weekend classes at four hours per day)* | 2 |

*Panel B: Adherence*

| | (1) Percent of regularly scheduled classes actually held (village-level) N=82 | (2) Percent of children attending each regularly scheduled class (village-level) N=82 | (3) Percent of regularly scheduled classes the child attends (child-level) N=2060 |
|---|---|---|---|
| Mean | 100% | 78.9% | 76.8% |
| SD | 0% | 8.7% | 27.8% |
| Distribution: | | | |
| *0* | 0 | 0 | 0.7% (15) |
| *>0 to 25%* | 0 | 0 | 10.1% (209) |
| *>25% to 50%* | 0 | 0 | 3.7% (77) |
| *>50% to 75%* | 0 | 30.5% (25) | 10.2% (210) |
| *>75% to 100%* | 100% (82) | 69.5% (57) | 74.6% (1537) |
| Missing | 0 | 0 | 0.6% (12) |

Note: Panel A tabulates the proportion of intervention villages with each of three different possible weekly implementation schedules. Panel B shows the proportion of children and villages meeting pre-specified adherence targets in terms of the proportion of regularly scheduled classes held at the village level (column 1), the proportion of enrolled children attending these classes (column 2) and the proportion of regularly scheduled classes children

attended (column 3). We give the number of observations for each adherence level in parentheses next to the proportion.

*3.6 Migration and attrition*

In this section, we briefly comment upon the retention of observations in our sample throughout the course of the trial. At the end of each academic year, we visited all trial villages with the goal of locating each child who was enrolled in the trial, asking her/his caregiver simple information about their activities in school the previous year, and recording whether or not the child was present in the village. In Table 5, we tabulate children's presence in the village across intervention and control groups and years of the study. We see that in the first year of the study, in both groups, less than seven percent of children are not present in the village. This increases to 8.4 and then 9.6 percent in the intervention villages in years two and three respectively, and to 10.4 and then to 15.1 percent in the control villages. At endline, attrition is 5.7 percentage points higher in control villages. Should higher-ability children be more likely to leave a village in the absence of the intervention, this could bias our estimates upwards. Nonetheless, given the small magnitude of this difference, no more than a small fraction of the large differences in performance we measure between intervention and control children could be explained by differential attrition.

**Table 5. Child presence in study village (migration) over course of trial**

| Variable | Intervention (N=2060) | | | Control (N = 2458) | | |
|---|---|---|---|---|---|---|
| | Year 1 | Year 2 | Year 3 | Year 1 | Year 2 | Year 3 |
| Present in village | 93.2% (1920) | 89.3% (1839) | 89.6% (1845) | 93.0% (2286) | 87.7% (2156) | 83.9% (2062) |
| Not present in village | 6.6% (135) | 8.4% (174) | 9.6% (197) | 6.7% (164) | 10.4% (256) | 15.1% (372) |
| Data missing | 0.2% (5) | 2.3% (47) | 0.9% (18) | 0.3% (8) | 1.9% (46) | 1.0% (24) |

Note: This table provides the proportion of children who were present in the village at the time of our annual visits, which took place at the end of each academic year, e.g., May-July 2016, May-July 2017, and February-May 2018, for Year 1, Year 2, and Year 3, respectively. The relevant number of observations is given in parentheses under the proportion.

## 4. Main results

In this section we present our main results. We begin by showing the primary outcome – children's performance on the endline test, calculated as the arithmetic mean of their scores on the reading test and the mathematics test. We then discuss performance on the various subsections of the test, which vary in difficulty, followed by comparisons of the primary outcome across pre-specified subgroups. We also present results for our pre-specified secondary outcomes: the child's enrollment and attendance in school, the household's education-related expenditure on the child, time use of the child and caregiver, and possible spillover of literacy or numeracy gains to the child's nearest older sibling, nearest younger sibling, and her caregiver. We conclude this section with a brief cost and cost-effectiveness analysis of our intervention.

### *4.1 Primary outcome*

We show the distribution of our primary outcome, the composite test score, in Figure 2, separately for intervention and control children. We show group-specific means of this variable and adjusted differences between them in the first row of Table 6. Intervention children score 46 percentage points higher on the test than control children (95% confidence interval: [43.3, 48.8]). This difference in means is highly statistically significant ($p < 0.0001$). In SD terms, this comprises a 3.2 SD difference; however, the SD measure is hard to interpret since the distributions of test scores in the two groups are far from Gaussian – with the consequence that the standard deviation is not a comprehensive description of dispersion – and markedly different between the two randomized groups. The second and third rows of the table provide analogous

results for the overall reading and mathematics test scores, separately; these show similar differences. In Figures 3.1 and 3.2, we present a bar chart showing mean scores, by intervention and control groups, for each component (also called a "subtask" or "task") of the mathematics and reading tests, respectively. As the number label on the subtask increases (e.g., from subtask 1 to subtask 2), so does its difficulty. The nature of each subtask is described in Table 1, and the full test papers are provided in Appendix D.

The patterns in this table show dramatic differences between control and intervention groups on all subtasks, from the easiest (letter recognition in reading and number recognition in math) to the most difficult (reading comprehension in reading, and two- and three-digit subtraction with borrowing in math). They also reveal very low learning levels in the control group. For the addition and subtraction subtasks on the mathematics test, the mean control group score is less than 20 percent correct for simple addition and less than 10 percent correct for advanced addition, simple subtraction, and advanced subtraction, respectively.[16] For the reading subtasks, the patterns are similar. For no reading subtasks does the control mean score exceed 30 percent correct answers and, for the five most difficult subtasks, in no case does the mean exceed six percent correct answers. In the intervention group, the mean mathematics subtask scores are between 80 and 95 percent correct answers for easier subtasks, and between 50 and 65 for the more difficult subtasks. For reading, these mean scores are between 47 and 69 percent correct.

[16] The higher control group scores from subtask 6 likely reflect the fact that the question is spoken to the child and relies less on school-based knowledge than the previous subtask, 5b, which requires parsing written, two-digit subtraction problems. Subtask 6 questions are also computationally simpler (e.g., single digit adding or subtraction) than those in 5b.

**Figure 2: Distribution of primary outcome for intervention and control children**

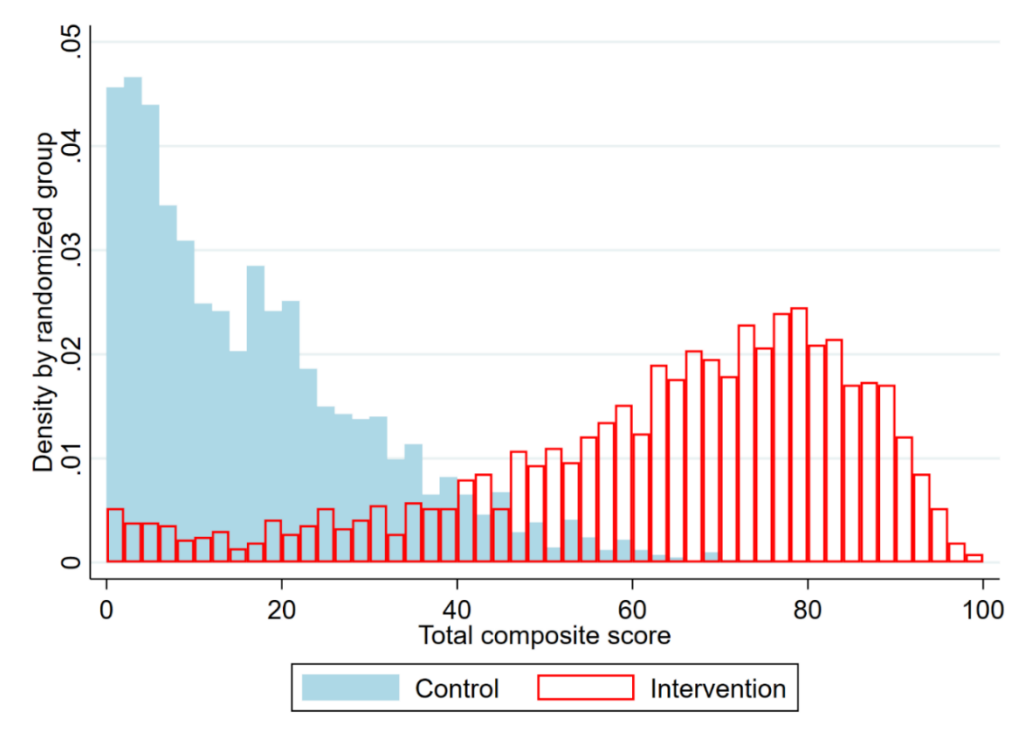



**Table 6. EGRA and EGMA test results**

| **Variable** | (1) **Intervention** | (2) **Control** | (3) **Adjusted difference [95% CI]** | (4) **P-value** |
|---|---|---|---|---|
| Composite test score | 63.3 (22.3) | 17.1 (14.2) | 46.0 [43.3, 48.8] | p<0.0001 |
| Composite test score difference in SD units | -- | -- | 3.23 [2.89, 3.63]* | -- |
| Mathematics test score, overall | 68.2 (21.8) | 24.7 (19.7) | 43.4 [40.2, 46.5] | p<0.0001 |
| Reading test score, overall | 58.3 (25.3) | 9.5 (11.2) | 48.7 [46.1, 51.4] | p<0.0001 |
| Number of observations** | 1815 | 2071 | -- | -- |

Note: In columns (1) and (2), group-specific estimates are shown in each cell, with standard deviations given below in parentheses. In column (3), we show the difference between column (1) and (2) adjusted for the randomization stratification factors with a 95% confidence interval (that takes into account the clustered design) in brackets below. In column (4) we present the p-value from the corresponding hypothesis test. *: Bootstrap confidence interval, bias corrected and accelerated, based on 2000 bootstrap samples of clusters with stratification by intervention/control. **: We have one additional observation in reading for the control group (a child who took the reading test, but not the mathematics test).

## Figure 3: Test performance, by subtask

*Figure 3.1: Mathematics subtask test scores for intervention and control children*

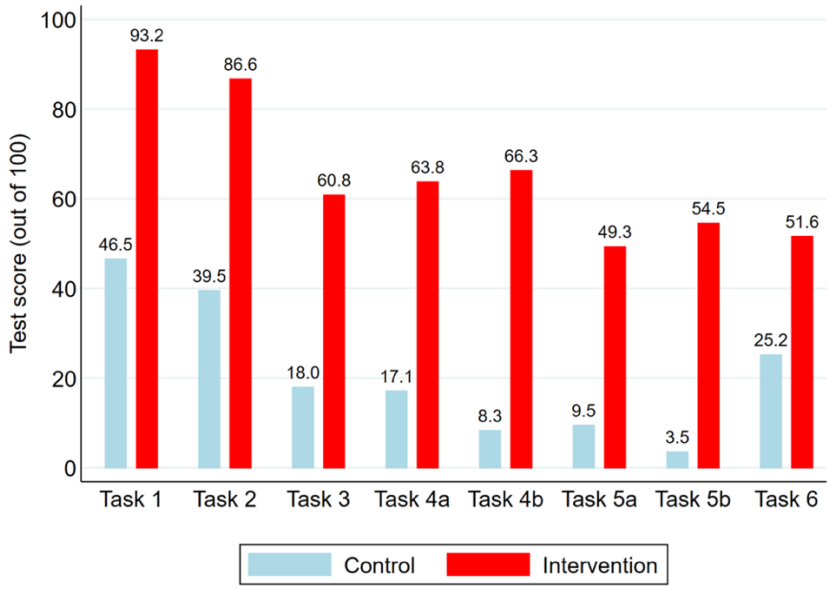
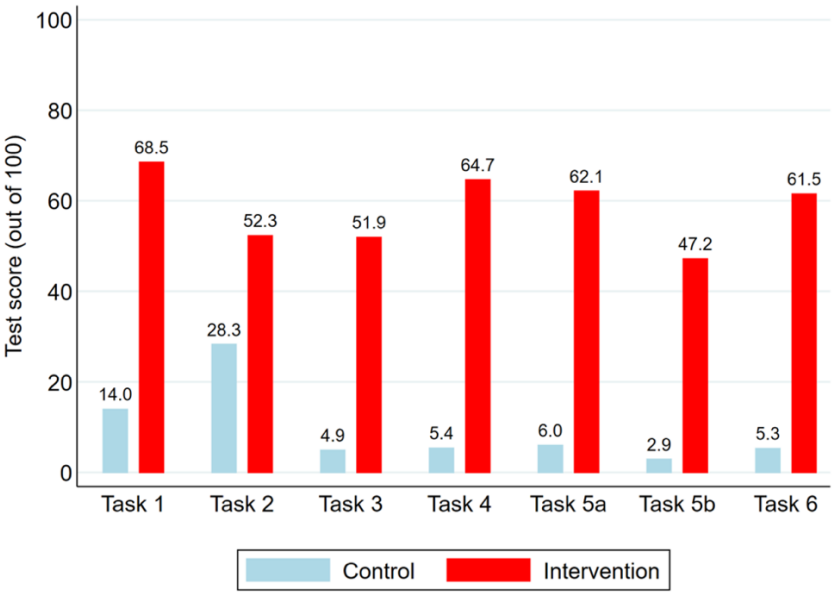



*Figure 3.2: Reading subtask test scores for intervention and control children*



Note: Figure 3.1 shows the mean test score of children in control and intervention groups on each of the mathematics subtasks described in Table 1, and Figure 3.2 shows the corresponding means for each of the reading subtasks described in Table 1.

We conduct ancillary analyses of the primary outcome to deal with differential uptake of the intervention across intervention villages, within the intervention group. These comprise a series of pre-specified per-protocol analyses to see how the treatment effect varies when restricting attention to only villages and children which held or attended classes, respectively, at a prespecified minimum adherence level. Given how close these are to our overall estimates, we report them in Appendix E.

*4.2 Subgroup analyses*

In Table 7, we present subgroup analyses of the primary outcome by pre-specified variables of interest based on characteristics of the child and village. We find no evidence of a differential impact of the intervention by any of the following variables: gender; wealth; ethnicity; the region in which the village is located (Lower River or North Bank); and distance of the village to the road. There is some evidence that the intervention may be marginally more beneficial for children with less-educated caregivers, though the pattern we observe is not statistically significant.

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

| Subgroup | (1) Intervention | (2) Control | (3) Adjusted difference [95% CI] | (4) P-value for interaction |
|---|---|---|---|---|
| Gender | | | | |
| *Male* | 62.4 | 16.4 | 45.9 | |
| (N: I=935, C=1022) | (23.1) | (14.1) | [42.9, 49.0] | p=0.86 |
| *Female* | 64.1 | 17.9 | 46.2 | |
| (N: I=879, C=1047) | (21.4) | (14.2) | [43.2, 49.2] | |

| | | | | |
|---|---|---|---|---|
| Wealth | | | | |
| *Category 1* | 60.9 | 15.2 | 45.6 | |
| (N: I=138, C=93) | (24.9) | (12.5) | [38.5, 52.7] | p=0.53 |
| *Category 2* | 63.5 | 16.7 | 46.7 | |
| (N: I=1227, C=1396) | (21.8) | (13.8) | [43.9, 49.6] | |
| *Category 3* | 63.7 | 18.6 | 45.1 | |
| (N: I=429, C=570) | (22.1) | (15.3) | [41.5, 48.7] | |
| Ethnicity | | | | |
| *Mandinka* | 63.4 | 16.7 | 46.5 | |
| (N: I=751, C=901) | (21.6) | (13.6) | [42.3, 50.6] | p=0.20 |
| *Wolof* | 63.8 | 16.4 | 47.4 | |
| (N: I=295, C=504) | (24.6) | (14.2) | [43.0, 51.8] | |
| *Fula* | 63.5 | 20.8 | 42.6 | |
| (N: I=467, C=411) | (21.9) | (14.9) | [38.8, 46.3] | |
| *Other* | 62.7 | 14.2 | 49.1 | |
| (N: I=281, C=243) | (21.6) | (13.6) | [43.2, 55.0] | |
| Region | | | | |
| *Lower River* | 63.4 | 19.5 | 43.9 | |
| (N: I=677, C=688) | (21.9) | (14.9) | [38.8, 48.9] | p=0.27 |
| *North Bank* | 63.2 | 15.9 | 47.2 | |
| (N: I=1138, C=1383) | (22.5) | (13.6) | [44.0, 50.4] | |
| Distance to road | | | | |
| *>median* | 63.9 | 16.2 | 47.6 | |
| (N: I=731, C=836) | (22.8) | (13.4) | [43.3, 52.0] | p=0.34 |
| *<median* | 62.8 | 17.8 | 45.0 | |
| (N: I=1084, C=1235) | (21.9) | (14.6) | [41.6, 48.3] | |
| Caregiver education | | | | |
| *None* | 63.0 | 16.5 | 46.5 | |
| (N: I=1364, C=1579) | (22.4) | (13.9) | [43.6, 49.3] | p=0.11 |
| *Pre-k or primary* | 64.0 | 17.6 | 46.4 | |
| (N: I=286, C=311) | (21.4) | (14.6) | [42.7, 50.1] | |
| *Junior secondary* | 66.6 | 22.6 | 44.0 | |
| (N: I=100, C=123) | (21.0) | (14.0) | [39.9, 48.1] | |
| *Secondary or higher* | 61.4 | 23.3 | 38.1 | |
| (N: I=39, C=43) | (21.9) | (15.4) | [30.4, 45.8] | |

Note: This table shows a series of pre-specified tests for heterogeneity in our primary outcome. As in Table 6, columns (1) and (2) show the mean test score for the group (e.g., in the first row, males in the intervention and control group, respectively) with standard deviations in parentheses below. In column (3), we show the difference between column (1) and (2) adjusted for the randomization stratification factors with a 95% confidence interval (that takes into account the clustered design) in brackets below. Column (4) shows the p-value for a test of the hypothesis that there is no interaction between the categorical heterogeneity variable and receipt of the intervention. The number of observations, by randomized group, is given under each predetermined characteristic, with I signifying the number for the intervention group and C for control.

*4.3 Pre-specified secondary analyses*

In this section, we present differences between the intervention and control groups on a series of pre-specified secondary analyses, including children's enrollment in school, attendance in school, time use, family resource expenditure on the child's schooling, and spillover to siblings and caregivers.

**Enrollment and attendance in school:** We study how the child's enrollment in school and attendance at school varies by treatment status and over time. Enrollment data come from caregiver surveys administered at the end of each academic year. Attendance data, as described below, come in two forms: from these same caregiver surveys and from administrative data collected by the child's school.

In Panel A, we show children's enrollment in school over the three years of the trial. We see that in years two and three of the trial, the odds of enrollment in school are 56% and 92% higher for intervention children than for control group children (raw differences of 9.9 and 11.4 percentage points, respectively). In Panel B, we tabulate which grades children are enrolled in during each of the three academic years. This shows that children in the intervention villages progressed more rapidly through school than those in control villages. In Panel C, we show

absence from school for children who are enrolled in grade 1 or above.[17] Here we have two measures: parents' report of the number of days in the past two weeks the child missed school (higher numbers indicate a greater number of days missed), and administrative data from the child's school on the percent of regularly scheduled school days the child attended (higher numbers mean a greater percent of days attended).[18] Overall, conditional on being enrolled in school, children in intervention villages are less likely to miss school than children in control villages, though the statistical significance of these differences varies from across different measures of attendance.

**Table 8. Enrollment and attendance in school**

*Panel A: Enrollment in School (grade 1 or above) in each academic year (AY)*

| **Enrollment in school** ***(grade 1 or above)*** | **(1) Intervention** | **(2) Control** | **(3) Odds ratio [95% CI]** | **(4) P-value** |
|---|---|---|---|---|
| Year 1 (AY 2015-16) | 47.5% (978) | 42.6% (1046) | 1.21 [0.92, 1.59] | p=0.179 |
| Year 2 (AY 2016-17) | 73.0% (1503) | 63.1% (1551) | 1.56 [1.18, 2.07] | p=0.002 |
| Year 3 (AY 2017-18) | 82.8% (1706) | 71.4% (1756) | 1.92 [1.50, 2.45] | p<0.001 |
| Number of observations | 2060 | 2458 | -- | |

Note: This shows the proportion of children enrolled in grade one or above in each academic year, by intervention status, with the number of enrolled children given below in parentheses. Columns (1), (2), (3), and (4) follow the convention of Table 6.

[17] Note that this parameter is estimated using only children who are enrolled in school. Since, as the trial progresses, the intervention induces some children to stay enrolled in school who otherwise might not have, its interpretation comprises both the attendance and enrollment effects of the intervention.

[18] There are fewer observations in the administrative data because in some cases we were unable either to find or to uniquely identify the child's name in the school register.

*Panel B: What grade child is enrolled in during each academic year*

| | (1) Intervention (N=2060) | (2) Control (N=2458) |
|---|---|---|
| *Year 1: AY 2015-16* | | |
| Not in school | 17.8% (367) | 15.3% (377) |
| ECD*/Nursery | 26.1% (538) | 33.5% (823) |
| 1 | 45.3% (934) | 40.2% (987) |
| 2 | 1.7% (34) | 2.0% (50) |
| 3 | 0.3% (7) | 0.3% (8) |
| 4 or 5 | 0.1% (3) | 0.0% (1) |
| Don't know | 1.8% (37) | 1.6% (40) |
| Missing | 6.8% (140) | 7.0% (172) |
| *Year 2: AY 2016-17* | | |
| Not in School | 11.7% (240) | 15.3% (377) |
| ECD/Nursery | 4.7% (96) | 9.3% (228) |
| 1 | 32.0% (659) | 35.8% (880) |
| 2 | 39.4% (811) | 26.4% (648) |
| 3 | 1.4% (29) | 0.9% (23) |
| 4 or 5 | 0.2% (4) | 0.0% (0) |
| Don't know | 0.0% (0) | 0.0% (0) |
| Missing | 10.7% (221) | 12.3% (302) |
| *Year 3: AY 2017-18* | | |
| Not in School | 2.0% (42) | 2.4% (59) |
| ECD/Nursery | 0.7% (15) | 3.3% (81) |
| 1 | 7.6% (156) | 17.1% (421) |
| 2 | 33.1% (681) | 32.6% (802) |
| 3 | 40.2% (829) | 20.9% (514) |
| 4 or 5 | 1.9% (40) | 0.8% (19) |
| Don't know | 0.0% (0) | 0.0% (0) |
| In school, but grade missing | 4.0% (82) | 6.8% (166) |
| Missing | 10.4% (215) | 16.1% (396) |

Note: In our statistical analysis plan we chose to tabulate these variables, but not to perform hypothesis tests, to minimize the risk of Type I error. As before, the number of observations for a given cell is given in parentheses next to the proportion. ECD stands for the (often informal) early child development classes held in some villages.

*Panel C: Attendance (conditional on enrollment in grade 1 or above)*

| | (1) Intervention | (2) Control | (3) Adjusted difference [95% CI] | (4) P-value |
|---|---|---|---|---|
| *Caregiver report of the number of days the child was absent from school in the two weeks prior to being surveyed* | | | | |
| Year 1: AY 2015-16 (N: I=977, C=1044) | 0.56 (1.65) | 0.75 (2.02) | -0.20 [-0.39, -0.00] | p = 0.047 |
| Year 2: AY 2016-17 (N: I=1500, C=1550) | 0.42 (1.42) | 0.52 (1.71) | -0.09 [-0.24, 0.06] | p = 0.247 |
| Year 3: AY 2016-17 (N: I=1701, C=1748) | 0.52 (1.54) | 0.56 (1.70) | -0.04 [-0.16, 0.08] | p = 0.514 |
| *School's record of child's attendance throughout study* | | | | |
| Percent of regularly scheduled classes child attended during the study (N: I=1565, C=1589) | 81.1 (21.6) | 75.1 (26.2) | 6.0 [1.2, 10.8] | p=0.016 |

Note: Columns (1), (2), (3), and (4) follow the convention of Table 6. The number of observations are given, by treatment arm, under the description of each variable. These vary with the enrollment of children in grade 1 or above across years and the missingness of data in different surveys. In AY 2015-16, there was one enrolled child from the intervention group and two from the control group who reported missing school, but for whom days missed were not recorded. In AY 2016-17, there was one intervention child and three control children with missing data for both of these questions. In AY 2017-18, there were four intervention children and eight control children with missing data for these questions. There was also one intervention child whose guardian reported the child missing school, but the number of days missed was not recorded. We pre-specified that we would also estimate bootstrap confidence intervals, bias corrected and accelerated, based on 2000 bootstrap samples of clusters with stratification by randomized group. These are included in the appendix tables, but because they are so similar to the conventional confidence intervals, we do not include them here.

*Panel D: Caregiver's intention to enroll the child in school in the 2018-19 academic year*

| | (1) Intervention | (2) Control |
|---|---|---|
| *Intention to enroll in AY 2018-19* | | |
| No | 1.1% (22) | 2.1% (51) |
| Yes | 88.3% (1820) | 81.4% (2002) |
| Don't Know | 0.1% (3) | 0.4% (9) |
| Missing | 10.4% (215) | 16.1% (396) |
| Observations | 2060 | 2458 |

Note: In our statistical analysis plan we chose to tabulate this variable but not to carry out a hypothesis test to minimize the risk of Type I error.

**Expenditure and time use**: we next present differences between intervention and control groups in terms of three key family inputs into education: financial expenditure on education, the proportion of the child's waking time spent on education, and the amount of time the child's caregiver spends helping the child with homework. These data were collected from the child's caregiver during the endline survey. We present results in Table 9. Families in the control group may spend slightly more money than those in intervention villages on school-related expenditures, but this difference is not significant at traditional levels.[19] As expected given the time-intensive nature of our intervention, caregivers of children in intervention villages report the child spending significantly more time in school-related activities. We find no evidence of a difference between intervention and control children in the amount of time the caregiver spends helping the child with schoolwork.

**Table 9. Caregiver spending on education, and school-related time use of parents and child**

| Variable | (1) Intervention Mean (SD) | (2) Control Mean (SD) | (3) Adjusted difference [95% CI] | (4) P-value |
|---|---|---|---|---|
| Total parental spend in past year (Gambian Dalasis*) (N: I=1803, C=2003) | 591 (438) | 659 (528) | -66 [-147, 14] | p=0.106 |
| School-related time use of child (proportion of child's waking hours)** (N: I=1845, C=2062) | 0.683 (0.123) | 0.553 (0.140) | 0.130 [0.113, 0.147] | p<0.001 |

[19] While this difference is not significant, its sign is consistent with evidence from other developing country contexts of substitution behavior by parents in response to education interventions (Das et al. 2013).

| | | | | |
|---|---|---|---|---|
| Number of hours caregiver spends helping child with homework per week (N: I=1803, C=2003) | 3.08 (4.27) | 2.99 (4.29) | 0.09 [-0.34, 0.53] | p=0.678 |

Note: Columns (1), (2), (3), and (4) follow the convention of Table 6. We pre-specified that we would also estimate bootstrap confidence intervals, bias corrected and accelerated, based on 2000 bootstrap samples of clusters with stratification by randomized group. These are included in the appendix tables, but because they are so similar to the conventional confidence intervals, we do not include them here. *Gambian Dalasis, average exchange rate over period of trial: 43.72 Dalasis per 1 US dollar. **: School-related time use of child measured as proportion of non-sleeping hours spent in school or on homework. Note that these data come from *all* children in our sample, not just those enrolled in school (as is the case for Table 8, Panel C).

**Spillover to siblings and caregiver:** Next, we analyze whether there were spillover learning gains from intervention children to either their siblings or caregiver. At baseline, we collected a roster of all children under age 17 in the household. At endline, we attempted to find the next older and next younger sibling of each study child according to this roster. If we could find the sibling, we asked the caregiver about the sibling's enrollment in school and school-related time use, and then administered a simple ASER-style literacy and numeracy test to the sibling (Pratham 2010). We also administered such a test to the caregiver.

In Table 10, we tabulate the enrollment of these siblings in school at the end of the trial, and in Table 11, we present our analysis of the endline sibling literacy and numeracy tests. Table 12 presents the caregiver-analogue to Table 11. In Table 10, we see little difference in enrollment in school for either sibling. Table 11 shows that literacy and numeracy for older siblings is similar in intervention and control groups. For younger siblings there is some evidence of improved performance in the intervention group, although the effects are not large. Further, the large number of missing observations among these children weakens our ability to infer any clear message from the statistical analysis. In Table 12, we see no evidence of differences in either literacy or numeracy among caregivers in the two groups.

**Table 10. Sibling enrollment in school and time use**

| Variable | (1) Intervention | (2) Control |
|---|---|---|
| | *Older sibling* | |
| Enrollment in school: | | |
| *Yes* | 61.8% (1076) | 59.8% (1255) |
| *No* | 14.8% (258) | 11.8% (247) |
| *Missing* | 23.4% (408) | 28.4% (596) |
| Observations | 1742 | 2098 |
| School-related time use: proportion (SD)* | 0.583 (0.088) | 0.577 (0.090) |
| Observations | 1076 | 1255 |
| | *Younger sibling* | |
| Enrollment in school: | | |
| *Yes* | 51.1% (923) | 46.4% (1031) |
| *No* | 25.7% (463) | 24.6% (546) |
| *Missing* | 23.2% (419) | 29.0% (643) |
| Observations | 1805 | 2220 |
| School-related time use: proportion (SD)* | 0.511 (0.110) | 0.504 (0.104) |
| Observations | 923 | 1031 |

Note: In our statistical analysis plan we chose to tabulate these variables, but not to perform hypothesis tests, to minimize the risk of Type I error. *: For both older and younger siblings, we have time use only for the siblings who are enrolled in school. Time use here is as in Table 9.

**Table 11. Sibling literacy and numeracy**

*Panel A: Older sibling*

| Variable | (1) Intervention | (2) Control | (3) Odds ratio* [95% CI] | (4) P-value |
|---|---|---|---|---|
| Older sibling literacy: | | | | |
| *Cannot read* | 17.3% (301) | 17.3% (362) | | |
| *Can read at least five letters, but less than five words* | 29.1% (507) | 27.7% (560) | | |

| | | | | |
|---|---|---|---|---|
| *Can read at least five words, but not a sentence* | 9.2% (161) | 10.5% (220) | 1.05 [0.76, 1.43] | p=0.78 |
| *Can read a sentence, but not a paragraph* | 5.2% (91) | 4.2% (89) | | |
| *Can read entire paragraph* | 9.8% (171) | 7.7% (161) | | |
| *Refused* | 0.1% (2) | 0.0% (1) | | |
| *Missing* | 29.2% (509) | 33.6% (705) | | |
| Older sibling numeracy: | | | | |
| *Cannot recognize single digits* | 5.3% (92) | 5.4% (114) | | |
| *Can recognize at least 4 single digit but not two digit numbers* | 13.1% (229) | 11.8% (247) | | |
| *Can recognize at least 4 two digit numbers, but can't add* | 11.8% (205) | 12.1% (253) | 1.02 [0.78, 1.35] | p=0.86 |
| *Can add single digit numbers, but cannot subtract* | 30.7% (534) | 28.4% (596) | | |
| *Can subtract two digit numbers* | 9.8% (171) | 8.7% (182) | | |
| *Refused* | 0.1% (2) | 0.0% (1) | | |
| *Missing* | 29.2% (509) | 33.6% (705) | | |
| Observations | 1742 | 2098 | | |

*Panel B: Younger sibling*

| **Variable** | **(1) Intervention** | **(2) Control** | **(3) Odds ratio* [95% CI]** | **(4) P-value** |
|---|---|---|---|---|
| Younger sibling literacy: | | | | |
| *Cannot read* | 60.6% (1093) | 59.0% (1310) | | |
| *Can read at least five letters, but no words* | 10.7% (194) | 7.6% (169) | | |
| *Can read at least five words, but not a sentence* | 0.6% (11) | 0.5% (10) | 1.39 [1.03, 1.88] | p=0.033 |
| *Can read a sentence, but not a paragraph* | 0.2% (4) | 0.2% (4) | | |

| | | | | |
|---|---|---|---|---|
| *Can read entire paragraph* | 0.2% (4) | 0.0% (1) | | |
| *Refused* | 0.5% (9) | 0.3% (6) | | |
| *Missing* | 27.1% (490) | 32.4% (720) | | |
| Younger sibling numeracy: | | | | |
| *Cannot recognize single digits* | 45.7% (825) | 45.4% (1007) | | |
| *Can recognize at least 4 single digit but not two digit numbers* | 20.3% (366) | 17.2% (382) | | |
| *Can recognize at least 4 two digit numbers, but can't add* | 2.8% (51) | 2.3% (51) | 1.22 [0.97, 1.54] | p=0.097 |
| *Can add single digit numbers, but cannot subtract* | 3.0% (54) | 2.2% (49) | | |
| *Can subtract double digit numbers* | 0.6% (10) | 0.2% (5) | | |
| *Refused* | 0.5% (9) | 0.3% (6) | | |
| *Missing* | 27.1% (490) | 32.4% (720) | | |
| Observations | 1805 | 2220 | | |

Note: This table shows how older and younger siblings of the study child perform on ASER-style literacy and numeracy tests administered at endline. *In column (3), we present the odds ratio and confidence interval from ordered logistic regression models, omitting those in the missing and refused categories, with the relevant p-value given in column (4).

## Table 12. Caregiver literacy and numeracy

| Variable | (1) Intervention | (2) Control | (3) Odds ratio* [95% CI] | (4) P-value |
|---|---|---|---|---|
| Caregiver literacy: | | | | |
| *Cannot read* | 72.3% (1489) | 67.9% (1669) | | |
| *Can read at least five letters, but no words* | 10.5% (216) | 10.1% (248) | | |
| *Can read at least five words, but not a sentence* | 2.9% (59) | 2.4% (58) | 1.02 [0.80, 1.31] | p=0.87 |

| | | | | |
|---|---|---|---|---|
| *Can read a sentence, but not a paragraph* | 0.9% (18) | 0.8% (20) | | |
| *Can read entire paragraph* | 3.1% (63) | 2.6% (65) | | |
| *Refused* | 0.0% (0) | 0.1% (2) | | |
| *Missing* | 10.4% (215) | 16.1% (396) | | |
| Caregiver numeracy: | | | | |
| *Cannot recognize single digits* | 41.4% (852) | 37.7% (926) | | |
| *Can recognize at least 4 single digit but not two digit numbers* | 22.6% (465) | 23.8% (584) | | |
| *Can recognize at least 4 double digit numbers, but can't add* | 7.5% (155) | 6.8% (166) | 1.01 [0.85, 1.20] | p=0.94 |
| *Can add single digit numbers, but cannot subtract* | 13.7% (282) | 11.6% (285) | | |
| *Can subtract double digit numbers* | 4.4% (91) | 4.0% (99) | | |
| *Refused* | 0.0% (0) | 0.1% (2) | | |
| *Missing* | 10.4% (215) | 16.1% (396) | | |
| Observations | 2060 | 2458 | | |

Note: This table follows the same format as Table 11. *:Odds ratios from an ordered logistic regression model, omitting those in the missing and refused categories.

Finally, we estimate whether the intervention had any impact on the activity of school management committees, which are local organizations that help with school operations. This is a test for whether the intervention increased demand for schooling in these communities. We see that all schools attended by children in both intervention and control villages have active committees, consistent with a high level of demand for schooling in these areas. These results are presented in Appendix B, Table 12.

*4.4 Costs and cost-effectiveness*

In this section, we discuss the cost of the intervention as it was implemented during the course of the study, how this may vary in future implementation, and the likely cost under government implementation. Our implementation involved entering the country and establishing a service delivery apparatus entirely independent of the government. Our initial calculation includes all the attendant costs that this incurred. While this strategy may be optimal in some contexts[20], in many others the government should be able to implement a version of this intervention using current staff, either after-school or during the school day, as in Piper et al. (2018).

Our main costs comprise the following components: teacher, monitor, and other staff salaries, benefits, and incidentals; design, piloting, printing, binding, and shipping of teaching and learning materials; purchasing, importing, maintaining, and fueling vehicles; the construction, renovation, and maintenance of a main office and a field office; staff training expenses (food, lodging, transport, per diems, and training materials); and various administrative costs (such as accountancy, taxes, insurance, HR) that come with running a stand-alone organization of roughly 150 employees. We capitalize vehicle costs and, separately, office construction and other capital expenses over an expected lifespan of 10 and 20 years, respectively. We express our figures in 2015 dollars and use an annual discount rate equivalent to the US Consumer Price Index for that year.[21]

Total expenditure for running the intervention was 1.493 million US dollars. We calculate the per-child cost of implementing this intervention by dividing the total expenditure by the 2,060 children enrolled at baseline in villages assigned to receive the intervention. This

[20] For example, in a recently completed study run in rural Guinea Bissau, because of the dysfunction of the school system, the authors decided to implement a similar intervention in lieu of – as opposed to in addition to – the government primary schools.

[21] Taken from www.usinflationcalculator.com/inflation/current-inflation-rates/ on May 15, 2019

generates a per-child cost of $724.77, or $241.59 per child per year. In this implementation, we focused primarily on maximizing the fidelity of the intervention and not on cost minimization. In future implementation at larger scale, we estimate that our per-child cost may decrease by up to 30 to 40 percent.

To estimate the cost for the government to implement the project, we use the following assumptions. First, we assume existing primary school teachers would implement our scripted lessons during some of the school day. Two, we assume that the only costs the government would incur in implementing this intervention would be related to 1) training teachers in the new materials; 2) the increased monitoring our intervention entails;[22] and 3) procuring the textbooks and other teaching and learning materials we use. For an initial rollout, we estimate this cost to be $137.11 overall ($45.70 per year). The four main components to this are training costs, monitoring costs (including hiring additional monitors), teaching materials costs (textbooks, lesson plans), and transport costs for both monitors and for training sessions. As teachers learn the material and as the government scales up (with which potential economics of scale in procurement and travel would accrue), we anticipate this cost too would drop substantially.

Using the "additional SD per $100" metric from Kremer et al. (2013), and using only the 1,815 children who took the endline test as the denominator, we estimate a 0.389 SD increase per $100 spent with the existing costs. We expect some loss of fidelity with government implementation (Bold et al. 2018), so we do not calculate this ratio for government implementation. While this is more expensive than many other interventions (Burde and Linden 2013; Lucas et al. 2014), we generate much larger absolute learning gains than these studies. Our gains are also larger than those of all other interventions cited in Kremer et al. (2013). Our costs

[22] The main costs here are 1) hiring new mid-level government employees to serve as a field monitor, at a rate of seven to eight villages per monitor, and 2) the costs of their transport between villages to conduct this monitoring.

in The Gambia are roughly 3.5 times larger than the per-child costs of the original intervention in India, and generate an SD impact 4.26 times larger (Lakshminarayana et al. 2013).

This model – very large gains earned in return for relatively large costs – is similar to the "graduation" model studied in Banerjee et al. (2015) which uses a multifaceted intervention to establish sustainable self-employment and generate lasting well-being improvements among the extremely poor. Together, these results suggest that greater expenditure may be necessary to reap such large learning gains in other particularly poor and remote areas.

## 5. Discussion

In this section, we address two questions. First, why are our impact estimates so large? Second, what do our results tell us about potential policy options to address the very low learning levels in contexts similar to the one we study?

### *5.1 Magnitude of effect estimates*

We estimate that the intervention had an extremely large impact on children's learning levels. We believe that the magnitude of this impact stems from two factors: one, the comprehensive nature of the intervention and possible complementarities between its component parts; and two, the particularly low learning levels in this context, which make such large gains possible.

The intervention combines at least three strategies shown in prior research to be effective in similar settings. First, the short-term nature of para teacher contracts has been shown to increase performance elsewhere, incentivizing teachers to exert greater effort (Banerjee et al. 2007; Duflo, Dupas, and Kremer 2015; Muralidharan and Sundararaman 2013). Second, we conducted high-frequency monitoring of our para teachers. The main purpose of this monitoring

was to improve teacher effectiveness through providing regular feedback on teaching methods and practice, also called "coaching" in recent research (Kraft, Blazar, and Hogan 2018; Muralidharan et al. 2017; Piper, Destefano, et al. 2018a). Three, we used a curriculum, shown to be effective elsewhere, that included scripted daily lesson plans and greater teacher-student interaction (Banerjee et al. 2017; Lakshminarayana et al. 2013). In light of the substantially smaller estimated effects of the prongs when implemented individually (e.g., Banerjee et al. 2017; Muralidharan et al. 2017; Muralidharan and Sundararaman 2013), we argue that our results provide some suggestive evidence of complementarity between these interventions, though our study was not designed to test this hypothesis. This is consistent with recent work from Tanzania showing complementarities between programs that change teacher incentives and those which provide additional resources to schools (Mbiti et al. 2019).

The other likely main contributor to these large relative gains is the extremely low learning levels in rural regions of The Gambia. The area in which we work is remote, difficult to reach, and very poor. Learning levels in these areas have been consistently low: government assessments of third grade children in villages similar to those in our trial show performance roughly similar to that of our control children. Other studies have found large learning gains from delivering interventions to similar settings and populations (Burde and Linden 2013; Lakshminarayana et al. 2013).

There are two important alternative explanations for the large effects we estimate. One is the potential for teaching to the test by our intervention team. The second is potential enumerator bias or leakage of the test paper. In terms of teaching to the test, our intervention used materials adapted from the materials used in Lakshminarayana et al. (2013) to match the Gambian curriculum. As described in Section 2, the Gambian Ministry of Basic and Secondary Education

regularly uses EGRA and EGMA tests to assess students and teachers. As a result, government teachers were incentivized to teach the specific skills assessed in the EGRA and EGMA tests. In addition, the higher-level skills tested here are quite general – word recognition, reading comprehension, and arithmetic. The performance gap between intervention and control children is greater for these skills than for the more basic skills, such as letter naming or number sequences, that might be considered more test-specific. This pattern runs contrary to what we would expect to see should there have been teaching to the test.

We took great care to minimize the risk of enumerator bias and leakage of the test paper. With regards to enumerator bias, EGRA and EGMA have rigid rules for implementation. Our training and supervision of the enumerators who administered the tests emphasized close adherence to these rules. Enumerators were recruited independently of our other research activities, and were not told of a village's assignment to either intervention or control. Furthermore, two authors (Eble and Hsieh) travelled to The Gambia to supervise the assessments in order to guard against such bias.

Finally, we took great lengths to ensure our test paper was not leaked. Specifically, while we conducted an initial pilot and adaptation session in The Gambia as per the EGRA and EGMA guidelines, the final items on the test were finalized in the UK by Hsieh and brought to The Gambia for printing only in May 2018, a few days prior to the training of enumerators. During the test administration, Eble and Hsieh conducted occasional interviews with children to test for possible cheating on the test. To do so, we identified high performers on the test and informally interviewed them to see whether they could answer spontaneous questions of similar difficulty to higher level tasks on the test, but with different content. In all cases, they performed in line with their performance on the test itself, further suggesting no evidence that the test paper was leaked.

Finally, in the data analysis, we did not notice any unusual outcomes that would suggest the test leaked to any specific clusters/classes, nor did we notice unusual results by question, e.g., that children performed, on average, more poorly on the easier questions than the more difficult ones.

*5.2 Lessons for education policy in rural areas of developing countries*

We argue that there are three main policy implications of our findings. The first is that this intervention offers the potential for large learning gains in this type of setting. It combines several best practices from the literature – altering teachers' contract structures; using improved curricular materials, including scripted lessons; and monitoring teachers frequently with the goal of improving their teaching practice – and has dramatically raised learning outcomes in two diverse contexts: central regions of The Gambia and southern Telangana, India. This work has definitively shown that contracting this type of intervention out to an NGO *can* lead to dramatic improvements in educational outcomes in such areas.

The second implication is that large gains may be possible within the confines of existing systems. While operationalizing this type of intervention is likely to be challenging for government (Bold et al. 2018), the component pieces of our intervention – a change in contract structure; improved, scripted curricular materials; and more frequent monitoring with a focus on pedagogic improvement – are all hypothetically implementable within the existing school hours and the existing school budgets of such systems. Recent work has shown that an intervention which focuses on raising literacy levels and which combines similar components was scalable in government schools in Kenya, albeit in areas with higher wealth and resource levels than rural Gambia (Piper, Destefano, et al. 2018b; Piper, Zuilkowski, and Mugenda 2014). The core limitation of this approach is that such changes may be demotivating or otherwise threatening to

existing teachers. These teachers are important stakeholders in the education system who may exert political pressure to oppose these changes. As a result, it may be challenging to maintain the large effect sizes of such programs when moving to government implementation (Bold et al. 2018).

Our findings suggest two possible paths for governments of countries facing similar problems. Obviously, government implementation of such a program within the existing funding system and school day would be ideal, and is worth aiming at. However, in nations where either 1) there exists strong political opposition to changes in teacher contracts and supervision, 2) government capacity, including that for financing, is low, or 3) both, government implementation may simply not be possible. Indeed, some of our collaborators in The Gambia have argued that it would be politically difficult to implement the type of restructuring this would imply for the Gambian education system. When such a change in bureaucracy is not possible, our results show that there remains room for large gains in learning levels through contracting these services out to NGOs. Even in this case, however, financing issues and perhaps others will remain.

Finally, our results suggest that large, supply-side interventions may be necessary to raise learning levels in particularly disadvantaged areas similar to the one we study. Several data points suggest high demand for education in the areas we study. These include the high levels of enrollment in school, both in our study and in the national gross enrollment data, and the common presence of school management committees, parent-run groups which aid local schools' functioning. Despite this, learning levels in these areas remain tragically low. Our study introduced experimental variation in the supply of education, and we find that this corresponds to massive learning gains. Together, we interpret these findings as evidence of the primary importance of the supply side in raising basic learning levels in remote areas of extreme poverty.

## 6. Conclusion

We find that a para teacher intervention combining frequent monitoring with an improved curriculum, shown to raise literacy and numeracy levels dramatically in rural India among primary-aged children, had even larger effects on literacy and numeracy levels among such children in rural parts of The Gambia. The gains we measure are very large, and provide important evidence of one way to substantially reduce the large rural/urban and developing/developed country gaps in learning. Our intervention combines three well-known methods for increasing learning levels: para teachers, frequent monitoring with a focus on improving teachers' practice, and an improved, scripted set of daily lesson plans. These, and low baseline learning levels in our study area, contribute to the dramatic results that we document here. While we cannot disentangle the relative importance of the different components, our findings in The Gambia – in addition to our prior study of a similar intervention in India – show that dramatic learning gains can be achieved across multiple challenging contexts through implementation of this para teacher-focused supplementary education program.

The key limitation of our study is that we do not know the potential for such an intervention to yield similar gains when implemented by the government, nor even if such implementation is possible in settings with strong political pressure in favor of the status quo arrangement for delivering primary education. We also do not have sufficient evidence to know conclusively whether it is one of the three prongs of the intervention, or their complementary nature, that generates these results.

Our findings have two clear implications for policy: one, that there exists a demonstrated way to reach the learning gains that many have called for in particularly disadvantaged areas. Two, that the choice facing governments hoping to reap such gains in similar areas is whether to

attempt to operationalize this within the government system, or contract it out. The most immediate research needed on this subject is assessing if and how government might do this in different contexts.

## References


Bandiera, Oriana, Robin Burgess, Erika Deserranno, Ricardo Morel, Imran Rasul, and Munshi Sulaiman. 2018. "Social Ties and the Delivery of Development Programs." *Working Paper*, October.

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland, and Michael Walton. 2017. "From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application." *Journal of Economic Perspectives* 31 (4): 73–102.

Banerjee, Abhijit, Shawn Cole, Esther Duflo, and Leigh Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *The Quarterly Journal of Economics* 122 (3): 1235–64. https://doi.org/10.1162/qjec.122.3.1235.

Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry. 2015. "A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries." *Science* 348 (6236): 1260799.

Blimpo, Moussa P., David K. Evans, and Nathalie Lahire. 2011. "School-Based Management and Educational Outcomes: Lessons from a Randomized Field Experiment." *Unpublished Manuscript*.

Blimpo, Moussa P., Ousman Gajigo, and Todd Pugatch. 2016. "Financial Constraints and Girls' Secondary Education: Evidence from School Fee Elimination in the Gambia." *The World Bank Economic Review* 33 (1): 185–208.

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

Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, Alice Ng'ang'a, and Justin Sandefur. 2018. "Experimental Evidence on Scaling up Education Reforms in Kenya." *Journal of Public Economics* 168 (December): 1–20. https://doi.org/10.1016/j.jpubeco.2018.08.007.

Boone, Peter, Alpha Camara, Alex Eble, Diana Elbourne, Samory Fernandes, Chris Frost, Chitra Jayanty, Maitri Lenin, and Ana Filipa Silva. 2015. "Remedial After-School Support Classes Offered in Rural Gambia (The SCORE Trial): Study Protocol for a Cluster Randomized Controlled Trial." *Trials* 16 (1): 574.

Bruhn, Miriam, and David McKenzie. 2009. "In Pursuit of Balance: Randomization in Practice in Development Field Experiments." *American Economic Journal: Applied Economics*, 200–232.

Burde, Dana, and Leigh L. Linden. 2013. "Bringing Education to Afghan Girls: A Randomized Controlled Trial of Village-Based Schools." *American Economic Journal: Applied Economics* 5 (3): 27–40.

Campbell, Marion K., Gilda Piaggio, Diana R. Elbourne, and Douglas G. Altman. 2012. "Consort 2010 Statement: Extension to Cluster Randomised Trials." *BMJ* 345: e5661.

Chaudhury, N., J. Hammer, M. Kremer, K. Muralidharan, and F. H. Rogers. 2006. "Missing in Action: Teacher and Health Worker Absence in Developing Countries." *The Journal of Economic Perspectives* 20 (1): 91–116.

Das, Jishnu, Stefan Dercon, James Habyarimana, Pramila Krishnan, Karthik Muralidharan, and Venkatesh Sundararaman. 2013. "School Inputs, Household Substitution, and Test Scores." *American Economic Journal: Applied Economics* 5 (2): 29–57.

Ding, Lin, Jia He, and Frederick Koon Shing Leung. 2014. "Relations between Subject Matter Knowledge and Pedagogical Content Knowledge: A Study of Chinese Pre-Service Teachers on the Topic of Three-Term Ratio." *The Mathematics Educator* 15 (2): 50–76.

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

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2015. "School Governance, Teacher Incentives, and Pupil–Teacher Ratios: Experimental Evidence from Kenyan Primary Schools." *Journal of Public Economics* 123 (March): 92–110. https://doi.org/10.1016/j.jpubeco.2014.11.008.

Evans, David K., and Anna Popova. 2016. "What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews." *The World Bank Research Observer* 31 (2): 242–70. https://doi.org/10.1093/wbro/lkw004.

Even, Ruhama. 1993. "Subject-Matter Knowledge and Pedagogical Content Knowledge: Prospective Secondary Teachers and the Function Concept." *Journal for Research in Mathematics Education*, 94–116.

Ganimian, Alejandro, and Richard Murnane. 2016. "Improving Education in Developing Countries: Lessons From Rigorous Impact Evaluations." *Review of Educational Research* 86 (3): 719–55.

Glewwe, Paul. 2002. "Schools and Skills in Developing Countries: Education Policies and Socioeconomic Outcomes." *Journal of Economic Literature* 40 (2): 436–482.

Glewwe, Paul, and Karthik Muralidharan. 2016. "Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications." In *Handbook of the Economics of Education*, 5:653–743. Elsevier.

Kraft, Matthew A., David Blazar, and Dylan Hogan. 2018. "The Effect of Teacher Coaching on Instruction and Achievement: A Meta-Analysis of the Causal Evidence." *Review of Educational Research* 88 (4): 547–588.

Kremer, Michael, Conner Brannen, and Rachel Glennerster. 2013. "The Challenge of Education and Learning in the Developing World." *Science* 340 (6130): 297–300.

Lakshminarayana, Rashmi, Alex Eble, Preetha Bhakta, Chris Frost, Peter Boone, Diana Elbourne, and Vera Mann. 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.

Liang, Kung-Yee, and Scott L. Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika* 73 (1): 13–22.

Lucas, Adrienne M., Patrick J. McEwan, Moses Ngware, and Moses Oketch. 2014. "Improving Early-Grade Literacy In East Africa: Experimental Evidence From Kenya And Uganda." *Journal of Policy Analysis and Management* 33 (4): 950–976.

Mbiti, Isaac, Karthik Muralidharan, Mauricio Romero, Youdi Schipper, Constantine Manda, and Rakesh Rajani. 2019. "Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania*." *The Quarterly Journal of Economics* 134 (3): 1627–73. 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.” *Review of Educational Research* 85 (3): 353–394.

Moher, D., S. Hopewell, K. F. Schulz, V. Montori, P. C. Gøtzsche, P. J. Devereaux, D. Elbourne, M. Egger, and D. G. Altman. 2010. “CONSORT 2010 Explanation and Elaboration: Updated Guidelines for Reporting Parallel Group Randomised Trials.” *BMJ* 340 (c869).

Muralidharan, Karthik, Jishnu Das, Alaka Holla, and Aakash Mohpal. 2017. “The Fiscal Cost of Weak Governance: Evidence from Teacher Absence in India.” *Journal of Public Economics* 145: 116–35.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2011. “Teacher Performance Pay: Experimental Evidence from India.” *Journal of Political Economy* 119 (1): 39–77.

———. 2013. “Contract Teachers: Experimental Evidence from India.” *NBER Working Paper 19440*.

Piper, Benjamin, Joseph Destefano, Esther M. Kinyanjui, and Salome Ong’ele. 2018a. “Scaling up Successfully: Lessons from Kenya’s Tusome National Literacy Program.” *Journal of Educational Change* 19 (3): 293–321.

———. 2018b. “Scaling up Successfully: Lessons from Kenya’s Tusome National Literacy Program.” *Journal of Educational Change* 19 (3): 293–321.

Piper, Benjamin, Yasmin Sitabkhan, Jessica Mejía, and Kellie Betts. 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*.

Piper, Benjamin, Stephanie Simmons Zuilkowski, and Abel Mugenda. 2014. “Improving Reading Outcomes in Kenya: First-Year Effects of the PRIMR Initiative.” *International Journal of Educational Development* 37: 11–21.

Platas, L. M., L. Ketterlin-Gellar, A. Brombacher, and Y. Sitabkhan. 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://books.google.com/books?hl=en&lr=&id=PQ72AAAAQBAJ&oi=fnd&pg=PR1&dq=pritchett+schooling+aint+learning&ots=uvSg4RtJhA&sig=1jSzmH3E1acmSrT3eRBDQCyjXwA.

Rajani, Rakesh. 2010. *Are Our Children Learning? Annual Learning Assessment Report Tanzania 2010*. Twaweza.

Salehi, Ahmad Shah, Abdul Tawab Kawa Saljuqi, Nadia Akseer, Krishna Rao, and Kathryn Coe. 2018. “Factors Influencing Performance by Contracted Non-State Providers Implementing a Basic Package of Health Services in Afghanistan.” *International Journal for Equity in Health* 17 (1): 128.

Shalem, Yael, Carola Steinberg, Hannchen Koornhof, and Francine De Clercq. 2016. “The What and How in Scripted Lesson Plans: The Case of the Gauteng Primary Language and Mathematics Strategy.” *Journal of Education* 66: 13–36.

Sprenger-Charolles, Liliane. 2008. “The Gambia : Early Grade Reading Assessment.” *World Bank Policy Report*. https://openknowledge.worldbank.org/handle/10986/12972.

The World Bank. 2018a. “Economic Data on The Gambia” Website, accessed December 16, 2018. URL: https://data.worldbank.org/country/gambia-the?view=chart.

———. 2018b. “Education Data on The Gambia” Website, accessed December 18, 2018. URL: http://datatopics.worldbank.org/education/country/gambia,-the.

Wood, Lesley, Matthias Egger, Lise Lotte Gluud, Kenneth F. Schulz, Peter Jüni, Douglas G. Altman, Christian Gluud, Richard M. Martin, Anthony JG Wood, and Jonathan AC Sterne. 2008. “Empirical Evidence of Bias in Treatment Effect Estimates in Controlled Trials with Different Interventions and Outcomes: Meta-Epidemiological Study.” *BMJ: British Medical Journal* 336 (7644): 601.

## Appendix – for online publication only

Appendix contents:

A. Further description of intervention components.

B. Full tables as specified in Statistical Analysis Plan

C. Sample size calculation

D. Final test papers

E. Per-protocol analyses

---

**Appendix A: Further description of intervention components.**

***Intervention set-up***

**Preparation of Teaching and Learning Materials (TLM)**

A pedagogic team developed TLM for the community educators and children by following three steps:

1. We broke down broad topics in the national curriculum in English and mathematics into sub-topics or smaller units to teach daily.
2. We wrote scripted lesson plans detailing what a teacher should do and say every day to teach a given concept or subtopic. These lesson plans included student activities based on the following concepts: the "cooperative reflective" approach; differentiation; and active learning strategies.
3. We created worksheets, reading cards, and additional materials for children to facilitate learning of these concepts.

**Schedule/format**

The intervention was designed as a series of after-school teaching sessions with two hours of instruction per session, but the curriculum and materials were designed to be flexible enough to accommodate different scheduling arrangements.

**Selection of Field Officers**

The community educators were supervised by a team of Field Officers. These individuals were required to have had prior experience of teaching and a primary teaching certificate. They also had to pass a written test and an in-person interview, and then successfully complete training before being selected.

**Selection of Community Educators**

With each community, we identified Community Educators (CEs), either from the recipient community itself or one nearby, who met age and education criteria. Identified candidates had to pass a test and successfully complete training (unlike Field Officers, there was no preliminary in-person interview for CEs) before being hired. Final hiring decisions were made at the end of the training.

**Responsibilities of Field Officers and Program Supervisors**

Each field officer was responsible for monitoring and giving pedagogic support to 6 – 7 CEs throughout delivery of the intervention. Allocation of CEs was done to minimize travel time between classrooms (and thus to maximize CE exposure to monitoring). Field Officers were tasked with monitoring and observing the CE, spending a full day in the village with the CE helping them before and after the academic class to prepare and teach the children. An additional tier of monitor – Program Supervisors – monitored the Field Officers, providing additional pedagogic support to the Field Officers and monitoring progress in children's learning levels.

### *Implementation*

**Recruitment**

As described above, recruitment of CEs and Field Officers followed a predetermined screening procedure. For those candidates who progressed to the final round of selection, we also required

successful completion of a multi-week teacher training, based on subject matter knowledge and pedagogic content knowledge. This involved further observation and evaluation. We provided a further week of initial monitoring and feedback to Field Officers.

**Community sensitization**

We sought to involve the community and seek their engagement in each stage of the process of establishing the intervention, including selecting the community educator(s) from their village, identifying a suitable venue for the classes to be held, and selecting a suitable time for the children to learn (because many children attended both government-run schools and Qur'anic school, selection of an appropriate time window in which to hold the classes was an important part of community sensitization).

**Training**

CEs received three types of training over the course of the intervention:

1. **Annual training:** before the beginning of each academic year, we conducted annual trainings with CEs to teach the grade-specific concepts they were responsible to teach that year. This involved extensive training in the scripted lesson plans, use of the worksheets and other teaching and learning materials, as well as more general teacher training.
2. **Review meetings:** regular review meetings were conducted with CEs. This included feedback sessions and input sessions. Feedback sessions included interaction with the CEs where best practices and issues identified during monitoring were discussed. Both good and bad practices were discussed in order that participating CEs could learn from the experience of other CEs. Input sessions were small training sessions which introduced CEs to the topics to be taught in the following month.
3. **In-cluster training:** once a quarter, we conducted in-cluster training with community educators to help them improve their subject matter knowledge in basic literacy and numeracy and to improve their teaching technique.

**Monitoring**

As described above, we monitored our teachers using three levels of staff – Field Officers, Program Supervisors, and the executive staff of our intervention team. While monitoring from Field Officers focused on specific pedagogic and curricular issues to improve CE teaching, monitoring from program supervisors was intended primarily to help the field officers and community educators in solving issues at the classroom level and community level as they arose.

**Enriched Teaching and Learning Materials**

We attempted to create a curriculum with easy to follow, scripted daily lesson plans, also known as "structured pedagogy" (Piper, Sitabkhan, et al. 2018; Shalem et al. 2016). These lesson plans were aimed to help CEs in structuring their days' lessons, with the hope that this would accomplish the following goals:

1. Facilitate training of concepts to CEs who may not otherwise fully master the material;
2. Free up CE mental bandwidth to more nimbly respond to child needs;
3. Help with child attention management. Additional learning materials for children are meant to make them more attentive to and interested in the academic classes.

**Community involvement**

As described in the body of the paper, we engaged the communities in which we worked from the start of this project, with the notion that greater community buy-in would ensure greater attendance and demand for our classes. We also held regular community meetings whenever we anticipated that community support was needed to solve an upcoming issue.

**Regular feedback**

We held monthly review meetings in which we provided feedback to CEs on their teaching, using both specific observations from monitoring each teacher as well as broader feedback about practices, offering a regular opportunity to correct mistakes and allowing CEs to learn from their peers, share best practices, and build morale.

**Assessment**

The intervention team also conducted regular classroom assessments of children's skill levels across key concepts as the trial progressed, and reallocated resources –CEs, supervision, and supplementary material – to help struggling students or centers as needed.

## Appendix B: Full tables as specified in Statistical Analysis Plan:

**Table1: Baseline characteristics of clusters**

| Variable | Intervention arm N=56 | Control arm N=55 |
|---|---|---|
| Region: North Bank/Lower River | 36:20 | 35:20 |
| Distance to road: <median/>median | 28:28 | 27:28 |
| Distance to road (km): mean (SD) | 2.00 (2.92) | 1.65 (2.80) |
| Randomised eligible children: mean (SD) | 36.8 (20.8) | 44.7 (35.7) |
| Villages per cluster: | | |
| 1 | 35 | 32 |
| 2 | 18 | 16 |
| 3 | 1 | 5 |
| 4 | 2 | 2 |
| Cluster population: mean (SD) | 1188 (556) | 1415 (1007) |

**Table 2. Baseline characteristics of individuals for body of paper**

| Variable | Intervention arm | | Control arm | |
|---|---|---|---|---|
| | Individual level N=2060 | Cluster level N=56 mean (SD) | Individual level N=2458 | Cluster level N=55 mean (SD) |
| Gender | | | | |
| Male | 1070 (51.9%) | 53.3% (8.8%) | 1239 (50.4%) | 50.6% (8.8%) |
| Female | 989 (48.0%) | 46.7% (8.8%) | 1217 (49.5%) | 49.4% (8.9%) |
| Not Known | 1 (0.0%) | 0.1% (0.5%) | 2 (0.1%) | 0.1% (0.4%) |
| Ethnic group: | | | | |
| Mandinka | 842 (40.9%) | 41.1% (38.0%) | 1040 (42.3%) | 48.9% (41.7%) |
| Wolof | 334 (16.2%) | 19.1% (28.6%) | 608 (24.7%) | 19.5% (33.2%) |
| Fula | 516 (25.0%) | 25.9% (28.8%) | 485 (19.7%) | 22.0% (29.9%) |
| Jola | 24 (1.2%) | 1.0% (2.4%) | 17 (0.7%) | 0.5% (1.4%) |
| Serere | 133 (6.5%) | 5.7% (11.7%) | 85 (3.5%) | 2.5% (5.7%) |
| Other | 155 (7.5%) | 7.2% (11.1%) | 177 (7.2%) | 6.7% (15.0%) |
| Missing | 56 (2.7%) | 2.6% (3.4%) | 46 (1.9%) | 2.0% (2.6%) |
| Child's main carer: | | | | |
| Biological mother | 1511 (73.3%) | 75.1% (12.3%) | 1833 (74.6%) | 75.6% (11.7%) |
| Biological father | 69 (3.3%) | 3.8% (6.9%) | 128 (5.2%) | 5.3% (7.6%) |
| Grandmother | 227 (11.0%) | 11.2% (6.4%) | 249 (10.1%) | 10.8% (6.7%) |
| Grandfather | 11 (0.5%) | 0.7% (1.8%) | 18 (0.7%) | 0.7% (1.7%) |
| Step/foster mother | 84 (4.1%) | 4.1% (3.8%) | 99 (4.0%) | 4.3% (3.5%) |
| Step/foster father | 11 (0.5%) | 0.8% (2.5%) | 13 (0.5%) | 0.4% (1.4%) |
| Other care-giver | 94 (4.6%) | 4.3% (4.7%) | 78 (3.2%) | 2.9% (3.3%) |
| Missing | 53 (2.6%) | 2.4% (3.5%) | 40 (1.6%) | 1.9% (2.9%) |

| | | | | |
|---|---|---|---|---|
| Mother's education: | | | | |
| No education | 1420 (68.9%) | 72.4% (16.7%) | 1772 (72.1%) | 73.6% (13.8%) |
| Pre-K/Nursery | 2 (0.1%) | 0.0% (0.3%) | 0 (0.0%) | 0.0% (0.0%) |
| Primary | 330 (16.0%) | 16.2% (11.5%) | 382 (15.5%) | 15.3% (9.2%) |
| Junior Secondary | 138 (6.7%) | 6.3% (6.6%) | 161 (6.6%) | 6.8% (7.0%) |
| Senior Secondary | 79 (3.8%) | 3.5% (4.3%) | 69 (2.8%) | 3.0% (4.3%) |
| Post-Secondary | 9 (0.4%) | 0.4% (1.0%) | 4 (0.2%) | 0.3% (2.1%) |
| Other | 9 (0.4%) | 0.5% (1.8%) | 3 (0.1%) | 0.1% (0.5%) |
| Don't know | 14 (0.7%) | 0.7% (1.7%) | 22 (0.9%) | 0.9% (1.6%) |
| Missing | 59 (2.9%) | 2.7% (3.8%) | 45 (1.8%) | 2.1% (3.0%) |
| Father's education: | | | | |
| No education | 1309 (63.5%) | 67.6% (15.4%) | 1732 (70.5%) | 73.2% (13.4%) |
| Pre-K/Nursery | 0 (0.0%) | 0.0% (0.0%) | 0 (0.0%) | 0.0% (0.0%) |
| Primary | 188 (9.1%) | 9.3% (6.5%) | 158 (6.4%) | 6.0% (4.7%) |
| Junior Secondary | 143 (6.9%) | 6.3% (5.4%) | 147 (6.0%) | 6.2% (6.9%) |
| Senior Secondary | 178 (8.6%) | 8.1% (6.4%) | 214 (8.7%) | 8.3% (6.9%) |
| Post-Secondary | 34 (1.7%) | 1.4% (2.9%) | 43 (1.7%) | 1.6% (2.8%) |
| Other | 65 (3.2%) | 3.0% (4.4%) | 60 (2.4%) | 2.2% (3.5%) |
| Don't know | 88 (4.3%) | 4.2% (4.6%) | 61 (2.5%) | 2.5% (3.2%) |
| Missing | 55 (2.7%) | 2.5% (3.5%) | 43 (1.7%) | 2.1% (3.0%) |
| Caregiver's education: | | | | |
| No education | 1520 (73.8%) | 76.8% (14.6%) | 1861 (75.7%) | 77.6% (12.8%) |
| Pre-K/Nursery | 1 (0.0%) | 0.0% (0.2%) | 0 (0.0%) | 0.0% (0.0%) |
| Primary | 317 (15.4%) | 15.3% (9.8%) | 352 (14.3%) | 14.3% (8.7%) |
| Junior Secondary | 110 (5.3%) | 4.8% (5.3%) | 142 (5.8%) | 5.7% (6.1%) |
| Senior Secondary | 41 (2.0%) | 2.1% (4.1%) | 51 (2.1%) | 2.1% (3.2%) |
| Post-Secondary | 7 (0.3%) | 0.3% (1.3%) | 3 (0.1%) | 0.2% (1.1%) |
| Other | 6 (0.3%) | 0.4% (2.4%) | 3 (0.1%) | 0.1% (0.5%) |
| Don't know | 2 (0.1%) | 0.2% (1.0%) | 0 (0.0%) | 0% (0.0%) |
| Missing | 56 (2.7%) | 2.6% (3.4%) | 46 (1.9%) | 2.0% (2.6%) |
| Child's age in September 2015 | 6.87 (SD=0.55) | 6.86 (0.13) | 6.88 (SD=0.55) | 6.87 (0.12) |
| Language spoken in home: | | | | |
| Mandinka | 869 (42.2%) | 42.6% (40.7%) | 1093 (44.5%) | 52.3% (43.6%) |
| Wolof | 379 (18.4%) | 21.2% (31.9%) | 691 (28.1%) | 20.6% (35.7%) |
| Fula | 506 (24.6%) | 25.5% (30.4%) | 440 (17.9%) | 20.6% (30.1%) |
| Jola | 4 (0.2%) | 0.1% (0.7%) | 3 (0.1%) | 0.1% (0.8%) |
| Serere | 110 (5.3%) | 4.3% (11.6%) | 27 (1.1%) | 0.9% (4.1%) |
| Other | 136 (6.6%) | 6.2% (11.1%) | 158 (6.4%) | 5.5% (14.8%) |
| Missing | 56 (2.7%) | 2.6% (3.4%) | 46 (1.9%) | 2.0% (2.6%) |
| Mother status at baseline | | | | |
| Alive | 1976 (95.9%) | 96.3% (3.7%) | 2384 (97.0%) | 96.6% (3.4%) |
| Dead | 31 (1.5%) | 1.3% (2.0%) | 34 (1.4%) | 1.4% (2.4%) |
| Not known | 53 (2.6%) | 2.4% (3.5%) | 40 (1.6%) | 1.9% (2.9%) |
| Father status at baseline | | | | |
| Alive | 1900 (92.2%) | 92.7% (6.0%) | 2262 (92.0%) | 90.8% (7.1%) |
| Dead | 107 (5.2%) | 5.0% (4.3%) | 156 (6.3%) | 7.2% (6.2%) |

| | | | | |
|---|---|---|---|---|
| Not known | 53 (2.6%) | 2.4% (3.5%) | 40 (1.6%) | 1.9% (2.9%) |
| Carer literacy at baseline: | | | | |
| Can't read | 1557 (75.6%) | 78.7% (12.3%) | 1915 (77.9%) | 79.4% (12.2%) |
| Can read at least one letter, but not an entire word | 204 (9.9%) | 8.9% (6.4%) | 209 (8.5%) | 9.2% (6.8%) |
| Can read at least one word, but not entire card | 80 (3.9%) | 4.0% (4.9%) | 93 (3.8%) | 3.4% (3.4%) |
| Read entire card slowly | 57 (2.8%) | 3.0% (3.2%) | 89 (3.6%) | 3.9% (5.3%) |
| Read entire card fluently | 106 (5.1%) | 5.4% (7.2%) | 105 (4.3%) | 4.1% (6.0%) |
| Refused | 0 (0.0%) | 0.0% (0.0%) | 1 (0.0%) | 0.0% (0.1%) |
| Missing | 56 (2.7%) | 2.6% (3.4%) | 46 (1.9%) | 2.0% (2.6%) |
| Wealth | | | | |
| Category 1 | 151 (7.3%) | 9.3% (15.6%) | 113 (4.6%) | 5.3% (9.1%) |
| Category 2 | 1373 (66.7%) | 67.9% (19.8%) | 1635 (66.5%) | 71.9% (18.5%) |
| Category 3 | 480 23.3%) | 22.8% (17.2%) | 664 (27.0%) | 22.8% (16.2%) |
| Missing | 56 (2.7%) | 2.6% (3.4%) | 46 (1.9%) | 2.0% (2.6%) |
| Older Sibling | | | | |
| Yes | 1742 (84.6%) | 84.7% (6.7%) | 2098 (85.4%) | 86.5% (8.0%) |
| No | 318 (15.4%) | 15.3% (6.7%) | 360 (14.6%) | 13.5% (8.0%) |
| Younger Sibling | | | | |
| Yes | 1805 (87.6%) | 88.2% (8.3%) | 2220 (90.3%) | 90.6% (8.9%) |
| No | 255 (12.4%) | 11.8% (8.3%) | 360 (9.7%) | 9.4% (8.9%) |

Note: In this and later appendix tables which provide descriptive statistics, the number of observations is given first, followed by the percentage of the sample in parentheses. We also present cluster means and standard deviations, as labeled in the column headings.

**Table 3. Class schedules and adherence, intervention arm only**

| Schedule | | | Number of villages |
|---|---|---|---|
| Six two-hour classes per week | | | 78 |
| Five two-hour classes per week (four at two hours, and one at four hours) | | | 2 |
| Four classes per week (two at two weekday classes at two hours per day, and two weekend classes at four hours per day) | | | 2 |
| | Adherence measure 1 (village-level) N=82 | Adherence measure 2 (village-level) N=82 | Adherence measure 3 (child-level) N=2060 |
| Mean | 100% | 78.9% | 76.8% |
| SD | 0% | 8.7% | 27.8% |
| 0 | 0 | 0 | 15 (0.7%) |
| >0 to 25% | 0 | 0 | 209 (10.1%) |
| >25% to 50% | 0 | 0 | 77 (3.7%) |
| >50% to 75% | 0 | 25 (30.5%) | 210 (10.2%) |
| >75% to 100% | 82 (100%) | 57 (69.5%) | 1537 (74.6%) |
| Missing | 0 | 0 | 12 (0.6%) |

**Table 4. Children resident in study village (migration)**

| Variable | Intervention arm N=2060 | | | Control arm N = 2458 | | |
|---|---|---|---|---|---|---|
| | Year 1 | Year 2 | Year 3 | Year 1 | Year 2 | Year 3 |
| Yes | 1920 (93.2%) | 1839 (89.3%) | 1845 (89.6%) | 2286 (93.0%) | 2156 (87.7%) | 2062 (83.9%) |
| No | 135 (6.6%) | 174 (8.4%) | 197 (9.6%) | 164 (6.7%) | 256 (10.4%) | 372 (15.1%) |
| Missing | 5 (0.2%) | 47 (2.3%) | 18 (0.9%) | 8 (0.3%) | 46 (1.9%) | 24 (1.0%) |

**Table 5. EGRA and EGMA test results**

| Variable | Intervention arm | | Control arm | | Difference (95% CI) p-value |
|---|---|---|---|---|---|
| | Individual level N: mean (SD) | Cluster level N=56 mean (SD) | Individual level N: mean (SD) | Cluster level N=55 mean (SD) | |
| Composite test score | 1815: 63.3 (22.3) | 63.5 (7.2) | 2071: 17.1 (14.2) | 16.8 (6.5) | 46.0 (43.3, 48.8) p<0.0001 |
| (SD units) | | | | | 3.23 (2.89, 3.63)* |
| | | | | | |
| Mathematics test score, overall | 1815: 68.2 (21.8) | 68.4 (7.7) | 2071: 24.7 (19.7) | 24.1 (8.6) | 43.4 (40.2, 46.5) p<0.0001 |
| | | | | | |
| Mathematics 1 | 1815: 93.2 (19.0) | 93.3 (5.1) | 2071: 46.5 (34.1) | 46.7 (15.3) | |
| Mathematics 2 | 1815: 86.6 (22.7) | 86.7 (5.8) | 2071: 39.5 (34.6) | 38.8 (13.8) | |
| Mathematics 3 | 1815: 60.8 (25.6) | 61.0 (9.1) | 2071: 18.0 (17.0) | 17.6 (7.9) | |
| Mathematics 4 | 1815: 65.0 (30.2) | 65.1 (11.0) | 2071: 12.7 (18.6) | 11.4 (7.1) | |
| Mathematics 4a | 1815: 63.8 (28.5) | 64.1 (8.9) | 2071: 17.1 (21.3) | 15.7 (8.4) | |
| Mathematics 4b | 1815: 66.3 (36.2) | 66.2 (13.9) | 2071: 8.3 (19.3) | 7.1 (6.6) | |
| Mathematics 5 | 1815: 51.9 (30.3) | 52.2 (12.3) | 2071: 6.5 (12.8) | 5.9 (5.0) | |
| Mathematics 5a | 1815: 49.3 (26.5) | 49.7 (9.2) | 2071: 9.5 (16.0) | 8.7 (5.8) | |
| Mathematics 5b | 1815: 54.5 (38.9) | 54.7 (16.2) | 2071: 3.5 (12.9) | 3.0 (5.0) | |
| Mathematics 6 | 1815: 51.6 (28.4) | 52.0 (11.4) | 2071: 25.2 (22.2) | 24.1 (8.0) | |
| | | | | | |
| Mathematics 1 (fluency) | 1815: 38.7 (18.1) | 39.0 (6.7) | 2071: 15.0 (14.1) | 14.6 (6.1) | |
| Mathematics 4a (fluency) | 1815: 18.8 (16.0) | 18.9 (5.2) | 2071: 5.6 (9.0) | 5.1 (2.6) | |
| Mathematics 5a (fluency) | 1815: 15.8 (15.2) | 16.0 (4.6) | 2071: 3.5 (7.4) | 3.1 (2.0) | |
| | | | | | |

| | | | | | |
|---|---|---|---|---|---|
| Reading test score, overall | 1815: 58.3 (25.3) | 58.6 (8.0) | 2072: 9.5 (11.2) | 9.6 (5.4) | 48.7 (46.1, 51.4) p<0.0001 |
| | | | | | |
| Reading 1 | 1815: 68.5 (24.8) | 68.9 (8.8) | 2072: 14.0 (17.7) | 14.5 (9.9) | |
| Reading 2 | 1815: 52.3 (27.4) | 52.5 (8.3) | 2072: 28.3 (22.4) | 28.8 (7.9) | |
| Reading 3 | 1815: 51.9 (29.9) | 52.6 (9.0) | 2072: 4.9 (13.6) | 5.3 (6.3) | |
| Reading 4 | 1815: 64.7 (34.1) | 64.8 (10.1) | 2072: 5.4 (14.4) | 5.2 (5.9) | |
| Reading 5a | 1815: 62.1 (31.5) | 62.5 (9.7) | 2072: 6.0 (13.7) | 5.8 (5.9) | |
| Reading 5b | 1815: 47.2 (32.3) | 47.4 (11.3) | 2072: 2.9 (8.5) | 2.6 (2.5) | |
| Reading 6 | 1815: 61.5 (38.5) | 61.7 (13.3) | 2072: 5.3 (15.7) | 4.6 (4.5) | |
| | | | | | |
| Reading 1 (fluency) | 1815: 70.0 (27.2) | 70.5 (10.1) | 2071: 14.0 (17.9) | 14.5 (9.8) | |
| Reading 3 (fluency) | 1815: 27.8 (18.0) | 28.2 (5.7) | 2072: 2.5 (7.1) | 2.7 (3.2) | |
| Reading 4 (fluency) | 1815: 39.8 (25.9) | 40.0 (9.0) | 2072: 2.8 (8.1) | 2.7 (3.1) | |
| Reading 5a (fluency) | 1815: 42.3 (25.0) | 42.5 (8.1) | 2072: 3.8 (9.0) | 3.7 (3.7) | |
| | | | | | |

*Bootstrap confidence interval, bias corrected and accelerated, based on 2000 bootstrap samples of clusters with stratification by randomized arm.

**Table 5a. EGRA and EGMA test results – composite test score, per protocol analyses**

| Variable | Intervention arm | | Control arm | | Difference (95% CI) p-value |
|---|---|---|---|---|---|
| | Individual level N: mean (SD) | Cluster level N: mean (SD) | Individual level N: mean (SD) | Cluster level N: mean (SD) | |
| Village level adherence > 75% | 1147: 65.2 (21.3) | 40 65.0 (7.7) | 2071: 17.1 (14.2) | 55 16.8 (6.5) | 47.9 (45.0, 50.9) p<0.0001 |
| (SD units) | | | | | 3.37 (3.00, 3.76)* |
| Child level adherence > 75% | 1525: 67.4 (18.9) | 56 67.3 (6.1) | 2071: 17.1 (14.2) | 55 16.8 (6.5) | 50.2 (47.7, 52.7) P<0.0001 |
| | | | | | 3.52 (3.19, 3.93)* |

*Bootstrap confidence interval, bias corrected and accelerated, based on 2000 bootstrap samples of clusters with stratification by randomized arm.

**Table 6. Composite test scores by subgroup, with interaction tests**

| Subgroup | Intervention arm | | Control arm | | Difference (95% CI) | p-value |
|---|---|---|---|---|---|---|
| | Individual level N: mean (SD) | Cluster level N: mean (SD) | Individual level N: mean (SD) | Cluster level N: mean (SD) | | |
| Sex | | | | | | |
| Male | 935: 62.4 (23.1) | 56: 62.8 (8.2) | 1022: 16.4 (14.1) | 55: 16.4 (6.9) | 45.9 (42.9, 49.0) | p=0.86 |
| Female | 879: 64.1 (21.4) | 56: 63.9 (9.5) | 1047: 17.9 (14.2) | 55: 17.0 (7.1) | 46.2 (43.2, 49.2) | |
| Wealth | | | | | | |
| Category 1 | 138: 60.9 (24.9) | 36: 57.7 (19.1) | 93: 15.2 (12.5) | 23 14.9 (9.9) | 45.6 (38.5, 52.7) | p=0.53 |
| Category 2 | 1227: 63.5 (21.8) | 56: 64.2 (8.0) | 1396: 16.7 (13.8) | 55: 16.6 (6.8) | 46.7 (43.9, 49.6) | |
| Category 3 | 429: 63.7 (22.1) | 51: 64.0 (11.4) | 570: 18.6 (15.3) | 51: 17.9 (10.0) | 45.1 (41.5, 48.7) | |
| Ethnicity | | | | | | |
| Mandinka | 751: 63.4 (21.6) | 41: 63.6 (8.6) | 901: 16.7 (13.6) | 41: 17.1 (8.7) | 46.5 (42.3, 50.6) | |
| Wolof | 295: 63.8 (24.6) | 25: 60.1 (20.6) | 504: 16.4 (14.2) | 24: 13.6 (5.9) | 47.4 (43.0, 51.8) | p=0.20 |
| Fula | 467: 63.5 (21.9) | 46: 61.6 (15.6) | 411: 20.8 (14.9) | 42: 17.8 (7.3) | 42.6 (38.8, 46.3) | |
| Other | 281: 62.7 (21.6) | 38: 62.0 (13.7) | 243: 14.2 (13.6) | 29: 13.4 (8.6) | 49.1 (43.2, 55.0) | |
| Region | | | | | | |
| Lower River | 677: 63.4 (21.9) | 20: 63.4 (8.0) | 688: 19.5 (14.9) | 20: 20.2 (7.3) | 43.9 (38.8, 48.9) | p=0.27 |
| North Bank | 1138: 63.2 (22.5) | 36: 63.5 (6.9) | 1383: 15.9 (13.6) | 35: 14.9 (5.3) | 47.2 (44.0, 50.4) | |
| Distance to road | | | | | | |
| >median | 731: 63.9 (22.8) | 28: 63.8 (8.6) | 836: 16.2 (13.4) | 28: 16.7 (7.0) | 47.6 (43.3, 52.0) | p=0.34 |
| <median | 1084: 62.8 (21.9) | 28: 63.2 (5.6) | 1235: 17.8 (14.6) | 27: 17.0 (6.2) | 45.0 (41.6, 48.3) | |
| Caregiver Education | | | | | | |
| None | 1364: 63.0 (22.4) | 56: 63.3 (8.0) | 1579: 16.5 (13.9) | 55: 16.8 (6.8) | 46.5 (43.6, 49.3) | p=0.11 |
| Pre-k or primary | 286: 64.0 (21.4) | 53: 64.6 (11.5) | 311: 17.6 (14.6) | 50: 15.8 (8.3) | 46.4 (42.7, 50.1) | |
| Junior secondary | 100: 66.6 (21.0) | 36: 66.7 (15.1) | 123: 22.6 (14.0) | 34: 20.0 (9.8) | 44.0 (39.9, 48.1) | |
| Secondary+ | 39: 61.4 (21.9) | 18: 59.6 (17.6) | 43: 23.3 (15.4) | 23: 23.1 (12.6) | 38.1 (30.4, 45.8) | |

Note: Not all categories (e.g. ethnicity) are represented in every cluster, hence the need for numbers of clusters to be reported for the cluster level analysis.

**Table 7. Enrollment and attendance in school**

| Variable | Intervention arm | | Control arm | | |
|---|---|---|---|---|---|
| Enrollment (grade 1 or above) | Individual level N =2060 | Cluster level N=56 mean (SD) | Individual level N =2458 | Cluster level N=55 mean (SD) | Odds ratio (95% CI) p-value |
| Year 1 | 978 (47.5%) | 49.7% (19.7%) | 1046 (42.6%) | 45.5% (22.2%) | 1.21 (0.92, 1.59) p=0.179 |
| Year 2 | 1503 (73.0%) | 74.7% (18.4%) | 1551 (63.1%) | 65.0% (20.0%) | 1.56 (1.18, 2.07) p=0.002 |
| Year 3 | 1706 (82.8%) | 83.7% (9.7%) | 1756 (71.4%) | 71.4% (15.4%) | 1.92 (1.50, 2.45) p<0.001 |
| | Intervention arm N=2060 | | Control arm N=2458 | | |
| Year 1 grade: Not in school | 367 (17.8%) | | 377 (15.3%) | | |
| Year 1 grade: ECD/Nursery | 538 (26.1%) | | 823 (33.5%) | | |
| Year 1 grade: 1 | 934 (45.3%) | | 987 (40.2%) | | |
| Year 1 grade: 2 | 34 (1.7%) | | 50 (2.0%) | | |
| Year 1 grade: 3 | 7 (0.3%) | | 8 (0.3%) | | |
| Year 1 grade: 4 | 3 (0.1%) | | 0 (0.0%) | | |
| Year 1 grade: 5 | 0 (0.0%) | | 1 (0.0%) | | |
| Year 1 grade: Don't know | 37 (1.8%) | | 40 (1.6%) | | |
| Year 1 grade: Missing | 140 (6.8%) | | 172 (7.0%) | | |
| Year 2 grade: Not in School | 240 (11.7%) | | 377 (15.3%) | | |
| Year 2 grade: ECD/Nursery | 96 (4.7%) | | 228 (9.3%) | | |
| Year 2 grade: 1 | 659 (32.0%) | | 880 (35.8%) | | |
| Year 2 grade: 2 | 811 (39.4%) | | 648 (26.4%) | | |
| Year 2 grade: 3 | 29 (1.4%) | | 23 (0.9%) | | |
| Year 2 grade: 4 | 4 (0.2%) | | 0 (0.0%) | | |
| Year 2 grade: 5 | 0 (0.0%) | | 0 (0.0%) | | |
| Year 2 grade: Don't know | 0 (0.0%) | | 0 (0.0%) | | |
| Year 2 grade: Missing | 221 (10.7%) | | 302 (12.3%) | | |
| Year 3 grade: Not in School | 42 (2.0%) | | 59 (2.4%) | | |
| Year 3 grade: ECD/Nursery | 15 (0.7%) | | 81 (3.3%) | | |
| Year 3 grade: 1 | 156 (7.6%) | | 421 (17.1%) | | |
| Year 3 grade: 2 | 681 (33.1%) | | 802 (32.6%) | | |
| Year 3 grade: 3 | 829 (40.2%) | | 514 (20.9%) | | |
| Year 3 grade: 4 | 36 (1.7%) | | 18 (0.7%) | | |
| Year 3 grade: 5 | 4 (0.2%) | | 1 (0.0%) | | |
| Year 3 grade: Don't know | 0 (0.0%) | | 0 (0.0%) | | |
| Year 3 grade: In School, but grade missing | 82 (4.0%) | | 166 (6.8%) | | |
| Year 3 grade: Missing | 215 (10.4%) | | 396 (16.1%) | | |
| Attendance (if enrolled at grade 1 or above) | Individual level N: mean (SD) | Cluster level N: mean (SD) | Individual level N: mean (SD) | Cluster level N: mean (SD) | Difference (95% CI) p-value |
| Parental Report Year 1 | 977*: 0.56 (1.65) | 55***: 0.51 (0.46) | 1044** 0.75 (2.02) | 54***: 0.87 (1.04) | -0.20 (-0.39, -0.00) (-0.40, -0.00)[#] |

| | | | | | |
|---|---|---|---|---|---|
| Days missed in past two weeks | | | | | p = 0.047 |
| Parental Report Year 2 Days missed in past two weeks | 1500[+++]: 0.42 (1.42) | 55***: 0.38 (0.31) | 1550[+]: 0.52 (1.71) | 54***: 0.52 (0.47) | -0.09 (-0.24, 0.06) (-0.24, 0.05)[#] p = 0.247 |
| Parental Report Year 3 | 1701[$] 0.52 (1.54) | 56: 0.53 (0.42) | 1748[$$] 0.56 (1.70) | 55: 0.53 (0.39) | -0.04 (-0.16, 0.08) (-0.16, 0.09)[#] p = 0.514 |
| School Report % days attended | 1565: 81.1 (21.6) | 53: [$$$] 82.0 (9.8) | 1589: 75.1 (26.2) | 52: [$$$] 75.6 (13.7) | 6.0 (1.2, 10.8) p=0.016 |
| | Intervention arm N=2060 | | Control arm N=2458 | | |
| Intention to enroll in year 4 No | 22 (1.1%) | | 51 (2.1%) | | |
| Yes | 1820 (88.3%) | | 2002 (81.4%) | | |
| Don't Know | 3 (0.1%) | | 9 (0.4%) | | |
| Missing | 215 (10.4%) | | 396 (16.1%) | | |

[#]Bootstrap confidence interval, bias corrected and accelerated, based on 2000 bootstrap samples of clusters with stratification by randomized arm.

*One enrolled child reported missing school, but days missed were not recorded.

** Two enrolled children reported missing school, but days missed were not recorded.

***One cluster in each arm had no enrolled children at grade 1 or above.

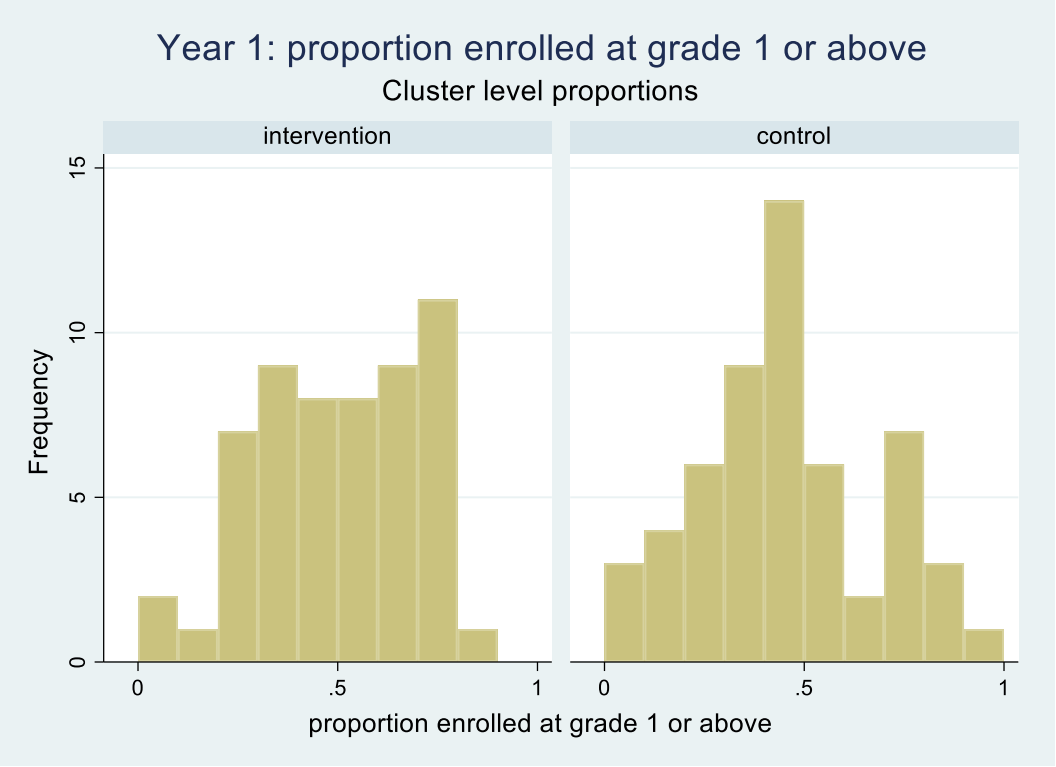
[+]One enrolled child with missing response

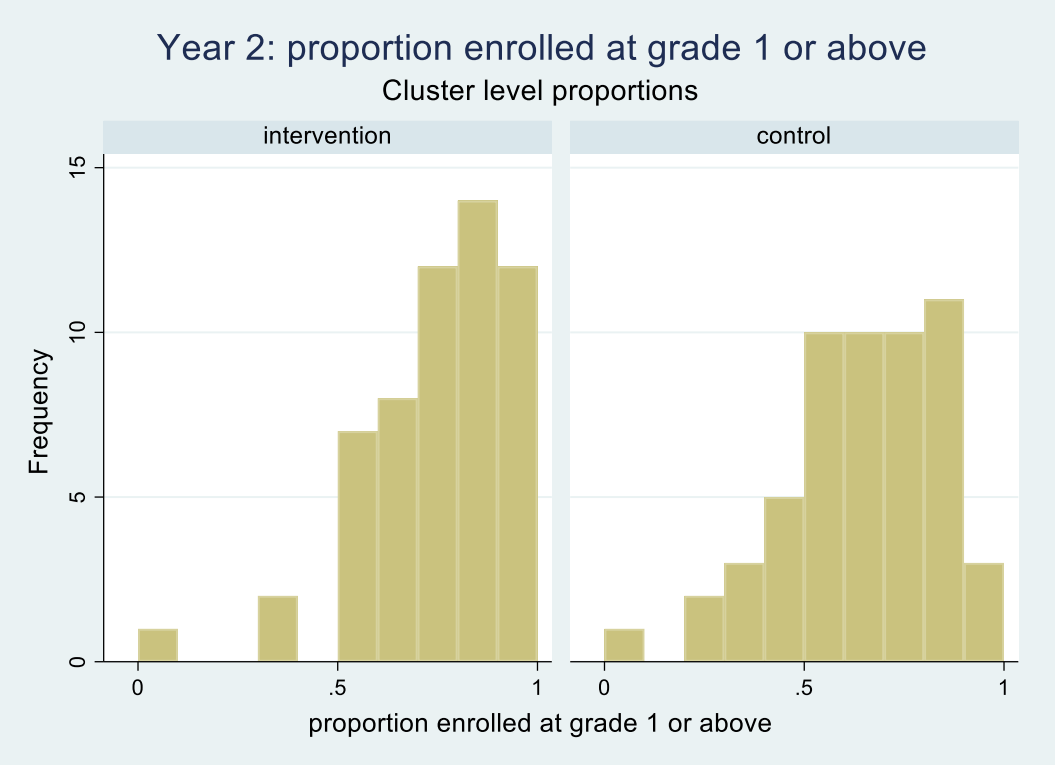
[+++]Three enrolled children with missing responses

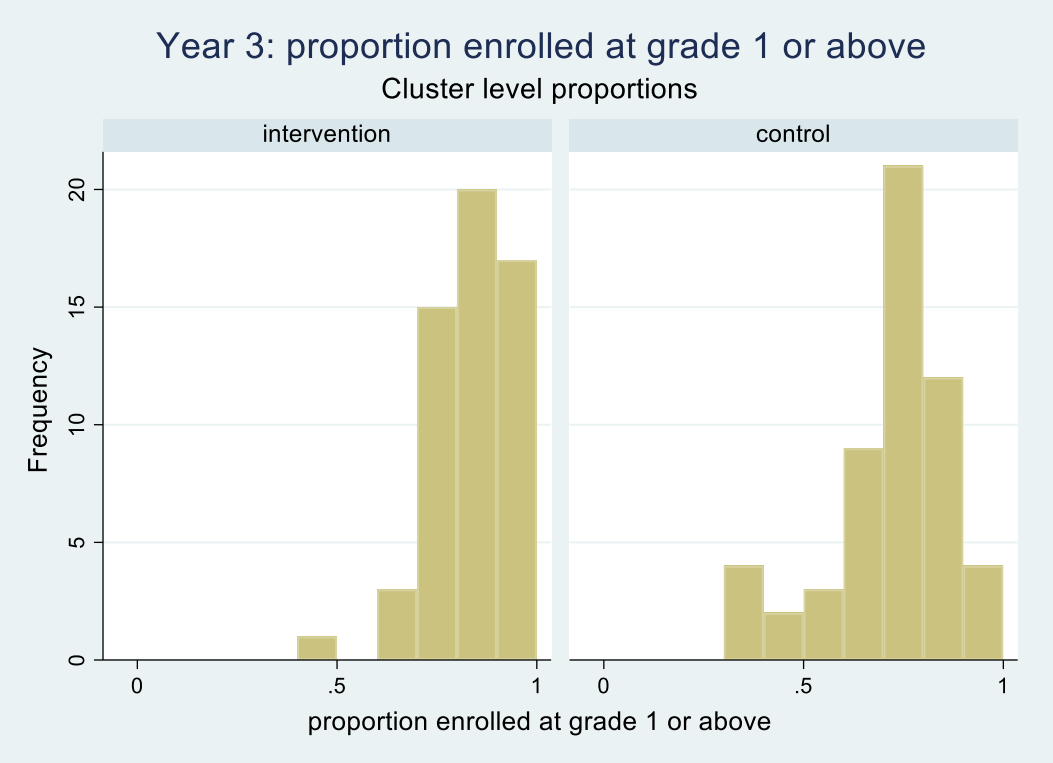
[$]Four enrolled children with missing response, and one reported missing school, but days missed were not recorded.

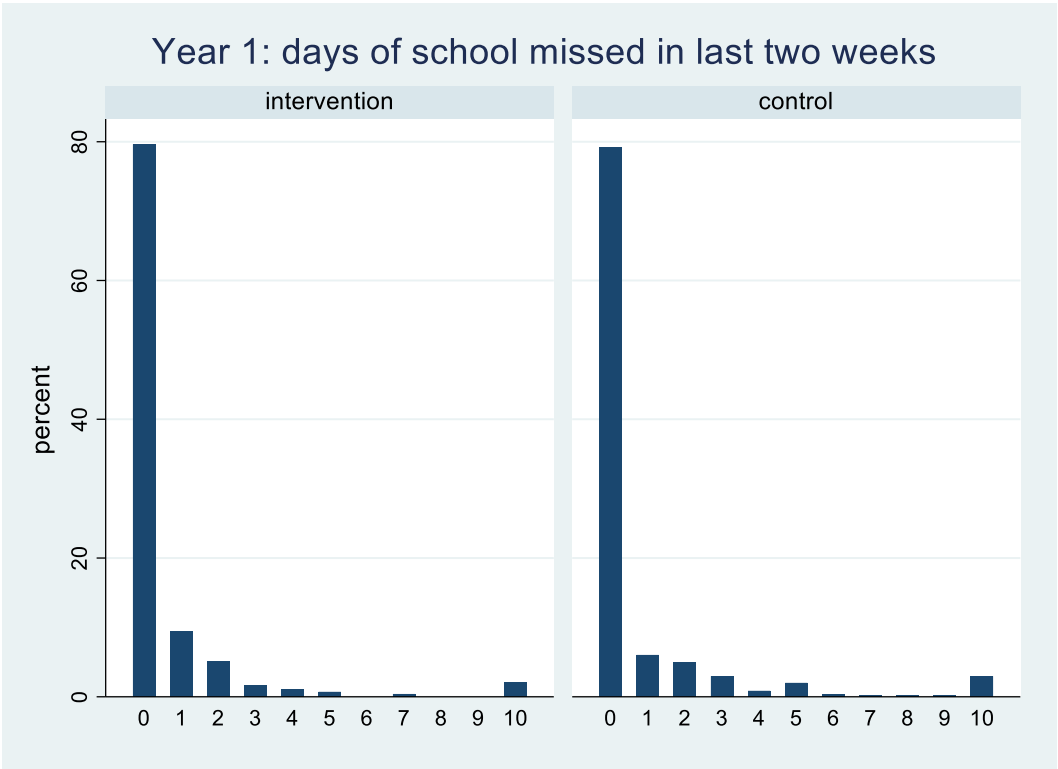
[$$]Eight enrolled children with missing response.

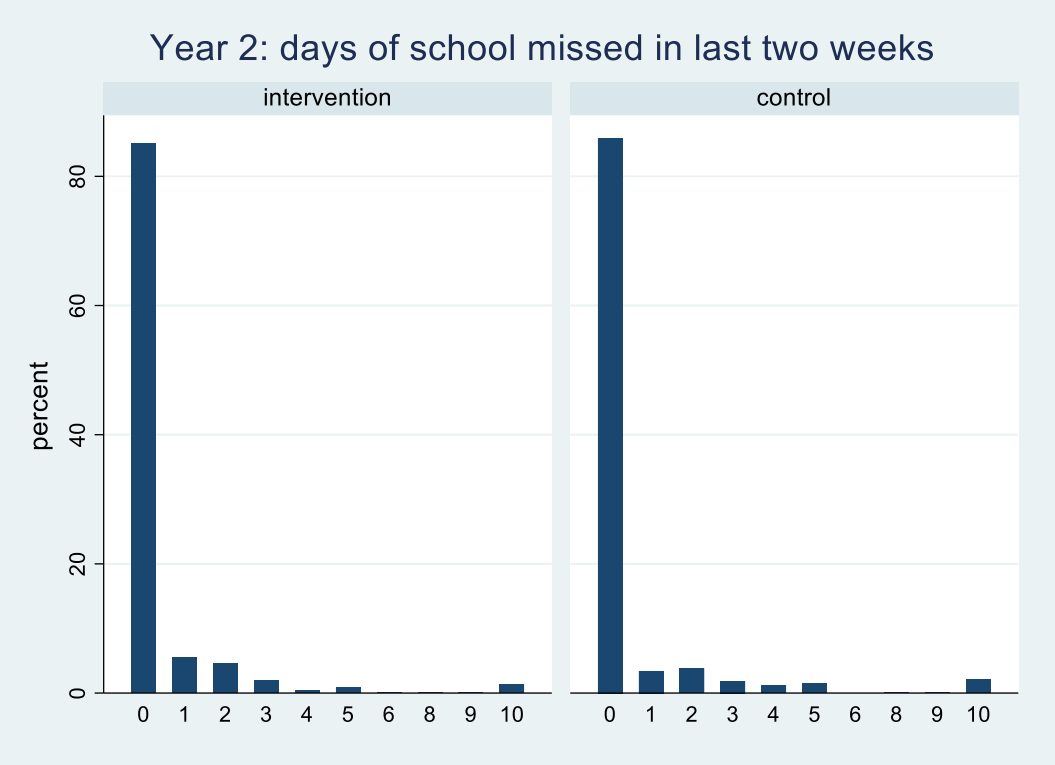
[$$$]Three clusters in each arm had no reported data.


Year 1: proportion enrolled at grade 1 or above
Cluster level proportions
intervention
control
Frequency
0
5
10
15
0
.5
1
proportion enrolled at grade 1 or above


Year 2: proportion enrolled at grade 1 or above
Cluster level proportions
intervention
control
Frequency
0
5
10
15
0
.5
1
proportion enrolled at grade 1 or above


Year 3: proportion enrolled at grade 1 or above
Cluster level proportions
intervention
control
Frequency
0
5
10
15
20
0
.5
1
proportion enrolled at grade 1 or above


Year 1: days of school missed in last two weeks
intervention
control
percent
0
20
40
60
80
0 1 2 3 4 5 6 7 8 9 10


Year 2: days of school missed in last two weeks
intervention
control
percent
0
20
40
60
80
0 1 2 3 4 5 6 8 9 10
0 1 2 3 4 5 6 8 9 10

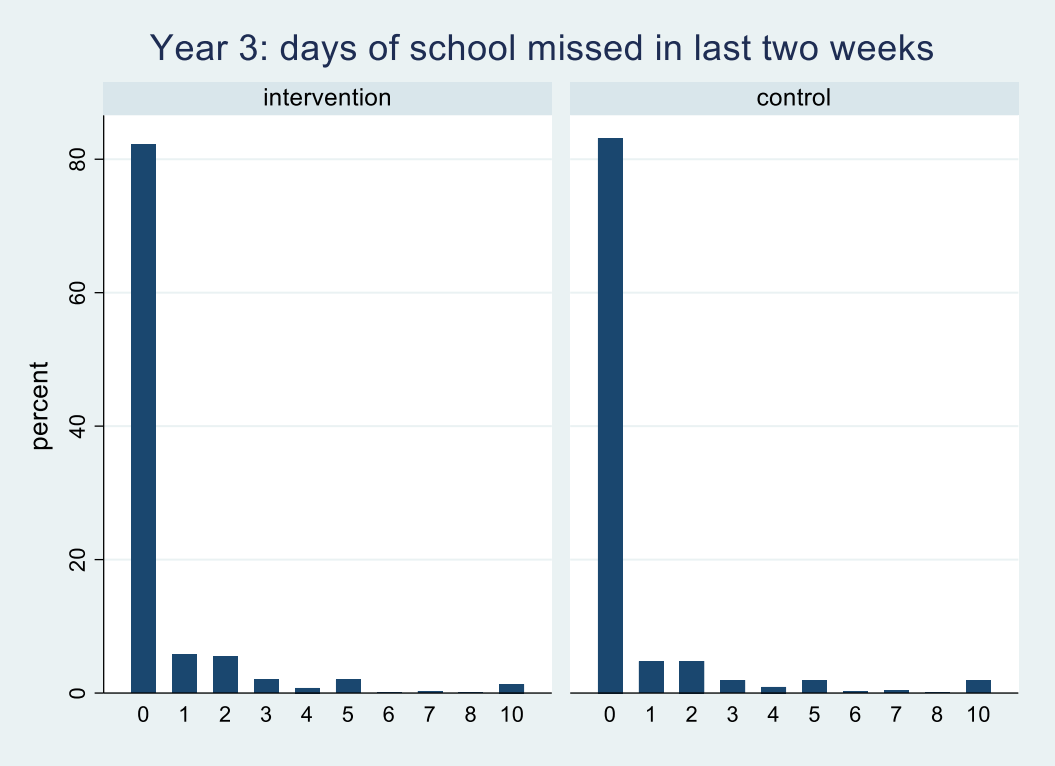
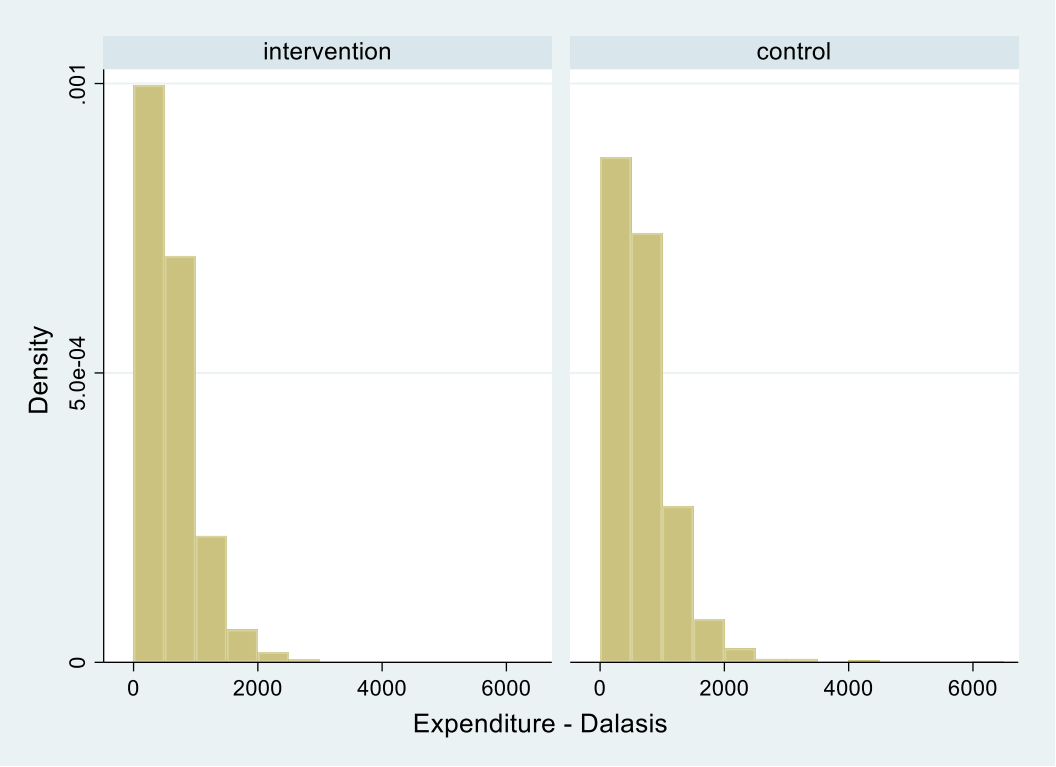
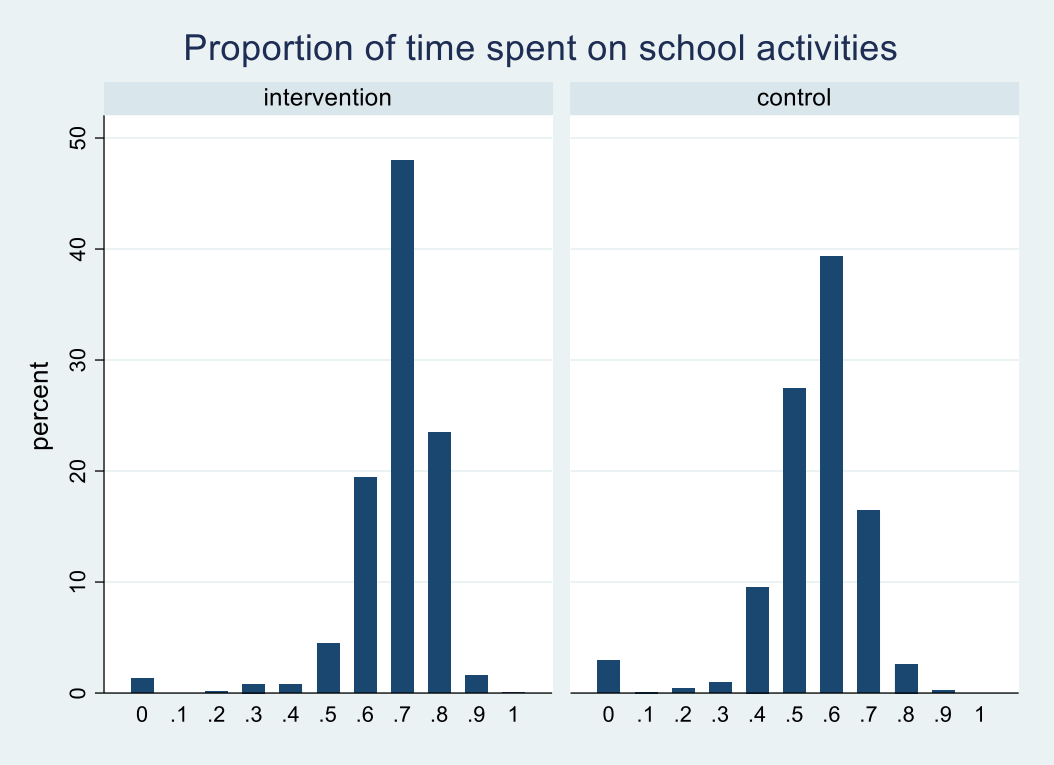

Year 3: days of school missed in last two weeks
intervention
control
percent
0
20
40
60
80
0 1 2 3 4 5 6 7 8 10
0 1 2 3 4 5 6 7 8 10

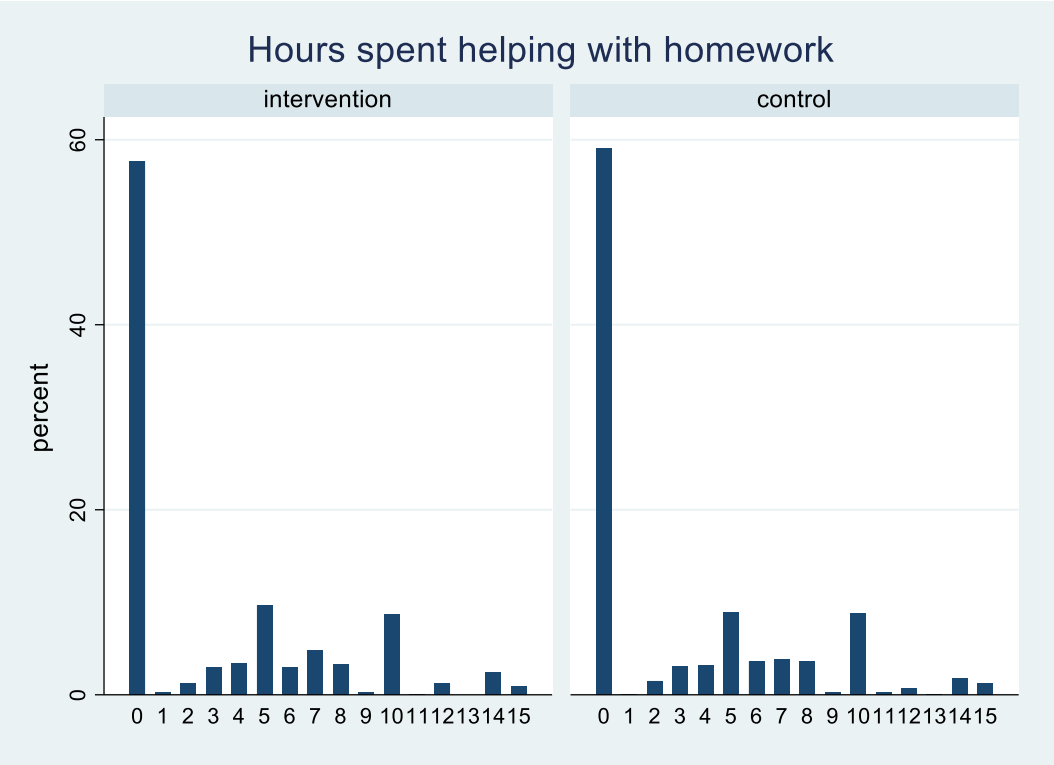
**Table 8. Parental spending on education, and school-related time use of parents and child**

| Variable | Intervention arm | | Control arm | | Difference (95% CI) p-value |
|---|---|---|---|---|---|
| | Individual level N: mean (SD) | Cluster level N=56 mean (SD) | Individual level N: mean (SD | Cluster level N=55 mean (SD) | |
| Total parental spend (Dalasis) | 1803: 591 (438) | 567 (198) | 2003: 659 (528) | 615 (207) | -66 (-147, 14) (-143, 15)* p=0.106 |
| School-related time use of child (proportion) | 1845: 0.683 (0.123) | 0.684 (0.039) | 2062: 0.553 (0.140) | 0.548 (0.062) | 0.130 (0.113, 0.147) p<0.001 |
| Number of hours caregiver spends helping child with homework per week | 1803: 3.08 (4.27) | 3.08 (1.29) | 2003: 2.99 (4.29) | 2.90 (1.26) | 0.09 (-0.34, 0.53) (-0.35, 0.52)* p=0.678 |

*Bootstrap confidence interval, bias corrected and accelerated, based on 2000 bootstrap samples of clusters with stratification by randomized arm.






Note: 15 here is 15 or more

**Table 9. Sibling enrollment in school and time use**

| Variable | Intervention arm | | Control arm | |
|---|---|---|---|---|
| Older sibling: | Individual level N=1742 | Cluster level N=56 mean (SD) | Individual level N=2098 | Cluster level N=55 mean (SD) |
| Enrollment in school | | | | |
| Yes | 1076 (61.8%) | 61.9% (16.1%) | 1255 (59.8%) | 59.1% (16.3%) |
| No | 258 (14.8%) | 14.8% (13.5%) | 247 (11.8%) | 11.3% (11.9%) |
| Missing | 408 (23.4%) | 23.3% (9.8%) | 596 (28.4%) | 29.6% (12.5%) |
| School-related time use: proportion [mean (SD)]* | 1076: 0.583 (0.088) | 0.578 (0.041) | 1255: 0.577 (0.090) | 0.572 (0.033) |
| Younger sibling: | Individual level N=1805 | Cluster level N: mean (SD) | Individual level N=2220 | Cluster level N: mean (SD) |
| Enrollment in school | | 56: | | 55: |
| Yes | 923 (51.1%) | 52.1% (16.1%) | 1031 (46.4%) | 49.4% (22.2%) |
| No | 463 (25.7%) | 25.1% (15.2%) | 546 (24.6%) | 21.2% (17.1%) |
| Missing | 419 (23.2%) | 22.8% (10.4%) | 643 (29.0%) | 29.4% (13.3%) |
| School-related time use: proportion [mean (SD)]* | 923: 0.511 (0.110) | 55:** 0.512 (0.048) | 1031: 0.504 (0.104) | 54:** 0.501 (0.045) |

*Restricted to those reported as being in school
**One cluster in each arm reported no data

**Table 10. Sibling literacy and numeracy**

| Variable | Intervention arm | Control arm | Odds ratio* (95% CI) p-value |
|---|---|---|---|
| Older sibling: | Individual level N=1742 | Individual level N=2098 | |
| Literacy: | | | |
| Cannot read | 301 (17.3%) | 362 (17.3%) | |
| Can read at least five letters, but no words | 507 (29.1%) | 560 (27.7%) | |
| Can read at least five words, but not a sentence | 161 (9.2%) | 220 (10.5%) | 1.05 (0.76, 1.43) p=0.78 |
| Can read a sentence, but not a paragraph | 91 (5.2%) | 89 (4.2%) | |
| Can read entire paragraph | 171 (9.8%) | 161 (7.7%) | |
| Refused | 2 (0.1%) | 1 (0.0%) | |
| Missing | 509 (29.2%) | 705 (33.6%) | |

| | | | |
|---|---|---|---|
| Numeracy: | | | |
| Cannot recognize single digits | 92 (5.3%) | 114 (5.4%) | |
| Can recognize ≥ 4 single digit but not two digit numbers | 229 (13.1%) | 247 (11.8%) | |
| Can recognize ≥ 4 double digit numbers, but can't add | 205 (11.8%) | 253 (12.1%) | 1.02 (0.78, 1.35) p=0.86 |
| Can add single digit numbers, but cannot subtract | 534 (30.7%) | 596 (28.4%) | |
| Can subtract double digit numbers | 171 (9.8%) | 182 (8.7%) | |
| Refused | 2 (0.1%) | 1 (0.0%) | |
| Missing | 509 (29.2%) | 705 (33.6%) | |
| Younger sibling: | Individual level N=1805 | Individual level N=2220 | |
| Literacy: | | | |
| Cannot read | 1093 (60.6%) | 1310 (59.0%) | |
| Can read at least five letters, but no words | 194 (10.7%) | 169 (7.6%) | |
| Can read at least five words, but not a sentence | 11 (0.6%) | 10 (0.5%) | 1.39 (1.03, 1.88) p=0.033 |
| Can read a sentence, but not a paragraph | 4 (0.2%) | 4 (0.2%) | |
| Can read entire paragraph | 4 (0.2%) | 1 (0.0%) | |
| Refused | 9 (0.5%) | 6 (0.3%) | |
| Missing | 490 (27.1%) | 720 (32.4%) | |
| Numeracy: | | | |
| Cannot recognize single digits | 825 (45.7%) | 1007 (45.4%) | |
| Can recognize ≥ 4 single digit but not two digit numbers | 366 (20.3%) | 382 (17.2%) | |
| Can recognize ≥ 4 double digit numbers, but can't add | 51 (2.8%) | 51 (2.3%) | 1.22 (0.97, 1.54) p=0.097 |
| Can add single digit numbers, but cannot subtract | 54 (3.0%) | 49 (2.2%) | |
| Can subtract double digit numbers | 10 (0.6%) | 5 (0.2%) | |
| Refused | 9 (0.5%) | 6 (0.3%) | |
| Missing | 490 (27.1%) | 720 (32.4%) | |

*Odds ratios from an ordered logistic regression model, omitting those in the missing and refused categories.

**Table 11. Caregiver literacy and numeracy**

| Variable | Intervention arm<br>Individual level<br>N=2060 | Control arm<br>Individual level<br>N=2458 | Odds ratio*<br>(95% CI)<br>p-value |
|---|---|---|---|
| Caregiver literacy: | | | |
| Cannot read | 1489 (72.3%) | 1669 (67.9%) | |
| Can read at least five letters, but no words | 216 (10.5%) | 248 (10.1%) | |
| Can read at least five words, but not a sentence | 59 (2.9%) | 58 (2.4%) | 1.02 |
| Can read a sentence, but not a paragraph | 18 (0.9%) | 20 (0.8%) | (0.80, 1.31)<br>p=0.87 |
| Can read entire paragraph | 63 (3.1%) | 65 (2.6%) | |
| Refused | 0 (0.0%) | 2 (0.1%) | |
| Missing | 215 (10.4%) | 396 (16.1%) | |
| Caregiver numeracy: | | | |
| Cannot recognize single digits | 852 (41.4%) | 926 (37.7%) | |
| Can recognize ≥ 4 single digit but not two digit numbers | 465 (22.6%) | 584 (23.8%) | |
| Can recognize ≥ 4 double digit numbers, but can't add | 155 (7.5%) | 166 (6.8%) | 1.01 |
| Can add single digit numbers, but cannot subtract | 282 (13.7%) | 285 (11.6%) | (0.85, 1.20)<br>p=0.94 |
| Can subtract double digit numbers | 91 (4.4%) | 99 (4.0%) | |
| Refused | 0 (0.0%) | 2 (0.1%) | |
| Missing | 215 (10.4%) | 396 (16.1%) | |

*Odds ratios from an ordered logistic regression model, omitting those in the missing and refused categories.

**Table 12. Activity of school mother's club (at cluster level)**

| Variable | Intervention arm<br>N=56<br>mean(SD) | Control arm<br>N=55<br>mean (SD) |
|---|---|---|
| Percentage of children attending a school with a mother's club | 100% (0%) | 100% (0%) |

**Appendix C: Sample size calculation**

Below is the text of the sample size calculation, also given in Boone et al. (2015):

*According to the regional directorates of the MoBSE in the Lower River and North Bank Regions, just under 10,000 grade 1 children are eligible to attend school in the combined North Bank and Lower River regions in the 179 public schools in these regions. The exact numbers of clusters and eligible students will not be known until after the first enumeration is carried out, but based on the above numbers we estimate that there will be around 150 clusters, each including an average of 40 students (6000 students in total). In the STRIPES trial the estimated effect was a 0.75 SD increase in mean score: however, effects of smaller magnitude than this would still be important to detect. Assuming that 60 % of the eligible children will take the test at the end of the trial and an intra-cluster correlation coefficient of 0.23 (as seen in the STRIPES trial), then a trial with 75 intervention villages and 75 control villages will give over 90 % power to detect a difference of 0.3 SD in the standardized score between intervention and control villages using a conventional 2-sided significance level of 5 %. If the treatment effect is of the order of that seen in the STRIPES trial then there will be reasonable statistical power to explore interactions by ethnicity, gender, wealth and geographic location.*

We ended up with fewer villages and children, but also lower attrition (<14%), which yielded more than enough observations to yield sufficient power to meet our original specification.

**Appendix D: Final test papers**

*(begins on next page)*

**SCORE | EGMA The Gambia, May 2018**

# Early Grade Math Assessment in The Gambia: Instructions for Enumerators and Children Response Form

## General Instructions

It is important to establish a playful and relaxed relationship with the child through an initial talk on topics of interest to the child (follow the text in bold below). The child should perceive the assessment more as a game rather than an evaluation. It is important that you ONLY read aloud the text in **bold**, slowly and clearly, so that the child can understand the exercises.

**Good morning. My name is _____. And you, what's your name? I like to _____. And you, what do you like to do? Now that you have done some reading games with my colleague, let's do some Maths game. Throughout this exercise, you can answer in the language that you prefer. Is that ok?** [wait until the child responds] **Are you ready? [**wait until the child responds] **Let's start.**

| Assessment start time: | _____ hh: _____ mm |
|---|---|

| Subtask 1. Number identification | Page 1 | 60 seconds |
|---|---|---|

**In this sheet there are some numbers. When I say "start", start here** [point to the first number]**, and read through the page** [sweep finger across first line]. **Point to each number and read out loud. I will use this timer and will tell you when to stop. Read as fast and the best you can. If there is one number you can't read, move to the next one. Put your finger in the first one [**make sure the child does so and prepare to time]**. Are you ready?** [wait until the child replies] **You can start.**

Start the timer when the child reads the first letter.

When the timer reaches 0, say "stop."

If the child hesitates for 5 seconds, say the number and then point to the next item and say "Go on". Mark the number that you provided as incorrect.

✍ ( / ) = Mark any incorrect number or no response with a slash ( / ).
(ø) Mark with a circle the self-corrections if you already marked as incorrect.
(⌋) = Mark the final number read with a bracket (⌋ ).

| | | | | |
|---|---|---|---|---|
| **2** | **9** | **0** | **12** | **30** |
| **22** | **45** | **39** | **23** | **48** |
| **91** | **33** | **74** | **87** | **65** |
| **108** | **245** | **587** | **731** | **989** |

✍ Time remaining on timer at completion (SECONDS):

| NA1: | NE1: |
|---|---|

✍ Which languages did the child use in this task? (circle all answers that apply)

English Pulaar Mandinka Olof Others (please specify) _____

**Thank you, let's move to the next task.**

| Subtask 2. Number discrimination (PRACTICE) | Page 2 | ✘ |
|---|---|---|

**Look at these numbers. Say which number is bigger** [the child can only be considered correct if he/she "says" the bigger number, pointing is not enough].

**8 4**

✓ [If the child answered 8, say] **Well done, 8 is bigger. Let's try another example.**

✘ [If the child did not answer 8, say] **The bigger number is 8.** [Point to 8] **This is 8.** [Point to 4] **This is 4. 8 is bigger than 4. Let's try another example.**

**Look at these numbers. Say which number is bigger.**

**10 12**

✓ [If the child answered 12, say] **Well done, 12 is bigger. Let's continue.**

✘ [If the child did not answer 12, say] **The bigger number is 12.** [Point to 10] **This is 10.** [Point to 12] **This is 12. 12 is bigger than 10. Let's continue**.

| Subtask 2. Number discrimination (TEST) | Page 3 | ✘ |
|---|---|---|

**Look at these numbers. Say which number is bigger.** [repeat for each item]

(✓) 1 = Correct (✓) 0 = Incorrect or without answer

(ø) Mark with a circle the self-corrections if you already marked as incorrect.

(⌋) = Mark the final answer provided with a bracket (⌋ ).

If the child makes 4 successive errors at any point, say "thank you", discontinue this subtask, mark below and move to the next subtask.

If the child hesitates for 5 seconds, provide the answer and then point to the next item and say "Go on". Mark the item that you provided answer as incorrect.

| | | | | | | |
|---|---|---|---|---|---|---|
| 1 | 7 | 5 | 7 | | 1 | 0 |
| 2 | 11 | 24 | 24 | | 1 | 0 |
| 3 | 47 | 34 | 47 | | 1 | 0 |
| 4 | 58 | 49 | 58 | | 1 | 0 |
| 5 | 65 | 67 | 67 | | 1 | 0 |
| 6 | 94 | 78 | 94 | | 1 | 0 |
| 7 | 146 | 153 | 153 | | 1 | 0 |
| 8 | 287 | 534 | 534 | | 1 | 0 |
| 9 | 623 | 632 | 632 | | 1 | 0 |
| 10 | 867 | 965 | 965 | | 1 | 0 |

Exercise discontinued because the child made 4 successive mistakes. ☐

| NA2: | NE2: |
|---|---|

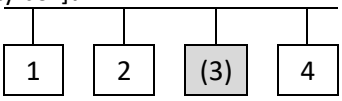
Which languages did the child use in this task? (circle all answers that apply)

English Pulaar Mandinka Olof Others (please specify) _____

**Thank you, let's move to the next task**

| Subtask 3. Missing Number (PRACTICE) | Page 4 | ✘ |
|---|---|---|

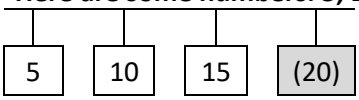
**P1** **Here are some numbers. 1, 2 and 4, what number goes here** [point to the empty box]**?**




✓ [If the child answered 3, say] **Well done, it's 3. Let's do another one.**

✘ [If the child did not answer 3, say] **The number 3 goes here. Say the numbers with me** [point to each number]. **1, 2, 3 and 4. 3 goes here. Let's try another one.**

**P2** **Here are some numbers. 5, 10 and 15, what number goes here?**




✓ [If the child answered 20, say] **Well done, it's 20. Let's continue**

✘ [If the child did not answer 20, say] **The number 20 goes here. Say the numbers with me** [point to each number]. **5, 10, 15 and 20. 20 goes here. Let's continue.**

| Subtask 3. Missing Number (TEST) | Page 5 and 6 | |
|---|---|---|

**Here are some numbers** [point to the box]**. What number goes here?** [repeat for each item]

(✓) 1 = Correct
(✓) 0 = Incorrect or without answer
(ø) Mark with a circle the self-corrections if you already marked as incorrect.
(⌋) = Mark the final answer provided with a bracket (⌋ ).

If the child makes 4 successive errors at any point, say "thank you", discontinue this subtask, mark below and move to the next subtask.

If the child hesitates for 5 seconds, provide the answer and then point to the next item and say "Go on". Mark the item that you provided answer as incorrect

1. 5 | 6 | 7 | (8) — 1 | 0
2. 14 | 15 | (16) | 17 — 1 | 0
3. 20 | (30) | 40 | 50 — 1 | 0
4. (200) | 300 | 400 | 500 — 1 | 0
5. 2 | 4 | 6 | (8) — 1 | 0
6. 348 | 349 | (350) | 351 — 1 | 0
7. 28 | (26) | 24 | 22 — 1 | 0
8. 30 | 35 | (40) | 45 — 1 | 0
9. 550 | 540 | 530 | (520) — 1 | 0
10. 3 | 8 | (13) | 18 — 1 | 0

Exercise discontinued because the child made 4 successive mistakes. ☐

| NA3: | NE3: |
|---|---|

Which languages did the child use in this task? (circle all answers that apply)

English Pulaar Mandinka Olof Others (please specify) _____

**Thank you, let's move to the next task.**

| Subtask 4a. Addition (level 1) | Page 7 and 8 | 60 seconds |
|---|---|---|

Paper and pencil

**In these two pages there are some addition questions** [glide hand from top to bottom on the two pages]. **You should start here** [point to the first problem]. **I will use the timer and will tell you when to start and when to stop. Say the answer for each question. If you don't know an answer, move to the next problem. If you want, you can use this paper and pencil. Are you ready?** [wait until the child responds and prepare to time] **Start.**

(✓) 1 = Correct
(✓) 0 = Incorrect or without answer
(ø) Mark with a circle the self-corrections if you already marked as incorrect.
(⌋) = Mark the final answer provided with a bracket (⌋ ).

| | | | |
|---|---|---|---|
| 1 | 3 + 2 = (5) | 1 | 0 |
| 2 | 1 + 3 = (4) | 1 | 0 |
| 3 | 4 + 5 = (9) | 1 | 0 |
| 4 | 6 + 2 = (8) | 1 | 0 |
| 5 | 8 + 1 = (9) | 1 | 0 |
| 6 | 3 + 3 = (6) | 1 | 0 |
| 7 | 7 + 3 = (10) | 1 | 0 |
| 8 | 3 + 9 = (12) | 1 | 0 |
| 9 | 2 + 8 = (10) | 1 | 0 |
| 10 | 9 + 3 = (12) | 1 | 0 |
| 11 | 7 + 8 = (15) | 1 | 0 |
| 12 | 4 + 7 = (11) | 1 | 0 |
| 13 | 7 + 5 = (12) | 1 | 0 |
| 14 | 8 + 6 = (14) | 1 | 0 |
| 15 | 9 + 8 = (17) | 1 | 0 |
| 16 | 6 + 7 = (13) | 1 | 0 |
| 17 | 8 + 8 = (16) | 1 | 0 |
| 18 | 8 + 5 = (13) | 1 | 0 |
| 19 | 10 + 2 = (12) | 1 | 0 |
| 20 | 8 + 10 = (18) | 1 | 0 |

The child used:

| | |
|---|---|
| | Fingers to count. |
| | Paper and pencil. |
| | Solved the question in his/her head. |

Tick✓ all answers that apply.

Start the timer when you say "start".

When the timer reaches 0, say "stop."

If the child makes 4 successive errors at any point, say "thank you", discontinue this subtask, mark below and move to the next subtask.

If the child hesitates for 5 seconds, provide the answer and then point to the next item and say "Go on". Mark the item that you provided answer as incorrect.

Time remaining on timer at completion (SECONDS):

Exercise discontinued because the child made 4 successive mistakes. ☐

| NA4a: | NE4a: |
|---|---|

Which languages did the child use in this task? (circle all answers that apply)

English Pulaar Mandinka Olof Others (please specify) _____

**Thank you, let's move to the next task.**

| Subtask 4b. Addition (level 2) | Page 9 | |
|---|---|---|

Paper and pencil

**Here are some addition questions** [glide hand from top to bottom]**. Tell me the answer for each question. If you do not know the answer, move to the next one. If you want, you may use this paper and pencil. Are you ready?** [wait until the child responds] **Start here** [point to the first problem]

(✓) 1 = Correct
(✓) 0 = Incorrect or without answer

| | | | |
|---|---|---|---|
| 1 | 13 + 6 = (19) | 1 | 0 |
| 2 | 18 + 7 = (25) | 1 | 0 |
| 3 | 14 + 25 = (39) | 1 | 0 |
| 4 | 22 + 37 = (59) | 1 | 0 |
| 5 | 38 + 26 = (64) | 1 | 0 |
| 6 | 234+512= (746) | 1 | 0 |

The child used:

| | |
|---|---|
| | Fingers to count. |
| | Paper and pencil. |
| | Solved the question in his/her head. |

Tick ✓ all answers that apply.

Skip this subtask if the child scores zero in level 1 Addition questions.

If the child makes 4 successive errors, say "thank you", discontinue this subtask, mark below and move to the next subtask.

➲ I If the child uses an inefficient strategy (e.g. tick marks), ask the child **"Do you know another way to solve the problem?** If "no", move to the next item after 5 seconds.

➲ If the child does not provide answer in 30, point to the next item and say "Go on". You may give additional 30 second if the child is still processing the question.

Exercise discontinued because the child made 4 successive errors. ☐

| NA4b: | NE4b: |
|---|---|

Which languages did the child use in this task? (circle all answers that apply)

English Pulaar Mandinka Olof Others (please specify) _____

**Thank you, let's move to the next task.**

| Subtask 5a. Subtraction (level 1) | Page 10 and 11 | 60 seconds |
|---|---|---|

Paper and pencil

Start the timer when you say "start".

**In these two pages there are some subtraction questions** [glide hand from top to bottom, showing the two pages]**. You should start here** [point to the first problem].**I will use timer and will tell you when to start and when to stop. Say the answer for each question. If you don't know an answer, move to the next question. If you want, you may use this paper and pencil. Are you ready?** [wait until the child responds and prepare to time] **Start.**

When the timer reaches 0, say "stop."

(✓) 1 = Correct (✓) 0 = Incorrect or without answer

(ø) Mark with a circle the self-corrections if you already marked as incorrect.

(⌋) = Mark the final answer provided with a bracket (⌋ ).

If the child makes 4 successive errors, say "thank you", discontinue this subtask, mark below and move to the next subtask.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 5 - 3 = (2) | 1 | 0 | 11 | 15 - 7 = (8) | 1 | 0 |
| 2 | 4 - 1 = (3) | 1 | 0 | 12 | 11 - 4 = (7) | 1 | 0 |
| 3 | 9 - 5 = (4) | 1 | 0 | 13 | 12 - 7 = (5) | 1 | 0 |
| 4 | 8 - 2 = (6) | 1 | 0 | 14 | 14 - 8 = (6) | 1 | 0 |
| 5 | 9 - 8 = (1) | 1 | 0 | 15 | 17 - 9 = (8) | 1 | 0 |
| 6 | 6 - 3 = (3) | 1 | 0 | 16 | 13 - 6 = (7) | 1 | 0 |
| 7 | 10 - 7 = (3) | 1 | 0 | 17 | 16 - 8 = (8) | 1 | 0 |
| 8 | 12 - 3= (9) | 1 | 0 | 18 | 13 - 8 = (5) | 1 | 0 |
| 9 | 10 - 2 = (8) | 1 | 0 | 19 | 12 - 10 = (2) | 1 | 0 |
| 10 | 12 - 9 = (3) | 1 | 0 | 20 | 18 - 8 = (10) | 1 | 0 |

If the child hesitates for 5 seconds, provide the answer and then point to the next item and say "Go on". Mark the item that you provided answer as incorrect

The child used:

| | |
|---|---|
| | Fingers to count. |
| | Paper and pencil. |
| | Solved the questions in his/her head. |

Tick ✓ all answers that apply.

Time remaining on timer at completion (SECONDS)

Exercise discontinued because the child made 4 successive mistakes. ☐

| NA5a: | NE5a: |
|---|---|

Which languages did the child use in this task? (circle all answers that apply)

English Pulaar Mandinka Olof Others (please specify) _____

**Thank you, let's move to the next task.**

| Subtask 5b. Subtraction (level 2) | Page 12 | |
|---|---|---|
| Paper and pencil | | Skip this subtask if the child scores zero in Level 1 subtraction questions. |
| **Here are some subtraction questions** [glide hand from top to bottom]**. Tell me the answer for each subtraction question. If you do not know an answer, move to the next one. If you want to, you may use this paper and pencil. Are you ready?** [wait until the child replies] **Start here** (point to the first problem] | | If the child makes 4 successive errors, say "thank you", discontinue this subtask, mark below and move to next task. |
| (✓) 1 = Correct<br>(✓) 0 = Incorrect or without answer | | If the child uses an inefficient strategy (e.g. tick marks), ask the child **"Do you know another way to solve the problem?** If "no", move to the next item after 5 seconds. |

| | | | |
|---|---|---|---|
| 1 | 19 - 6 = (13) | 1 | 0 |
| 2 | 25 - 7 = (18) | 1 | 0 |
| 3 | 26 - 14 = (12) | 1 | 0 |
| 4 | 59 - 37 = (22) | 1 | 0 |
| 5 | 64 - 26 = (38) | 1 | 0 |
| 6 | 746 - 512= (234) | 1 | 0 |

The child used:

| | |
|---|---|
| | Fingers to count. |
| | Paper and pencil. |
| | Solved the questions in his/her head. |

Tick ✓all answers that apply.

If the child does not provide answer in 30, point to the next item and say "Go on". You may give additional 30 second if the child is still processing the question.

Exercise discontinued because the child made 4 successive mistakes. ☐

| NA5b: | NE5b: |
|---|---|

Which languages did the child use in this task? (circle all answers that apply)

English Pulaar Mandinka Olof Others (please specify) _____

**Thank you, let's move to the next task.**

| Subtask 6. Word problems (PRACTICE) | ✘ | ✘ |
|---|---|---|
| Counters, paper and pencil. | | |
| **I am going to read some problems for you to solve them. If you want you can use these counters, paper and pencil. Listen carefully to each problem. If you need, I can repeat once. Are you ready?** [wait until the child replies] **Let's start.** | | ✘ |
| **There are 3 children in the classroom** [pause and check]<br>**1 child gets out of the classroom.** [pause and check]<br>**How many children stay in the classroom**? | | |

| | |
|---|---|
| ✓ [If the child answers 2, say] **Well done, 2 children stayed in the classroom. Let's continue.**<br>✗ [If the child does not answer 2, Put 3 counters on top of the table and say] **Imagine that these counters are children. One of the children gets out of the classroom. Show me the child getting out of the classroom. How many children stayed in the classroom?**<br>**Well done, two children stayed in the classroom. Let's continue.** | |

| Subtask 6. Word Problems (TEST) | | |
|---|---|---|
| Counters, paper and pencil. | | |
| **Now I will read some more problems for you.** | | |
| (✓) 1 = Correct (✓) 0 = Incorrect or no response | | |
| **1. There is 1 child in the classroom. Another 3 children get inside the classroom. How many children are now in the classroom?** | **(4)** 1 0 | [pause and check] at the end of each sentence to make sure that the child understands what you have said before continuing. You can ask "Do you understand?" when in doubt. If the child requests, you may repeat the question ONCE only. |
| **2. There are 8 balls in the bag. 2 are white and the rest are red. How many red balls are inside the bag?** | **(6)** 1 0 | |
| **3. Demba has 3 oranges. Awa has 6 oranges. How many oranges do I have to give to Demba so that they have the same number of oranges?** | **(3)** 1 0 | |
| **4. There were 8 children in the classroom. Some more children got inside the classroom. Now there are 14 children in the classroom. How many children got inside the classroom?** | **(6)** 1 0 | If the child makes 4 successive errors, say "thank you", discontinue this subtask and mark below. |
| **5. I have 15 bananas to share between 3 children. How many bananas should I give to each child so that all of them get the same number of bananas?** | **(5)** 1 0 | If the child has worked on the problem for more than 60 seconds and not provided an answer, say "let us try another one" and move on to the next item and mark the item as incorrect. |
| **6. There are 6 tables in the classroom. At each table there are 2 children seated. How many children are in the classroom altogether?** | **(12)** 1 0 | |

The child used (Tick all answers that apply):

| | |
|---|---|
| | Fingers to count. |
| | Counter |
| | Paper and pencil. |
| | Solved the problems in his/her head. |

Exercise discontinued because the child made 4 successive errors. ☐

| NA6: | NE6: |
|---|---|

| ✍ Which languages did the child use in this task? (circle all answers that apply) |
|---|
| English Pulaar Mandinka Olof Others (please specify) _____ |

**Thank you, you did a good job. Now please return to your own classroom/you can go home.**

| ✍ Which language(s) did you use to apply this test? (circle all answers that apply) | |
|---|---|
| English Pulaar Mandinka Olof Others (please specify) _____ | |
| Assessment end time: | _____ hh: _____ mm |

| Does the child have any visible/noticeable disability? (circle as appropriate) |
|---|
| No Yes (please specify)_____ |

# Early Grade Reading Assessment in The Gambia: Instructions for Enumerators and Children Response Form

## General Instructions

It is important to establish a playful and relaxed relationship with the child that will be assessed through an initial talk on topics of interest to the child (see example below). Use this time to identify whether the child is comfortable with the national language you use. The child should perceive the assessment more as a game rather than an evaluation. It is important that you do not deviate from the guidelines and ONLY read aloud the text in **bold**, slowly and clearly, so that the child can understand the exercises.

**Good morning/afternoon. My name is _____ and I work at Effective Intervention. And you, what's your name?** [wait until the child responds] **How is your family?** [wait until the child responds] **When I am not at work, I like to _____. And you? What do you most enjoy doing when you are not at school?** [wait until the child responds]

## Verbal Consent

- **Let me tell you why I am here today. I am working with a project of Effective Intervention. We came today to your school to do an exercise to help us better understand how children learn how to read and do mathematics, and you were chosen to help us.**
- **We would like to ask for your help. But you do not have to take part if you do not want to.**
- **We are going to play reading and mathematics games. I am going to ask you to read letters, words and a short story out loud. Then you will go to my friend/colleague sitting at the other side** (point to the direction of the EGMA enumerator**), and he/she will ask you to identify numbers, do some calculations and solve a few problems.**
- **Sometimes I will use this timer to time how long it takes you to complete some of the tasks. If you hear it beeps, please do not pay attention to it.**
- **This is NOT a test and it will not affect your grade at school.**
- **Once we begin, if you would rather not answer a question, that's all right.**
- **Can we start?** [wait until the child responds]

If the oral consent is obtained, please tick: ☐
If the oral consent is not obtained, please make a note on the student list.

| | |
|---|---|
| Assessment start time: | _____ hh: _____ mm |

| Subtask 1. Letter Sound Identification | Page 1 | 60 seconds |
|---|---|---|

**Here is a page with many English letter sounds. Please tell me the SOUNDS of as many letters as you can- not the NAMES of the letters, but the SOUNDS.**

**For example,** [Point to "A"] **this letter sound is /a/.**

**Let's practice. [**Point to "T"] **Tell me what letter sound this is.**
✓ [If the child read /t/, say] **Very good, this letter sound is /t/.**
× [If the child did not read /t/, say] **This letter sound is /t/.**

[Point to "b"]. **Now let's try another one. Tell me what letter sound this is.**
✓ **[**If the child read /b/, say] **Very good, this letter sound is /b/.**
× [If the child did not read /b/, say] **This letter sound is /b/.**

**Have you understood?** [wait until the child replies]

**When I say "start", start here** [point to the first letter]**, and read through the page** [sweep finger across first line]. **I will use this timer and will tell you when to stop. Point to each letter and read out loud the letter sound. Read as fast and the best you can. If there is a letter sound you can't read, move to the next one.**

**Put your finger on the first letter** [make sure the child does so]. **Are you ready?** [wait until the child responds and prepare to time] **You can start.**

Start the timer when the child reads the first letter. Stop the timer when the child reads the last letter.

➲ If the child hesitates for 3 seconds, read that letter and then point to the next letter and say "Continue". Mark the letter you read as incorrect.

When the timer reaches 0, say "stop."

If the child does not provide a single correct response on the first line, say "Thank you!", discontinue this subtask, check the box at the bottom, and go on to the next subtask.

✍ ( / ) Mark any incorrect words with a slash ( / ).
(ø) Mark with a circle the self-corrections if you already marked as incorrect.
(⌋) Mark the final letter read with a bracket (⌋).

Examples: A T b

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
|---|---|---|---|---|---|---|---|---|---|---|
| L | i | h | R | S | y | E | O | w | T | (10) |
| i | e | T | m | G | t | a | d | n | B | (20) |
| h | O | A | E | U | r | L | e | R | u | (30) |
| g | R | e | N | i | r | m | t | s | r | (40) |
| S | T | E | C | p | A | F | c | a | E | (50) |
| y | s | K | A | O | C | O | h | t | P | (60) |
| e | A | e | s | M | F | n | u | R | t | (70) |
| A | y | H | N | S | i | g | m | i | L | (80) |
| b | i | L | O | i | o | E | p | r | x | (90) |
| N | v | c | D | e | d | J | z | O | n | (100) |

✍ Time remaining on timer at completion (SECONDS): _____

✍ Exercise discontinued because the child had no correct answers in the first line. ☐

| NA1: | NE1: |
|---|---|

**Thank you, let's move to the next task.**

Subtask 2: Letter Sound Discrimination

**In this exercise, you will listen to the English words that I read. I will read three words and one of them starts with a different sound. I will read twice. Tell me which one starts with a different sound.**

**For example:**

**"cat", "car", "hot"; "cat", "car", "hot" which one starts with a different sound?**

✓ [If the child answered "hot", say] **Very good, "hot" starts with a different sound.**

✗ [If the child did not answer "hot", say] **"cat", "car", "hot". "hot" starts with a different sound than "cat" and "car".**

**Now let's try again:**
**"light", "count", "learn"; "light", "count", "learn", which one starts with a different sound?**

✓ [If the child answered "count", say] **Very good, "count" starts with a different sound.**

✗ [If the child did not answer "count", say] **"light", "count", "learn". "count" starts with a different sound than "light" and "learn".**

**Did you understand?** [wait until the child responds] **Are you ready?** [wait until the child responds] **Let's start.**

If the child does not provide a correct answer in the first 5 items, say "Thank you!", discontinue this subtask, check the box at the bottom, and go on to the next subtask.

If the child hesitates for 5 seconds, provide the answer. Mark the item that you provided answer as "no response".

(✓) 1 = Correct
(✓) 0 = Incorrect
(✓) . = No answer

| | …… which one starts with a different sound? | | | Correct answer | Correct | Incorrect | No response |
|---|---|---|---|---|---|---|---|
| 1. | **book** | **dog** | **boy** | [dog ] | 1 | 0 | . |
| 2. | **like** | **eat** | **egg** | [like] | 1 | 0 | . |
| 3. | **do** | **get** | **go** | [do] | 1 | 0 | . |
| 4. | **say** | **pay** | **sad** | [pay] | 1 | 0 | . |
| 5. | **apple** | **candle** | **ant** | [candle] | 1 | 0 | . |
| 6. | **sun** | **red** | **run** | [sun] | 1 | 0 | . |
| 7. | **bag** | **ball** | **kick** | [kick] | 1 | 0 | . |
| 8. | **is** | **if** | **of** | [of] | 1 | 0 | . |
| 9. | **from** | **drum** | **drive** | [from] | 1 | 0 | . |
| 10. | **fly** | **good** | **food** | [good] | 1 | 0 | . |

✍ Exercise discontinued because the child had no correct answers in the first 5 items. ☐

NA2:

NE2:

**Thank you, let's move to the next task.**

Subtask 3. Nonword Reading | Page 2 | 60 seconds

**In this sheet there are some made-up words. Read as many words as you can. Do not spell the words, but read them.**

**For example** [Point to the word "ut"], **this made up word is "ut"**.

**Let's practice.** [Point to the word "dif"] **Read this word.**

✓ [If the child answered "dif", say] **Very good, this made up word is "dif"**.
× [If the child did not answer "dif", say] **This made up word is "dif".**

[Point to the word "mab"] **Now let's try another one. Read this word.**

✓ [If the child answered "mab", say] **Very good, this made up word is "mab"**.
× [If the child did not answer "mab", say] **This made up word is "mab".**

**When I say "start", start here** [point to the first word]**, and read through the page** [sweep finger across first line]. **I will use this timer and will tell you when to stop. Point to each word and read out loud. Read as fast and the best you can. If there is one word you can't read, move to the next one. Put your finger on the first word [**make sure the child does so]**. Are you ready?** [wait until the child responds and prepare to time] **Start.**

Start the timer when the child reads the first word. Stop the timer when the child reads the last word.

If the child hesitates for 3 seconds, say the word and then point to the next word and say "Continue". Mark the word that you provided as incorrect.

When the timer reaches 0, say "stop."

If the child does not provide a single correct response in the first line (5 words), say "Thank you!", discontinue this subtask, check the box at the bottom, and go on to the next subtask.

( / ) Mark any incorrect words with a slash ( / ).
(ø) Mark with a circle the self-corrections if you already marked as incorrect.
(⌋) Mark the final word read with a bracket (⌋).

Examples: ut dif mab

| 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|
| ri | loz | yat | zam | tob | (5) |
| zom | hon | mon | jaf | git | (10) |
| bas | af | ked | ig | el | (15) |
| tig | om | dop | pif | ip | (20) |
| fe | ral | mip | kag | vif | (25) |
| lut | sig | zop | zir | naf | (30) |
| riz | yot | wab | lat | jep | (35) |
| wub | dod | ik | vit | nux | (40) |
| pek | zel | bef | wab | hix | (45) |
| wof | ib | mig | zek | vok | (50) |

Time remaining on timer at completion (SECONDS): _____

Exercise discontinued because the child had no correct answers in the first line. ☐

| NA3: | NE3: |
|---|---|

**Thank you, let's move to the next task.**

Subtask 4. Familiar Word Reading | Page 3 | 60 seconds

**In this sheet, there are some English words. Read as many words as you can. Do not spell the words, but read them.**

**For example,** [Point to the word "cat"] **this word is "cat".**

**Let's practice.** [Point to the word "mat"]**. Read this word.**
✓ [If the child answered "mat", say] **Very good, the word is "mat".**
✗ [If the child did not answer "mat", say] **This word is "mat".**

**Now let's try another one.** [Point to the word "top"]
✓ [If the child answered "top", say] **Very good, the word is "top".**
✗ [If the child did not answer "top", say] **This word is "top".**

**When I say "start", start here** [point to the first word]**, and read through the page** [sweep finger across first line]. **I will use this timer and will tell you when to stop. Point to each word and read out loud. Read as fast and the best you can. If there is one word you can't read, move to the next one. Put your finger on the first word [**make sure the child does so]**. Are you ready? [**wait until the child responds and prepare to time] **Start.**

Start the timer when the child reads the first word. Stop the timer when the child reads the last word.

If the child hesitates for 3 seconds, provide the word and then point to the next word and say "Continue". Mark the word that you provided as incorrect.

When the timer reaches 0, say "stop."

If the child does not provide a single correct response on the first line (5 words), say "Thank you!", discontinue this subtask, check the box at the bottom, and go on to the next subtask.

( / ) Mark any incorrect words with a slash ( / ).
(ø) Mark with a circle the self-corrections if you already marked as incorrect.
(⌋) Mark the final word read with a bracket (⌋).

Example: cat mat top

| 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|
| but | time | in | the | also | (5) |
| make | no | its | said | where | (10) |
| came | very | do | after | long | (15) |
| water | run | all | for | paper | (20) |
| her | was | three | been | more | (25) |
| that | must | can | ear | it | (30) |
| jump | words | back | called | work | (35) |
| could | an | him | on | see | (40) |
| that | get | not | zip | what | (45) |
| you | if | their | teacher | when | (50) |

✍ Time remaining on timer at completion (SECONDS): _____

✍ Exercise discontinued because the child had no correct answers in the first line. ☐

NA4: | NE4:

**Thank you, let's move to the next task.**

## Subtask 5a: Passage Reading

Page 5.

Show to the children the page of the stimulus booklet while you read the instructions.

**Here is a short story. I would like that you read this story aloud, quickly but carefully. I will use this timer and will tell you when to begin and when to stop. If there is a word that you cannot read, go to the next one. When you finish, I will ask you some questions about the story. Ready?** [wait until the child responds and prepare to time] **You can start.**

( / ) Mark any incorrect words with a slash ( / ).
(ø) Mark with a circle the self-corrections .
(⌋) Mark the final word read with a bracket (⌋).

| | |
|---|---|
| Ali told his friend Ida to go to uncle Musa's farm. | 11 |
| Ali was hungry and wanted to steal bananas in the farm. | 22 |
| Ida was angry and said: "We cannot do that, to steal is very wrong. | 36 |
| Let's just ask." They found uncle Musa and asked him nicely. He gave them one banana each. | 53 |
| They were glad that they did the right thing. | 62 |

Time remaining on timer at completion (SECONDS): _____

## 60 seconds

Start the timer when the child reads the first word.

If the child hesitates or stops more than 3 seconds on a word, move to the next word and say "Continue".

When the timer reaches 0, say "stop."

If the child does not read any word correctly before the boxed word farm mark below and move to the next task.

If the child says "I don't know", mark incorrect.

Ask the last question even if the child only reads up to word 53.

Exercise discontinued because the child did not read any word correct before the boxed word. ☐

## Subtask 5b: Reading Comprehension

When the child finishes reading, REMOVE the passage from the child's view and ask the first question.

Ask the child only the questions related to the text read. The child should have read the part of the text that corresponds to the question. If a child does not give an answer after 10 seconds, mark "no response" and move to the next question. Do not repeat the questions. Consider all sensible answers the child provides as correct.

**Now I am going to ask you about the story you just read. Answer the questions the best you can.**

| Questions [Answers] | Correct | Incorrect | No response |
|---|---|---|---|
| 1. **Who went with Ali to the farm?** [Ida] | 1 | 0 | . |
| 2. **What did Ali want to do in uncle Musa's farm?** [To steal bananas] | 1 | 0 | . |
| 3. **Why was Ida angry?** [Because to steal is very bad; because Ali wanted to steal] | 1 | 0 | . |
| 4. **How did Ali and Ida get the bananas?** [They asked nicely, they asked uncle Musa, uncle Musa gave to them] | 1 | 0 | . |
| 5. **How would uncle Musa feel if he found out what Ali wanted to do?** [Sad; angry; disappoint] | 1 | 0 | . |

| NA 5a: | NE 5a: | NA 5b: | NE 5b: |
|---|---|---|---|

Which languages did the child use in this task? (circle all answers that apply)
English Pulaar Mandinka Wolof Others (please specify) _____

**Thank you, let's move to the next task.**

| Subtask 6. Listening comprehension | | |
|---|---|---|
| | | Remove the passage from the child's view. Do not allow the child to look at the passage or the questions. If a child says "I don't know", mark as incorrect. |

**I am going to read you a short story aloud ONCE and then ask you some questions. Please listen carefully and answer the questions as best as you can. You can answer the questions in whichever language you prefer. Ready?** [wait until the child responds]

**Demba was very sad when he lost one of his goats. He could not go to look for the goat, because he had to watch the other goats. Demba's grandfather helped and found the goat. Demba was very happy.**

| **Now I am going to ask you some questions related to the story:** | Correct | Incorrect | No response |
|---|---|---|---|
| **Why was Demba sad?** [He lost his goat; he could not go to look for it; he cannot see his goat] | 1 | 0 | . |
| **Who helped to look for the goat?** [Demba's grandfather, his grandfather, grandfather] | 1 | 0 | . |
| **Why was Demba happy?** [Grandfather returned with his goat; his goat is back; Grandfather found the goat, he sees/saw the goat etc] | 1 | 0 | . |

Which languages did the child use in this task? (circle all answers that apply)

English Pulaar Mandinka Wolof Others (please specify) _____

**Thank you for doing this exercise with me**. [Follow the instruction on the enumeration manual]

Which language(s) did you use to apply this test? (circle all answers that apply)

English Pulaar Mandinka Wolof Others (please specify) _____

| Assessment end time: | _____ hh: _____ mm |
|---|---|

Does the child have any visible/noticeable disability? (circle as appropriate)

No Yes (please specify) _____

## Appendix E: Per-protocol Analyses

We pre-specified three measures of adherence to the intervention to allow for the possibility that take-up of the intervention would be uneven and we would need to conduct a per-protocol analysis to estimate a treatment-on-the-treated effect. These are:

1. At village level: the proportion of planned class days on which intervention classes were held.
2. At village level: the mean proportion of eligible children attending classes on days when a class was held.
3. At child level: the proportion of intervention classes attended (including, in the denominator, scheduled classes that did not take place).

For inclusion in each per-protocol analysis, we pre-specified that adherence on the relevant level be at least 75%. In Table D1, we present two pre-specified per-protocol analyses, restricting our intervention sample to only those villages or children who participated in a given number of intervention classes and recalculating our primary outcome. As anticipated, we estimate larger effects when focusing on only villages / children who adhered to the intervention. Given the high levels of adherence overall, however (as shown in Table 4), these results are not dramatically different from the results shown in Table 6 for the entire intervention sample.

**Table E1. EGRA and EGMA test results – composite test score, per protocol analyses**

| Adherence measure used | (1) Intervention | (2) Control | (3) Adjusted difference [95% CI] | (4) P-value |
|---|---|---|---|---|
| Only intervention villages with at least 75% child attendance | 65.2 (21.3) | 17.1 (14.2) | 47.9 [45.0, 50.9] | p<0.0001 |
| Difference in SD units: | -- | -- | 3.37 | -- |

| | | | | |
|---|---|---|---|---|
| | | | [3.00, 3.76]* | |
| Observations | 1147 | 2071 | -- | -- |
| Only intervention children with at least 75% attendance | 67.4 (18.9) | 17.1 (14.2) | 50.2 [47.7, 52.7] | p<0.0001 |
| Difference in SD units: | -- | -- | 3.52 [3.19, 3.93]* | -- |
| Observations | 1525 | 2071 | -- | -- |

Note: Column (1) shows the mean test scores for the intervention group, restricting the intervention sample to two separate, pre-specified minimum adherence measures, respectively. Column (2) shows the same control group means as in Table 6. In column (3), we show the difference between column (1) and (2) adjusted for the randomization stratification factors with a 95% confidence interval (that takes into account the clustered design) given in brackets below. In column (4) we present the p-value from the corresponding hypothesis test. *Bootstrap confidence interval, bias corrected and accelerated, based on 2000 bootstrap samples of clusters with stratification by intervention status.