EDUCATION POLICY AND DATA CENTER

Making sense of data to improve education for development

Effectiveness of Read-Aloud Instruction on Reading and Math Outcomes

EVIDENCE FROM NORTHERN NIGERIA

EPDC Research Paper No. 18-01

Wael Moussa, Emily Koester, and Olusola Alonge

April 2018

Education Policy and Data Center, FHI 360 | 1825 Connecticut Avenue NW, Washington, D.C., USA

Effectiveness of Read-Aloud Instruction on Reading and Math Outcomes: Evidence from Northern Nigeria

Wael Moussa[†,‡], Emily Koester[‡], and Olusola Alonge[*]

Abstract

This paper evaluates the net causal effects of Read-Aloud (RA) story lessons on reading and math learning outcomes in 199 second grade classrooms in northern Nigeria. We employ a cluster-randomized controlled trial to assess the efficacy of two types of RAs. The first type uses local language stories, accompanied by vocabulary instruction and comprehension questions. The second follows the same structure, but is math-themed and includes a set of math exercises. The treatments are delivered to students in the treatment schools for 35 minutes weekly over the duration of the school year. The findings of this paper suggest that regular RAs are effective in improving reading outcomes, whereas math-themed RAs improve math-related outcomes. We estimate that the regular RA lessons increase oral reading fluency by 3.12 correct words per minute and improves the literacy rate by 11 percent. Math RAs improve listening comprehension, missing number identification, and math word problem scores by 8.8, 18, and 21 percent, respectively.

Keywords: Read-Aloud, Reading, Mathematics, Early Grade

[†] Corresponding author. e: wmoussa@fhi360.org; t: 202-884-8206
[‡] Research and Evaluation, Education Policy and Data Center, FHI 360, 1825 Connecticut Avenue NW, Washington, DC 20009.
[*] FHI 360-AHNI, Plot 1073 J. S. Tarka Street, Garki, Area 3, P.M.B. 44, Abuja, Nigeria.

1 Introduction

Literacy and numeracy skills are essential for education attainment in the short-term and human capital accumulation in the long-term (Hanushek et al., 2013; Montenegro and Patrinos, 2014). Much of the developing world, however, is facing a substantial deficit in reading and math skills. Across at least 39 reading interventions funded by the United States Agency for International Development (USAID), almost half of all children in these programs cannot read a single word; that number reaches about 72 percent in Sub-Saharan African countries (Gove et al., 2015; USAID, 2017). Developing countries also face a substantial deficit in basic mathematics. Recent administrations of the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) and the Programme for the Analysis of Education Systems (PASEC), show that about half of the children in early primary have less than a basic level of competency in mathematics, and 16 percent of those children do not display any competency at all.[1] Although international donors have recently invested heavily in early grade reading programs, these have been criticized for their narrow focus on phonics over other types of reading skills, such as comprehension and concepts of print (Bartlett and Frazier, 2015). Many of these international early grade programs neglect subjects outside of literacy, such as science and mathematics, preferring to prioritize their funding for reading.

One promising strategy for addressing learning challenges in this environment is reading aloud to young children. In many parts of the world, Read-Aloud (RA) books are widely recommended for home and school settings (Campbell, 2001). RA lessons are promoted for developing students' emergent reading skills and motivating students through reading enjoyment (Fisher et al., 2004; Trelease, 2013).[2] RAs have been shown to positively impact children's oral language skills and concepts of print (National Early Literacy Panel, 2010; Swanson, Wanzek, Petscher, Vaughn, Heckert, Cavanaguh, Kraft, & Tacket, 2011) and numerous studies have shown that RAs can enhance students' vocabulary (Biemiller & Boote, 2006, Brett, Rothlein, & Hurley, 1996; Penno, Wilkinson, & Moore, 2002; Senechal, 1997; American Academy of Pediatrics, 2017). More limited research indicates that RAs positively impact listening comprehension outcomes (Meyer, Wardrop, Stahl, & Linn, 1994).

The bulk of the research literature on RAs has focused on literacy and language outcomes. However, there is increasing interest in the potential of storybooks to improve outcomes in other subjects, including mathematics (Van den Heuvel-Panhuizen et al., 2016). Children's literature can be used to build context for mathematics, make connections

[1] The PASEC assessment defines basic/sufficient mathematics competency as being able to recognize numbers up to 100, compare them, complete logical series and perform operations with numbers under 50. Students with basic competencies are also able to solve basic problems with numbers under 20 using reasoning skills. For more information on the PASEC assessment visit: http://www.pasec.confemen.org/wp-content/uploads/2015/12/Rapport_Pasec2014_GB_webv2.pdf

[2] Note that for this paper, we will use the terms RA lessons and RAs interchangeably.

between mathematics and children's lives, generate interest around mathematics, and promote the use of mathematical processes like problem-solving, reasoning, and inquiry (Van den Heuvel-Panuizen and Elia, 2012). A few small studies have indicated that math RAs, implemented in conjunction with related instructional activities, positively affect math interest and basic math skills (Hong, 1996; Jennings, Jennings, Richey, & Dixon-Krauss, 1992; Young-Loveridge, 2004). In a recent study of 384 kindergartners, Van den Heuvel-Panhuizen, Elia and Robitzsch (2016) found that math RAs implemented twice weekly for three months resulted in 22% gains in basic math skills over the control group.

The research on the effects of RAs on listening comprehension and math outcomes is sparse. The small number of studies that investigated math-themed RAs have found positive effects on math interest and basic math skills.[3] The studies described above examined the impacts of regular and math-themed RAs on either literacy outcomes or math outcomes, exclusively. In addition, these studies suffer from small sample sizes. Moreover, all studies conducted around RAs include between 20 and 384 preschool or kindergarten students in developed countries, the majority of whom include fewer than 100 participants (Swanson et al., 2011). The inferences made in these studies may not all be construed as causal nor do they necessarily generalize to different socioeconomic and geographic contexts.

The current study addresses important gaps in the research on RAs. First, this study takes place in northern Nigeria, thus expanding the evidence base around classroom-based programs in primary schools from similarly low-income regions. Second, we compare two interventions, a standard language RA intervention and a math RA intervention, simultaneously, against the same reference group.[4] None of the existing research have examined language RAs and math RAs within the same context. As such, this paper creates an improved evaluation design of RAs using a multi-treatment cluster-randomized controlled trial (CRT) design. The study also examines how language and math RAs affect both math and reading outcomes, which we identify as a gap in the research literature. Lastly, this study employs a cluster-randomized controlled trial with a sample size of 2,244 second grade children from 199 schools. This large sample size enables us to disaggregate results by gender, socioeconomic status, and school type.[5] The latter grouping is relevant directly to the Northern Nigerian context where the program operates, where about a third of primary age children attend a Quranic school.

We evaluate the effectiveness of RA lessons on reading and math outcomes in 199 second grade classrooms in Northern Nigeria, provided as part of a larger early primary

[3] Hong (1996), Jennings et al. (1992), Young-Loveridge (2004), Van den Heuvel-Panhuizen et al. (2016).
[4] We define standard language RAs as read aloud lessons that are not math-themed, and will be referred to as language RAs from this point forward.
[5] There are two main types of schools in Northern Nigeria, public schools and integrated Qur'anic schools (IQS). The latter school type integrates the national curriculum with Islamic religious study.

intervention, Reading and Numeracy Activity (RANA). RANA is an educational intervention implemented by FHI 360, funded by UNICEF and DfID as part of the Girls' Education Program Phase 3 (GEP3), to boost literacy and numeracy skills among first through third grade students. As such, the focus of this paper is the RA lessons, and RANA is the context within which this evaluation takes place. This means that all 199 RANA schools are implementing the RANA curriculum and receiving equal support from the intervention for teacher professional development and teacher learning circles. However, in second grade, half of all schools are randomly assigned to either the language RA or math RA group, and the remaining half are assigned to the control group receiving no RA lessons whatsoever.

RA lessons are thus provided by teachers in two modes. The first delivers language RA lessons once per week, for 35 minutes, using regular children's stories written in the local language, Hausa, and created with the Northern Nigerian context in mind. This mode is the more common method of RA delivery in early grade settings where teachers read a story to the students and then engage them in a discussion with questions and answers (Swanson et al., 2011). The second mode follows the same structure as the language RA lessons but are math-themed followed by math exercises, also delivered in the local language, which we refer to as the math RAs. The treatments, language RAs and math RAs, are thus randomly assigned to 50 schools each, and the remaining 100 schools are assigned to the control group.

We find that both the language and math RAs are successful in increasing student literacy and numeracy outcomes. We provide evidence that regular RAs are most effective in improving reading outcomes, while math RAs are most effective in improving listening comprehension and math outcomes. Specifically, we estimate that the language RAs lead to a 3.12 correct word per minute (cwpm) increase in oral reading fluency (ORF) scores and a 7 percentage point decline in the proportion of non-readers. The math RAs lead to about a 11.8 percentage point gain in word problem solving, an 5.8 percentage point gain in missing number identification, and almost a 6.9 percentage point increase in listening comprehension scores. The RA treatments are also effective in closing the achievement gap between boys and girls in reading, but not in terms of comprehension and numeracy outcomes. When disaggregating the results by SES, we find that RAs, although effective across SES groups, yield higher returns among high SES students, which is in line with existing international research evaluating program/intervention impacts.[6] Finally, our results show that RAs had the larger impact on reading and math learning outcomes among students in public schools, while RAs were more successful among Integrated Qur'anic Schools students in terms of listening comprehension performance.

[6] Glewwe et al. (2013) provide a summary of recent evaluations of international education interventions showing that these interventions can have dis-equalizing effects between SES groups.

2 Context and Experiment Design

2.1 The Reading and Numeracy Activity Program

This study takes place as part of the Reading and Numeracy Activity (RANA), a three-year literacy and numeracy program, funded by the UK Department for International Development (DfID) through the United Nations Children's Fund (UNICEF), and implemented by FHI 360. The program operates in 199 primary-level schools in the northwestern Nigerian states of Katsina and Zamfara. Approximately 60 percent (120 schools) of RANA schools are public schools, while about 40 percent (79 schools) are Integrated Quranic Schools (IQS). IQS are traditional Islamic learning centers that have recently expanded to include elements of the core national curriculum (Solomon, 2015). RANA aims to improve student literacy and numeracy outcomes in grades 1, 2, and 3 and reading and math instructional practices in those same grades. To achieve these aims, RANA developed a set of student books, teacher guides, and in-class RA lessons for grades 1, 2 and 3. RANA provides teacher training, monitoring, and support to help teachers properly use these materials in class. All materials and training activities are conducted in the local language, Hausa.

The RANA program began implementation midway through the 2015-16 academic year and followed a phased roll-out schedule of the intervention as follows. In its first year, RANA implemented 3 phonics lessons per week, which included brief RAs (roughly five minutes long) in grades 1 through 3. In Year 2, RANA implemented a more robust RA program. This revised program eliminated the 5-minute RA, and added a fourth lesson that was devoted entirely to RAs. As such, this paper aims to estimate the net effect of the RANA RA intervention and not necessarily identify the learning mechanism by which RAs affect children's cognitive ability. However, in grade 2, as we will detail in the following section, the RANA program divides 100 classrooms, randomly and equally, between math-themed RAs and standard language RAs, while the 100 are assigned to the control group, where the control group receives no RA lessons. This is to evaluate the effectiveness of RAs and to ascertain whether math RAs provide any value-added over and above those from regular RAs.[7]

2.2 RA Lessons

RANA RA stories are developed with special attention to the local and cultural contexts of the teaching curriculum among RANA schools. In a series of writing workshops with local authors, trainers, educators, math specialists, and RANA project staff, the project developed 55 different math RA stories for Primary 1, 2 and 3. Each grade features 24 stories (with some overlap in stories across grades), or 8 stories per term, each written in

---

[7] All RANA classrooms feature numeracy themed RA stories with the exception of grade 2 classrooms, where we conduct the experiment analyzed in this paper.

Hausa. RANA-developed RA stories are based on research advocating for explicit vocabulary teaching and sufficient discussion (Biemiller and Boote, 2006; Dickinson and Smith, 1994). Math RA stories have typically been accompanied by corresponding exercises (Van den Heuvel-Panhuizen et al., 2016). As such, each RANA-developed RAs are accompanied by a picture, two highlighted vocabulary words, three listening comprehension questions, and a set of math exercises that correspond to the story theme.

These stories are designed to be engaging works of literature that motivate students to listen and learn. [8] All math themes and exercises align with the Nigerian national mathematics curriculum, thus complementing what pupils are learning in their regularly scheduled math classes. For example, during the week that teachers are expected to teach double digit subtraction in their regular math classes, Hausa teachers are expected to read a RANA story aloud on double digit subtraction. Given the unregulated school contexts of northern Nigeria, however, it is possible that the math curriculum is not followed as designed in most schools.

All RA stories are compiled by grade level and ordered by curricular topic. Each story has one illustration that is large enough to be seen in a crowded classroom. Teachers are expected to teach one RA class per week (totaling 24 RA classes per year, 8 per term). Each RA class is expected to last approximately 35 minutes. During the RA class, teachers are expected to follow the following seven steps:

1) Introduce the title and show the cover picture
2) Ask pupils to predict what the story will be about
3) Introduce two new vocabulary words in the story. Pupils are expected to raise up their thumb when they heard the word in the story.
4) Read the story aloud, using good expression. Check to make sure pupils raise their thumb when each vocabulary word is RA.
5) Ask the three comprehension questions listed at the end of the story
6) Lead the class in the math exercises listed at the end of each story
7) Ask pupils what they learned that day.

RANA provides teachers with three days of training (one day per term) to help teachers become comfortable with these steps. During each training, RANA facilitators model RA lessons, facilitate practice in reading with expression, connect reading aloud with oral storytelling traditions, and provide practice time for leading math exercises and full lessons.

The following grade 2 math topics, taken from the Nigerian curriculum, are included in RANA's RA stories:

- Counting up to 1,000

[8] Refer to the Appendix for examples of regular and math RA stories used among grade 2 RANA classrooms.

- Place value
- Odd and even numbers
- Counting by 5s and 10s
- Fractions
- Addition and subtraction of two digit and three digit numbers
- Multiplication
- Recognition of money
- Length, area, capacity, weight
- Time and days of the week
- Two and three-dimensional shapes

3 Experiment Design and Data

The main objective of this paper is to identify the effects of RAs, and isolate these effects from the RANA program effects and other confounders, which all 199 schools are a part of, on early grade literacy, numeracy, and comprehension outcomes. Specifically, we estimate the causal impact of RA lessons on the probability of being a non-reader, oral reading fluency, listening comprehension, missing number identification, and math word problem solving. We implement a cluster-randomized experimental research design with multiple treatments. The treatment is the delivery of RA lessons once per week, over the course of an academic year, and is divided into two treatment arms. The first relies only on regular RA stories and the second uses math-themed RA stories, both of which are presented in the local language, Hausa. We follow a cluster-randomized approach as the RANA program is implemented in clusters made up of 4-6 schools that are within 5 km of each other—the intended purpose of school clustering within the RANA context is to facilitate teacher professional development and support activities and foster shared knowledge, teaching practices, and accountability spillovers within each cluster. The cluster-randomization ensures that any spillover gains from the school clusters are the same across all RANA schools, and are thus differenced out in our empirical strategy.

To ensure that the language and math RA treatment effects are not influenced by these spillovers, all schools in each cluster will either be implementing one of the treatments or none, i.e. there is no within-cluster variation in the treatment assignment. We also note that randomization was only possible in grade 2 due to project and budgetary constraints, and as such, the focus of this paper will be on grade 2. Even though the second grade curriculum has overlapping elements with other early primary grades, the findings of this research are primarily applicable to grade 2. To test whether different RA types produce different results, we randomly assign 25 percent of all schools to deliver language RAs, another 25 percent to deliver math RAs, and the remaining half are assigned to the control group where no RA stories were provided. We further supplement the multi-treatment experimental design, by stratifying the impact estimates by student gender, socioeconomic status, and school type to test for heterogeneous treatment effects.

Although randomization of the treatments will, asymptotically, balance the treatment and control groups across observable and unobservable characteristics of the schools and students, it is possible to achieve imbalance between the treatment groups from a single draw. We formally test for balance in the first condition and adjust for any deviations to ensure the robustness of our impact estimates via standard matching techniques.[9] Matching in a cluster-randomized controlled trial such as this ensures sample balance, but also minimizes bias in the estimated treatment effects and leads to more conservative tests of inference (Campbell et al., 2007; Donner et al., 2007). In the context of our identification strategy, to provide causal inference on the RA effects, we also test for pre-treatment differences in test scores to ensure that the randomization was successful.

We use student-level data collected by the RANA project over the duration of the 2016-17 academic year from grade 2 classrooms in 191 schools (out of 199) implementing the RANA program. The original sampling framework for the RANA program monitoring and evaluation data collection includes a random sample of about 10 students from one grade 2 classroom in each of the RANA schools at the beginning of the school year, and again at the end of the year.[10] The dataset itself contains rich information on school characteristics, classroom lesson observations, teacher interviews, and early grade reading and mathematics assessments (EGR/MA).[11] In addition to student reading and math test results, the EGR/MA data include data on individual students' socioeconomic background, home literacy environment, availability of learning materials, and background demographic information.

To complete the analysis, we restrict the analytic dataset to those student observations with complete data across both terms, and across all demographic, school, classroom, and teacher characteristics. Here, each student in the analytic dataset is tracked longitudinally and observed twice, once at the beginning of the year (at the end of term 1) and again at the end of the academic year (end of term 3). This restriction yields a final analytic sample of 2,244 grade 2 student-term observations (1,122 unique students) across 50 schools providing language RAs, 49 providing math RAs, and 92 control schools. Specifically, 1,099 students are randomly assigned to the control group, 581 to the language RA treatment group, and 564 to the math RA treatment group.

[INSERT FIGURE 1 HERE]

[9] To ensure balance in the observable characteristics between the treatment and control groups, we employ propensity score matching with multiple treatments using multinomial logistic regression to determine propensity scores.

[10] Almost all schools implementing RANA had a single classroom per grade, as such, all students that were sampled are from the same class in each school.

[11] EGR/MA tests are early grade literacy and numeracy assessments designed by RTI International. For more information see Gove and Wetterberg (2011).

Figure 1, plots the overall student composition within RANA schools in terms of gender, socioeconomic status (SES), school type (public and integrated Quranic), and whether the student lives with their teacher.[12] The descriptive analysis here shows that our sample of student is fairly representative of the larger student population of Northern Nigeria, where the majority of student enrollment is composed of boys, the majority of students are considered low SES, and about two-thirds of students attend a government funded school (NEMIS, 2016). Figure 2 displays the student test score distributions, aggregated to the school level, by subtest and treatment status. Across the treatment groups, we can see that average test scores are similarly distributed with the treatment groups showing slightly higher peaks closer to the right tails of each distribution. Meaning that, in the treatment period, treatment schools were somewhat outperforming the control schools.

[INSERT FIGURE 2 HERE]

4 Results

4.1 Experiment Validation: Pre-Treatment Differences

Figure 3 plots the percentage of students who are identified as non-readers, average oral reading fluency scores, and average listening comprehension scores from grade 2 students attending schools that were later assigned to the control, language RA, and math RA groups the following year. It is important to note that data collected in the first year of the project only administered EGRA without the mathematics component. As such, we are unable to test for pre-treatment differences in math learning outcomes.

The first panel of Figure 3 shows that prior to the treatment, students attending the control schools were averaging about 10.1 words per minute, while those in the language RA and math RA groups averaged 10.6 and 9.5 correct words per minute, respectively. Simple t-tests show that the differences between the treatment groups and the control group means are statistically insignificant. The second panel describes a similar test conducted on pre-treatment non-reader rates among second grade students. The figure shows that prior to the introduction of the 35-minute language and math RAs, schools that were eventually assigned to the treatment had a higher rate of non-readers relative to the control group. Although only the mean difference between the math RAs and the control group is statistically significant. This means, that the randomization was successful at least in not favoring the treatment groups where both RA groups are slightly lower achieving, in terms of percent of readers, at baseline.

[12] Student SES is determined using a principal components analysis across 7 survey items that determine household asset ownership. Students that live with their teacher is an indicator that the student does not live with their biological parents, or does not have a legal guardian—in certain cases, this can be used as a proxy for orphanhood, however, it is not generalizable.

We also test for baseline differences in listening comprehension which is an EGRA subtest measured as the percentage of questions answered correctly. We find that, in the pre-treatment period, students in the control group and math RA group have a statistically similar performance on listening comprehension, but students in language RAs schools have a 5-percentage point lower score, on average. Although this difference is statistically significant, this could bias our treatment effects downward, meaning that at worst, our impact estimates on comprehension outcomes would serve as a lower bound.

[INSERT FIGURE 3 HERE]

Finally, we test for sample balance between the treatment groups and the control group using simple t-tests with standard errors clustered at the school level to account for the original sampling framework for the RANA project. Table 1 presents the mean student, teacher, and classroom characteristics by treatment status without making any adjustments to the samples. In this case, hypothesis testing is carried out in pairwise combinations that compare each treatment group to the control group separately. Comparing the regular RAs group to the control group, we find that students are statistically different in terms of the likelihood of teacher gender encouragement during in-class lesson delivery, by 6 percentage points.[13] In terms of the math RA group, we find that students differ from the control group along home literacy environment, wealth, teacher educational attainment, and class attendance rate. We note that the samples are somewhat imbalanced despite the randomization of the treatment assignments. However, because the schools were assigned to the treatments strictly using a random process, i.e. all schools had an equal probability of being selected to the treatment prior to selection, these differences are a function of the data generating process and not some deterministic function.

[INSERT TABLE 1 HERE]

Although the differences in the observed characteristics between the treatment and control groups are due to the randomization, these differences can still influence estimates of the language RA and math RA effects. This is because the differences between groups are also correlated with the outcome variables, and would therefore bias the impact estimates. To ensure that the impact estimates are not influenced by observable confounders, we balance the samples by employing propensity score matching using Kernel density weights. This technique reweights observations ensuring that student, teacher, and classroom characteristics are statistically indistinguishable between the treatment and control groups. Specifically, we employ a multinomial logit model to

[13] Teacher encouragement data are collected via lesson observations carried out by RANA staff to assess their fidelity of program implementation. However, lesson observations were not carried out for any RAS or math RA activities.

determine probabilities of belonging to either treatment group and calculating inverse probability weights to match the samples as follows.

$$w_i = \sum_j \frac{T_i^j}{p_i^j}; \text{ where } j = c, t_1, t_2 \qquad [1]$$

Here, w_i corresponds to the inverse probability weight that is calculated as the inverse of the probability of being assigned to either the control, language RA, or math RA groups for student i in the analytic sample. T^j thus, takes on a value of 1 if student i belongs to group j, which could be either the control, language RA, or math RA group, and takes on a value of zero otherwise. As such the treatment effects are estimated as follows.

$$\hat{\delta}^j = \frac{\sum_i \sum_j w_i T_i^j Y_i(X)}{\sum_i \sum_j w_i T_i^j} \qquad [2]$$

where, Y_i denotes the literacy or numeracy outcome of interest and X is a matrix of control variables that include student, teacher, and classroom observed characteristics. Equation (2) follows from the following propensity score weighted regression.

$$Y_i = \hat{\delta}^j \sum_j T_i^j + X_i \beta + \varepsilon_i \qquad [3]$$

Table 2 tests for sample balance after applying inverse probability weighting using the propensity scores produced by the multinomial logit. In this case, we find that only students receiving the language RA differ from the control group in terms of their age, again almost by one full year. However, all other differences between the three groups are no longer statistically significant following the reweighting scheme. We can thus conclude that the samples are now balanced and have minimized influence from confounders in estimating the impacts of RA stories.

[INSERT TABLE 2 HERE]

4.2 Read Aloud Effects

In the following analyses, we present point estimates of $\hat{\delta}$ for the language and math RA lessons, separately. $\hat{\delta}$ represents the effect of the language or math RAs on the probability of being a non-reader, oral reading fluency, listening comprehension, missing number identification, and math word problems. The results are then stratified by student subgroups: gender, school type, and SES.

4.2.1 Reading Outcomes

Table 3 displays the average effects of language RAs and math RAs on the probability that a student is a non-reader. Here, we define non-readers as students who took the oral reading fluency test and received a score of zero. Table 4 displays the point estimates of the conditional differences in oral reading fluency performance (measured as correct words per minute) between students in the language RA, math RA, and the control groups.[14]

[INSERT TABLE 3 HERE]

Overall, we find that language RA lessons are successful in lowering the percentage of non-readers by 7 percentage points. Relative to the counterfactual non-reader rate, this is equivalent to increasing the percentage of students who can read at least one word by almost 11 percent. We also find that the language RA lessons have the largest effects on girls, low SES students, and public school students.

The percentage of girls who are non-readers is lowered by 9.5 percentage points, from 71.1 percent down to 61.6 percent. Although the point estimate is only marginally insignificant at 12 percent level, the magnitude is meaningful. Relative to the control group, girls exposed to the language RAs are 13.3 percent less likely to be identified as non-readers. Boys experience a similar effect lowering the percentage of non-readers, but to a lesser extent. Language RAs lower the percentage of non-readers among boys by only 3.7 percentage points, and the estimate is not statistically significant. The combination of these findings has implications on the gender reading gap. In the absence of the treatment, the gender literacy gap stands at 10 percent in favor of boys, i.e. there are 10 percent fewer boys who cannot read than girls. The language RAs narrow the gap from 10 percent to 4.6 percent. The math RAs do not show any statistically significant changes in the probability of being a non-reader among boys or girls.

Across SES student groups, the language RAs lower the percentage of non-readers only among the low SES group, by 9.3 percentage points. The high SES group, however, effectively experience a net zero effect from the language RAs. From an education equity perspective, this means that the preexisting reading gap of 9.8 percent in favor of high SES students is lowered to 0.3 percent, essentially eliminating the SES achievement gap in reading. On the other hand, our estimates of the math RAs on both the low and high SES groups are positive and not statistically significant.

When stratifying the analysis by school type, we find that students in public schools benefitted the most from the regular RA lessons. The percentage of non-readers among public school students is lowered by 10 percentage points in the language RA group, lowering the percentage of non-readers from 75.6 percent to 65.6 percent, which is equivalent to approximately a 14 percent decline in the probability that a student cannot read when exposed to reading instruction with weekly standard language RA stories. The same effect in IQS is only a 4.4 percentage point decline in the probability of non-reading, and is not statistically significant. On the other hand, the results show that the math RA

[14] For simplicity, we refer to reading fluency as measured by the EGRA oral reading fluency test that measures the number of correct words in a grade appropriate passage read per 60 seconds.

lessons do not influence the probability that a student can read at least one word in a statistically significant manner, overall or by student strata.

[INSERT TABLE 4 HERE]

The average treatment effect of the language RA lessons adds about 3.12 cwpm to students' reading fluency over the mean reading fluency where students are not exposed to any RA lessons. Relative to the 18.1 cwpm standard deviation for the ORF score distribution, this effect translates to an effect size of .17 SD. Girls exhibited the highest response to the language RA lessons, with an increase of 5.41 cwpm on the oral reading fluency test. The language RAs more than doubled their level of reading fluency had they not been exposed to any RA lessons, increasing it to little over 10 cwpm. Equivalently, the ORF boost experienced by girls translates to an effect size of .30 SD. The effect on boys, on the other hand, is modest with an increase of 1.41 cwpm that is also statistically insignificant, which is an effect size of .08 SD. This translates to the gender gap being closed by the language RA lessons from a state where the gender gap was 3.04 cwpm in favor of boys to about 1 cwpm in favor of girls, although this difference is not statistically significant.

Table 4 also presents the effects of the math RA lessons, where we find that the only statistically significant effects are among girls and public school students. Moreover, we estimate that girls' ORF benefitted from the math RAs with a 3.97 cwpm increase in their reading fluency, an effect size of .22 SD, whereas boys only saw a 0.63 cwpm increase from the same RA lessons, a statistically insignificant effect size of .03 SD. Like the language RAs, the gender reading fluency gap between boys and girls is effectively closed by the math RAs, where girls' average ORF score following the treatment is 8.59 cwpm and 8.28 cwpm for boys. When examining the effects of the language and math RAs on IQS and public school students, the results show that they positively impact public school students only, while all estimated effects on IQS students are close to zero and statistically insignificant. Public school students exposed to language or math RAs similarly experience about 4.7 cwpm increase in their reading fluency scores, which is equivalent to a .26 SD effect size. Lastly, the results show that both the high and low SES groups benefit from the language RA almost equally showing improvements in ORF by 4.4-4.6 cwpm, which is an effect size of .24-.25SD.

These findings shed light on the use of a RA lesson structure to positively influence reading outcomes, both at the margin of being able to read at all and in enhancing fluency. Overall and across most student subgroups, language RA lessons are more impactful in improving reading outcomes than the math RAs, as these are generally more focused on improving literacy and the alternative are focused on numeracy. The exact mechanism behind these findings is not clear, as there may not be a direct link from exposure to RAs that would directly translate into improved reading skills. However, the results show that

RAs may have an indirect link to reading and fluency, whereby children exposed to RAs improve upon an intermediary skill that consequently enhances reading skill acquisition.

4.2.2 Listening Comprehension

The effects of language and math RA lessons on listening comprehension are presented in Table 5. The listening comprehension test is a list of 5 questions following a short passage that test takers listen to, and is scored as the percentage of questions answered correctly. We find that math RA lessons, more so than the language RAs, are successful in boosting listening comprehension across most students and subgroups.

[INSERT TABLE 5 HERE]

Students in the math RA group outperform their control group counterparts by 6.9 points (out of 100), about 9 percent relative to the control group, while the Language RA group outperform the control group by 2.1 points, which is statistically insignificant. When stratifying the RA effects by gender, we find that girls showed larger improvements in comprehension relative to boys in the language RA group, however, we find that girls do not benefit from the math RA lessons. The effect on girls is 5.6 points and virtually zero for boys; however, both point estimates are not significant at the conventional levels. We find that the math RA lessons have a significant and positive effect on boys' listening comprehension scores, however. Boys outperform the control group by 9.1 points, which is equivalent to approximately a 12 percent increase relative to the control group, whereas the girls show no effects from the same RA lessons.

Table 5 shows that the language and math RAs are both equally successful in improving listening comprehension among the low SES group, with estimated effects of 6.7 and 7.6 points relative to the control group. This translates to an increase in listening comprehension performance by between 8.9 and 9.7 percent relative to the control group mean of 75.2 percent. Among high SES students, we find that only the math RAs increased listening comprehension scores by 16.3 points, whereas the effect of language RAs was a 2.6 point decline, but is statistically insignificant. In the absence of any RA treatment, the listening comprehension gap between SES groups was minimal. However, with the introduction of language RAs, the low SES students are now outperforming their high SES counterparts. Upon introduction of the math RAs, we find that although both groups improve, the high SES group still outperforms the low SES group by a statistically significant margin of 9 points.

Among IQS, the results show that both language and math RA lessons are successful in improving listening comprehension scores by 8.9 and 15.8 points, respectively. This translates to an 11 percent and 20 percent improvements over control group performance, in listening comprehension among students in IQS. Conversely, we do not find statistically significant effects of either type of RA lessons on listening comprehension scores of public school students. We estimate a 1.9 and 3.2 point improvement for students in the language RA and math RA groups, respectively. However, both point estimates are small in magnitude and statistically insignificant.

These findings are not surprising as the learning mechanism behind RAs is practicing listening to the teacher tell/read a story, as such, we find positive effects on listening comprehension scores. What is interesting in these results is that the structure of the RA story appears to make a difference in terms of effectiveness, where the numeracy themed RA stories induce a larger positive response from their students relative to those exposed to standard language RA stories.

4.2.3 Missing Number Identification

The following two subsections detail the effects of RA lessons on math outcomes. In this section, we discuss the findings of our experiment on missing number identification scores. This EGMA subtest entails giving students a sequence of numbers where one of the numbers is missing, and the correct answer is identifying the correct missing number. Each student is given 10 such items, thus, scoring is measured as the percent of questions answered correctly.

[INSERT TABLE 6 HERE]

Table 6 shows that both the language and math RA lessons are successful in boosting missing number identification relative to the control group, with similar effect sizes. Specifically, we estimate that the regular RA group outperforms its counterpart by 5.4 points (out of 100), while the math RA group outperforms the control group by 5.8 points. Relative to the control group average of 32.1 points, the estimated effects are equivalent to 17 and 18 percent improvements in missing number identification performance, respectively. Similarly, the analysis shows that girls from both treatment groups exhibited a substantial gain in missing number identification relative to the control group. Girls in the Language RA group improved their missing number scores by 10.2 points, whereas girls in the math RA group improved by 7.4 points. Although the effect size difference between the Language RA effect and the math RA effect on girls is about 2.8 points, this difference is not statistically significant. In terms of the evaluating the same effects on boys, we find that the math RA lessons had a higher positive impact than the Language RA lessons, with boys in the first group improving by 6.7 points relative to 3.2 points in the latter group.

The missing number subtest results are similar to those from the listening comprehension analysis in terms of disaggregating the RA estimates by SES group. The low SES group exhibits improvements in missing number scores in both the language and math RAs relative to the control group. We estimate that the language and math RA treatment effects increase missing number identification scores by 11.2 and 9.5 points, respectively. Relative to the control group mean of 27.7 points, these effects translate to a 40 percent increase among the language RA group and 34 percent among the math RA group. With the high SES group, we find that the language RA group experiences a small and statistically insignificant decline in missing number scores. The high SES math RA group, however, exhibits a 14.8 point increase in missing number scores relative to the control group. The SES stratification shows that although both groups are benefitting from RAs, the high SES group is benefitting more from the math RA lessons and exacerbating a 3.4 point gap between high and low SES students in the absence of the treatment to 8.7 points.

Lastly, we show that missing number performance of students attending IQS is not affected by the language RA lessons, but similar students in the math RA group display a 10 point improvement over the control group. We note that the math RA treatment effect is not statistically significant although the effect size itself is substantial in magnitude. Among public school students, who also have the lowest average missing number identification score of all the subgroups identified in this study, exhibit a 9 and 9.8 point score increases for students in the Hausa and math RA groups, respectively. Similar to the findings from the reading outcomes, we do not expect a direct link between RA stories, math-themed or otherwise, and missing number identification as that skill is not explicitly taught during the RA lesson. However, we hypothesize that there is an indirect link, such that students exposed to RAs are acquiring an intermediate skill that, in turn, improves students' mathematical logic and sequence identification.

4.2.4 Word Problem Solving

Table 7 presents the final set of results showing the impacts of language and math RA lessons on math word problem solving. The math word problem subtest is structured similarly to the listening comprehension subtest, but is math-themed and students are required to answer six numeracy and math questions. The word problem subtest is scored on a scale of zero to 100 points.

[INSERT TABLE 7 HERE]

Examining the effects on the full sample, we find that both the language and math RA groups outperform the control group, with the math group showing the larger impact. Moreover, the treatment effect in the math RA group is an 11.8 point increase in math word problem solving while that of the regular Language RA group is estimated to be a 3.3 point improvement. Relative to the counterfactual mean, these effects translate to a 21.3 percent and a 6 percent increase in math word problem scores. In terms of boys, students attending IQS, and students attending public school, we find a consistent pattern such that both RA types positively influence word problem performance, with math RAs exhibiting the larger impact. Specifically, math RAs increase word problem scores by 13.8, 14.9, and 16.6 points among boys, IQS students, and public school students, respectively. Whereas, the regular Language RA improves word problem scores by 2.1, 12, and 7.4 points, respectively.

The treatment effects on the low SES group are positive for both treatments, although the effect is larger in magnitude for the math RA group. The standard language and math RAs both increased math word problem solving scores by 7.6 and 14.3 points, which translate to increases of 14.6 and 27.6 percent, respectively. Among the high SES sample, we find that only the math RA lessons positively affected math word problem solving, by 25.5 points. Relative to the control group mean, that is a 52 percent improvement over no RA exposure. In terms of ascertaining the effects of RAs on girls, we also find that girls in both treatment groups show positive effects in word problem solving. The effect size of the language RA effect is larger than the math RA effect, where girls in the first group exhibit an 11.8 point increase in word problem scores while those in the second group show a 3.9 point increase. However, in testing for the difference in effects, we find that the language and math RA lessons are not different statistically.

These results are unsurprising as the RAs structure, in general, trains children's listening skills. With the infusion of the RA stories with math themes, this is translating into tangible effects in terms of being able to listen to a mathematical passage and identify and synthesize numbers, patterns, combinations, and comparisons.

4.3 Study Limitations

Even with a randomized experimental setting, there are still limitations to this study. First, the randomization of the treatments was not perfectly successful as some observable characteristics differed between the treatment and control groups. The differences we find in observables could bias the impact estimates of RAs. As a result, we rely on matching techniques to balance the samples, which may in turn alter the balance in unobservable factors that could influence student assessment performance. Second, as discussed in the experiment design section of the paper, our data lack student numeracy performance prior to the treatment period. This means that we can only assume that numeracy outcomes are similar between the treatment and control groups prior to the treatment to ensure that our numeracy impact estimates are causal.

RA lesson observation data were not collected and it is possible that language RAs and math RAs were not delivered with full fidelity. We are thus unable to provide further insight into the mechanisms by which children are improving their learning outcomes via the RA lessons. Additionally, we would expect a certain level of heterogeneity in classroom RA delivery which could potentially influence our effect size estimation. The National Literacy Panel in 2010, among numerous other studies, discuss that style and frequency, and more generally the quality, of RA delivery can alter its impacts on student learning outcomes.[15] In this case, we argue that less than perfect fidelity in RA delivery would bias our estimates downward, which would not modify this paper's conclusions.

---

[15] Also see Dickinson and Smith (1994), Hemmati et al. (2015), and Piasta et al. (2012).

The experiment is conducted in only one school year (9 months in total). As a result, our current data and research design do not allow us to test for any long-term or temporal effects of RAs. This does, however, provide an avenue for further research to add to the contributions of this paper with richer data and a longer timeframe. For instance, including student and teacher longitudinal elements and maintaining the experiment over a longer period would enable the identification of the longer-term effects of RAs on learning outcomes and learning growth. Lastly, the geographical context of this paper means that the external validity of our findings can be questioned as these effects may not be applicable universally. We can only generalize our findings to settings that are similar to Northern Nigeria. As such, these results are relevant mostly to other Sub-Saharan African countries that are mostly rural, with a high level of poverty and illiteracy among the population.

5 Policy Implications

This research revealed that the delivery of RA lessons in both the standard format and with math themes can be effective in advancing literacy and numeracy outcomes in one of the poorest regions of a low-income country. This evaluation also provides some evidence that RAs can be effective under low-resource classroom conditions.[16] Globally, as many as 250 million children are reported as unable to read (Education for All Global Monitoring Report, 2014). The evidence generated from this evaluation can have potentially far-reaching policy implications with a modest modification to existing literacy and numeracy interventions. Even with donor funding focused on improving reading skills and not necessarily numeracy or math, the implementation of RAs in both the standard and numeracy-themed formats can aid in the larger goal of bridging the literacy and numeracy gap. Our findings are in line with stated objectives of agencies such as USAID who have dedicated almost 900 million USD in official development assistance (ODA) as of 2016, reaching approximately 50 million children, to aid and develop primary education in developing countries.[17] The RANA RA approach, especially with the addition of math-themed RAs, is innovative and can be applied to supplement existing and new reading and numeracy programs in a straightforward manner.

Our RA impact estimates suggest that early grade reading and math interventions, at least internationally, should include both types of RAs to maximize their benefits on student learning outcomes. However, it is important to consider the budgetary constraints that such an approach may adversely affect. In such cases, our research suggests that math RAs can have a relatively broader impact on learning outcomes than the standard language RAs. We show that math RAs significantly improve listening comprehension

[16] Northern Nigerian public schools boast average class sizes of approximately 75 students in early primary education, and classes as large as 270 students (Smiley and Moussa, 2018).
[17] Latest USAID education spending and budget shares are retrieved from: https://www.usaid.gov/results-and-data/budget-spending

and numeracy skills, on average and across almost all student groups, in addition to improving reading outcomes for girls. Regular RAs had clear positive and relatively larger impacts on reading outcomes across all students.

Lastly, the findings of this paper also shed light on the equity implications of primary school reading and math programs such as RANA that are delivered in similar countries in Sub-Saharan Africa. We find that inclusion of RAs into the teaching curriculum is beneficial in improving girls' learning outcomes, thus bridging the gender learning gap. Our findings also show that such interventions can also exacerbate learning inequalities based on socioeconomic status, where students from relatively advantaged backgrounds benefit more than their less advantaged counterparts. It is therefore important to keep these heterogeneous effects in mind as the delivery of RAs can be altered to ensure parity in terms of the benefits all students can attain.

6 Discussion

The objective of this paper is to provide evidence of the effectiveness of RA lessons in improving early grade reading and mathematics learning outcomes. We employ a cluster-randomized controlled trial to identify the impacts of regular and math-themed RA stories on learning outcomes such as oral reading fluency, listening comprehension, missing number identification, and math word problems. Because we are unable to identify the exact mechanisms that aid students in gaining reading and math skills, the design of our experiment detects the net causal effect of RA instruction on learning outcomes. The empirical results show that both types of RAs are, to a degree, effective in boosting reading and math learning outcomes. With regular RAs producing larger effect sizes on reading and math-themed RAs yield larger effects on listening comprehension and math outcomes.

Specifically, we estimate that regular RA lessons increase the probability that a child can read at least a single word by 11 percent overall, and by about 13 percent for girls and low SES students. The regular RAs also produced significant gains in oral reading fluency. Relative to a standard deviation of 18.1, language RAs yielded effect sizes of about .17 SD for the overall sample, and .30 SD for girls. We also find that RAs were effective regardless of SES group, although the effect size was larger for the high SES group, .24 SD compared to .16 SD. It is worth noting that these effect sizes are in line with impact estimates from similar contexts. For instance, Jukes et al. (2017) estimate the impact of a teacher professional development intervention on literacy outcomes and find effect sizes up to .64 SD. Other impact studies of literacy and numeracy interventions implemented in Sub-Saharan Africa estimate effect sizes between .14 SD and .65 SD (McCoy et al., 2017, Aber et al., 2017).[18] It is important to note that these evaluations are of treatments that are

[18] Refer to Piper (2017) and Gove (2017) for a full synthesis of these studies.

typically a combination of different activities such as teacher professional development, teacher mentoring and coaching, and lesson plans.

The math RA lessons, on the other hand, were relatively more effective than the language RAs in improving listening comprehension, missing number identification, and math word problem scores. Math RAs increased listening comprehension scores by 8.8 percent, on average, and by up to 21 percent for the high SES group. Further, we estimate that missing number identification scores are improved by 18 percent, overall, by 25 percent among girls, and by 34 percent among low SES students. Similarly, we find that math RAs increase math word problem solving scores by 21 percent, on average, and by between 7 and 52 percent depending on the student subgroup.

The results of this study show that the inclusion of RA lessons into what is otherwise a typical early grade reading intervention can yield gains in student learning outcomes over and above the potential gains from the intervention as a whole.[19] As such, this study contributes a unique perspective to the early grade research literature as we measure impacts of a subset of activities. This approach is unlike typical evaluations that evaluate the broader package of activities that makes up a reading or mathematics intervention. Our findings suggest that a straightforward modification to a typical intervention that includes regular and math-themed RA lessons can yield substantial gains in reading and math skill acquisition.

[19] Early grade education interventions typically include elements of teacher professional development, teacher coaching and mentorship, and pre-specified lesson plans.

7 References

Aghion, P., L. Boustan, C. Hoxby, and J. Vandenbussche. 2009. The causal impact of education on economic growth: Evidence from US.

Bartlett, Lesley, and Julia Frazier. 2015. Literacy and development. *Routledge Handbook of International Education and Development*.

Biemiller, Andrew, and Catherine Boote. 2006. An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology* 98 (1): 44.

Campbell, MJ, A. Donner, and N. Klar. 2007. Developments in cluster randomized trials and statistics in medicine. *Statistics in Medicine* 26 (1): 2-19.

Campbell, Robin. 2001. *Read-alouds with young children*. ERIC.

Dickinson, David K., and Miriam W. Smith. 1994. Long-term effects of preschool teachers' book readings on low-income children's vocabulary and story comprehension. *Reading Research Quarterly*: 105-22.

Donner, A., and N. Klar. 2004. Pitfalls of and controversies in cluster randomization trials. *American Journal of Public Health* 94 (3) (Mar): 416-22.

Fielding, Linda, and Cathy Roller. 1992. Making difficult books accessible and easy books acceptable. *The Reading Teacher* 45 (9): 678-85.

———. 1992. Making difficult books accessible and easy books acceptable. *The Reading Teacher* 45 (9): 678-85.

Fisher, Douglas, James Flood, Diane Lapp, and Nancy Frey. 2004. Interactive Read-Alouds: Is there a common set of implementation practices? *The Reading Teacher* 58 (1): 8-17.

Gove, A., C. Chabbott, A. Dick, J. DeStefano, S. King, J. Mejia, and B. Piper. 2015. Early learning assessments: A retrospective. background paper for education for all global monitoring report 2015.

———. 2015. Early learning assessments: A retrospective. background paper for education for all global monitoring report 2015.

Gove, Amber, and Anna Wetterberg. 2011. *The early grade reading assessment: Applications and interventions to improve basic literacy*. ERIC.

Gove, Amber. 2017. What we are learning about early education in sub-Saharan Africa. Journal of Research on Educational Effectiveness 10 (3) (07/03): 530-4.

Hanushek, Eric A. 2013. Economic growth in developing countries: The role of human capital. *Economics of Education Review* 37: 204-12.

Hemmati, Fatemeh, Zeinab Gholamrezapour, and Gholamreza Hessamy. 2015. The effect of teachers' storytelling and reading story aloud on the listening comprehension of iranian EFL learners. *Theory and Practice in Language Studies* 5 (7): 1482.

Hong, Haekyung. 1996. Effects of mathematics learning through children's literature on math achievement and dispositional outcomes. *Early Childhood Research Quarterly* 11 (4): 477-94.

Jennings, Clara M., James E. Jennings, Joyce Richey, and Lisbeth Dixon-Krauss. 1992. Increasing interest and achievement in mathematics through children's literature. *Early Childhood Research Quarterly* 7 (2): 263-76.

Lee, Valerie E., and Tia Linda Zuze. 2011. School resources and academic performance in sub-saharan africa. *Comparative Education Review* 55 (3): 369-97.

Montenegro, Claudio E., and Harry A. Patrinos. 2014. Comparable estimates of returns to schooling around the world.

National Early Literacy Panel. 2010. *Developing early literacy: Report of the national early literacy panel*. Washington, DC: National Institute for Literacy.

Penno, Julie F., Ian AG Wilkinson, and Dennis W. Moore. 2002. Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the matthew effect? *Journal of Educational Psychology* 94 (1): 23.

Piasta, Shayne B., Laura M. Justice, Anita S. McGinty, and Joan N. Kaderavek. 2012. Increasing young children's contact with print during shared reading: Longitudinal effects on literacy achievement. Child Development 83 (3): 810-20.

Piper, Benjamin. 2017. Improving learning in sub-saharan africa using rigorous research designs. Journal of Research on Educational Effectiveness 10 (3) (07/03): 535-40.

Sénéchal, Monique. 1997. The differential effect of storybook reading on preschoolers' acquisition of expressive and receptive vocabulary. *Journal of Child Language* 24 (1): 123-38.

Swanson, Elizabeth, Sharon Vaughn, Jeanne Wanzek, Yaacov Petscher, Jennifer Heckert, Christie Cavanaugh, Guliz Kraft, and Kathryn Tackett. 2011. A synthesis of read-aloud interventions on early reading outcomes among preschool through third graders at risk for reading difficulties. *Journal of Learning Disabilities* 44 (3): 258-75.

Trelease, Jim. 2013. *The read-aloud handbook*. Penguin.

van den Heuvel-Panhuizen, Marja, and Iliada Elia. 2012. Developing a framework for the evaluation of picturebooks that support kindergartners' learning of mathematics. *Research in Mathematics Education* 14 (1): 17-47.

van den Heuvel-Panhuizen, Marja, Iliada Elia, and Alexander Robitzsch. 2016. Effects of reading picture books on kindergartners' mathematics performance. *Educational Psychology* 36 (2): 323-46.

Young-Loveridge, Jennifer M. 2004. Effects on early numeracy of a program using number books and games. *Early Childhood Research Quarterly* 19 (1): 82-98.

8 Tables and Figures

Table 1. Sample summary, by treatment status

	Control Cluster	Language Read-Aloud	Math Read-Aloud
Student Characteristics:			
Female	.420	+0.00	+.023
Age	9.88	-.642	-.502
Ate breakfast before class	.895	-.032	+.031
Lives with teacher	.136	-.045	+.008
Has school book	.828	+.010	+.056
Is read to at home	.466	+.054	-.130*
Gets help with homework	.456	+.036	-.099*
Has learning materials at home	.538	+.043	-.063
Wealth Index	.085	+.033	-.222*
Speaks Hausa at home	.976	-.007	-.010
Teacher Characteristics:			
Teacher is Female	.106	+.071	-.010
Age	36.1	-1.25	+1.22
Experience (years)	10.4	-.617	+1.32
Educational attainment	.571	+.031	-.166*
Full-time teacher	.786	+.066	+.067
Classroom Characteristics:			
Multi-grade classroom	.019	-.010	-.004
Class Size	67.1	-5.15	+12.0
Attendance (%)	.798	-.015	-.075*
Walls show students' work	.522	+.120	+.086
Observations	***1,099***	***581***	***564***

Notes: Figures under "Control Cluster" headings indicate the mean value for each variable, while figures under Hausa and Math Read-Aloud headings are mean differences from the control group.

Table 2. Matched sample summary, by treatment status

	Control Cluster	Language Read-Aloud	Math Read-Aloud
Student Characteristics:			
Female	.394	+.052	+.067
Age	9.94	-.242	-.528
Ate breakfast before class	.874	-.015	+.014
Lives with teacher	.116	-.051	-.034
Has school book	.830	+.043	+.039
Is read to at home	.489	+.062	-.028
Gets help with homework	.470	+.036	+.005
Has learning materials at home	.561	+.034	+.018
Wealth Index	.085	+.025	-.134
Speaks Hausa at home	.976	-.012	-.008
Teacher Characteristics:			
Teacher is Female	.139	+.108	+.053
Age	35.0	-.376	-.497
Experience (years)	9.91	-.503	-.380
Educational attainment	.558	+.085	-.083
Full-time teacher	.866	+.062	+.004
Classroom Characteristics:			
Multi-grade classroom	.012	-.006	-.005
Class Size	63.9	-5.99	+5.52
Attendance (%)	.798	+.001	-.003
Walls show students' work	.567	+.131	+.008
Observations	***1,099***	***581***	***564***

Notes: Figures under "Control Cluster" headings indicate the mean value for each variable, while figures under Hausa and Math Read-Aloud headings are mean differences from the control group. Asterisks indicate statistically significant mean differences at the .01 level. Number of observations in the "Matched" panel refer to the weighted observations using propensity scores as inverse probability weights. Statistical significance is determined using a simple t-test.

Table 3. Effect of Read Alouds on the Probability of Non-Reading (ORF score = 0)

	Language RA	Math RA	Lang - Math	Control	Observations
Full Sample	-0.070**	+0.038	-0.108**	0.643	2,046
Female	-0.095*	-0.028	-0.067	0.711	872
Male	-0.037	+0.043	-0.080	0.607	1,169
Low SES	-0.093***	+0.087	-0.180***	0.705	1,241
High SES	-0.001	+0.023	-0.024	0.617	773
IQS	-0.044	+0.052	-0.150	0.472	556
Public	-0.100***	-0.009	-0.091*	0.756	1,369

Notes: Non-readers are identified as receiving a score of zero on ORF test. Coefficients are estimated using propensity scores as inverse probability weights in regression analysis. Figures under Control Group heading refer to the conditional mean for outcomes (counterfactual). All regression specifications include student characteristics, teacher characteristics, and classroom characteristics as control variables. Asterisks denote statistical significance as follows.
*** p < .01, ** p < .05, and * p < .10

Table 4. Effect of Read Alouds on Oral Reading Fluency (cwpm)

	Language RA	Math RA	Lang - Math	Control	Observations
Full Sample	+3.116***	+1.328	+1.788	6.627	2,046
Female	+5.409***	+3.967*	+1.442	4.617	877
Male	+1.413	+0.628	+0.785	7.654	1,169
Low SES	+4.587***	-0.424	5.010**	5.298	1,271
High SES	+4.372	+2.337	+2.035	8.637	779
IQS	-0.109	-3.124	-0.170	8.522	524
Public	+4.737*	+4.706*	+0.031	6.537	1,391

Notes: Oral reading fluency outcomes are measured as correct words per minute. Coefficients are estimated using propensity scores as inverse probability weights in regression analysis. Figures under Control Group heading refer to the conditional mean for outcomes (counterfactual). All regression specifications include student characteristics, teacher characteristics, and classroom characteristics as control variables. Asterisks denote statistical significance as follows.
*** p < .01, ** p < .05, and * p < .10

Table 5. Effect of RAs on Listening Comprehension (% correct)

	Language RA	Math RA	Lang - Math	Control	Observations
Full Sample	+0.021	+0.069***	-0.048**	0.782	2,030
Female	+0.056	-0.004	+0.060	0.789	867
Male	+0.001	+0.091***	-0.090***	0.786	1,163
Low SES	+0.067***	+0.073**	-0.006	0.752	1,220
High SES	-0.026	+0.163***	-0.190***	0.764	778
IQS	+0.089*	+0.158***	-0.070	0.8	541
Public	+0.019	+0.032	-0.012	0.756	1,372

Notes: Listening comprehension scores are measured as percent of questions answered correctly. Coefficients are estimated using propensity scores as inverse probability weights in regression analysis. Figures under Control Group heading refer to the conditional mean for outcomes (counterfactual). All regression specifications include student characteristics, teacher characteristics, and classroom characteristics as control variables. Asterisks denote statistical significance as follows.
*** p < .01, ** p < .05, and * p < .10

Table 6. Effect of Read Alouds on Missing Number Identification (% correct)

	Language RA	Math RA	Lang - Math	Control	Observations
Full Sample	0.054***	0.058**	-0.004	0.321	2,023
Female	0.102**	+0.074	+0.028	0.293	865
Male	+0.032	0.067***	-0.035	0.339	1,158
Low SES	0.112***	0.095***	+0.016	0.277	1,211
High SES	-0.028	0.148*	-0.176***	0.311	780
IQS	-0.013	+0.100	-0.113	0.415	545
Public	0.090***	0.098***	-0.008	0.261	1,365

Notes: Missing number scores are measured as percent of questions answered correctly. Coefficients are estimated using propensity scores as inverse probability weights in regression analysis. Figures under Control Group heading refer to the conditional mean for outcomes (counterfactual). All regression specifications include student characteristics, teacher characteristics, and classroom characteristics as control variables. Asterisks denote statistical significance as follows.
*** p < .01, ** p < .05, and * p < .10

Table 7. Effect of Read Alouds on Math Word problems (% correct)

	Language RA	Math RA	Lang - Math	Control	Observations
Full Sample	0.033*	0.118***	-0.085***	0.552	2,055
Female	0.118***	+0.039	0.080*	0.56	880
Male	+0.021	0.138***	-0.117***	0.549	1,175
Low SES	0.076**	0.143***	-0.067	0.519	1,240
High SES	-0.003	0.255***	-0.258***	0.488	783
IQS	0.120*	0.149**	-0.029	0.66	545
Public	0.074**	0.166***	-0.092**	0.487	1,395

Notes: Word problem scores are measured as percent of questions answered correctly. Coefficients are estimated using propensity scores as inverse probability weights in regression analysis. Figures under Control Group heading refer to the conditional mean for outcomes (counterfactual). All regression specifications include student characteristics, teacher characteristics, and classroom characteristics as control variables. Asterisks denote statistical significance as follows. *** p < .01, ** p < .05, and * p < .10

Figure 1. Student composition of RANA schools

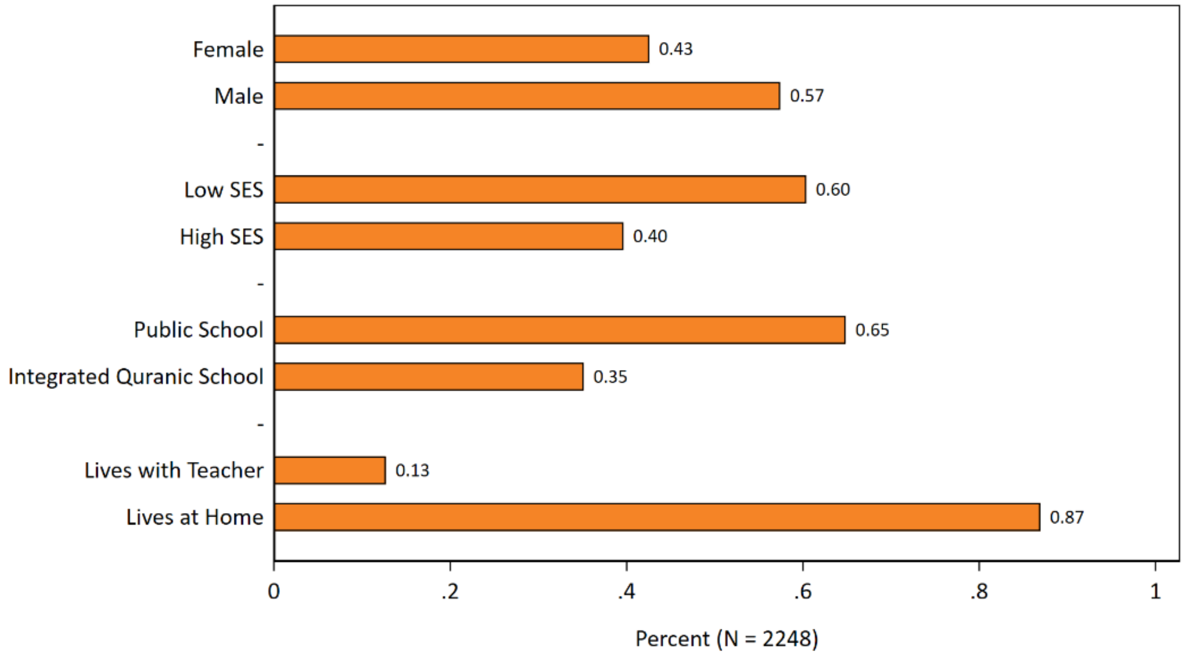
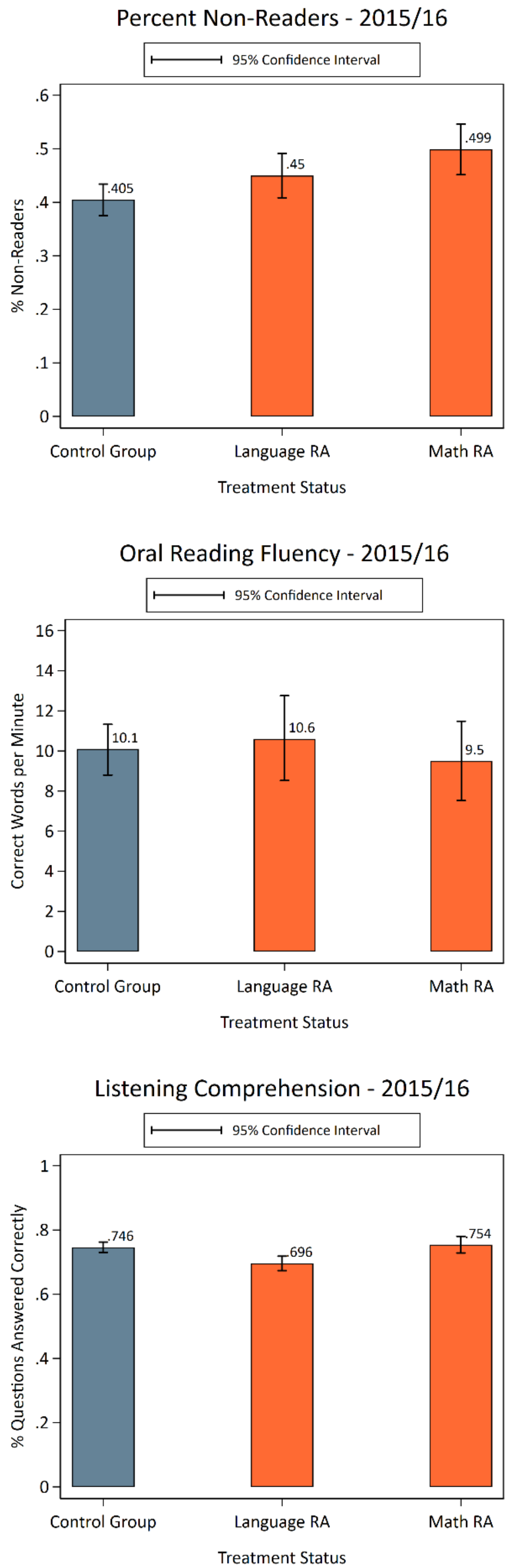


Figure 2. Test score distributions, by treatment status

Figure 3. Pre-treatment mean test scores, by treatment status



9 Appendix

A1. Sample story, English translation (vocabulary words underlined):

Once upon a time, there lived an old man who had only one daughter named Hadiza. One day, the old man called Hadiza to him. “I am very old and will not be around forever,” he told her. “I want to make sure you are happy even when I am gone.” Hadiza asked her father, “How can I be happy if you are not here?” Hadiza’s father answered, “When you live in a town with 10 mango trees at the entrance, then you will be happy.”

Hadiza searched far and wide for a town with 10 mango trees. She could not find one. She asked her friends and neighbors about a town with 10 mango trees. They could not tell her. Hadiza began to grow <u>weary</u>. What if she never found the town? But suddenly, a thought occurred to her. “Why can’t *this* town be the town with 10 mango trees?” she thought.

Hadiza went and found 10 young mango trees. One by one, she planted each tree at the entrance of her town: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. The years passed, and the trees grew and grew. Soon they had big leafy branches, and the townspeople came and sat under the shade. Soon the trees produced fruit in <u>abundance</u>. The townspeople gathered together and ate their fill. As Hadiza sat among her neighbors and ate the fruit, she thought, “I have found the town with 10 mango trees. And I am truly happy.”

Comprehension Questions

How many trees did Hadiza’s father tell her to find?

Why did Hadiza choose to plant the trees herself?

What does this story teach us about happiness?

Math Exercises

Write the number 10 on the board. Ask students to write it in the air with you. Then ask a student to write it on the board.

Ask 10 students to come to the front of the room. Each student should put up their arms as if they are tree branches. Tell the class to pretend they are Hadiza’s trees. Then ask the class to count the total trees. Then ask a student to count to 10 using a different set of objects in the room.

Write a number (0, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10) on the board. Ask students to identify the number and hold up the correct number of fingers. Erase and repeat.

EXTENSION: Read the story again. When students hear a number, they should hold up the correct number of fingers.