Endline Report

IMPACT OF REMEDIATION ON LEARNING OUTCOMES IN READING: EVIDENCE FROM A RANDOMIZED CONTROLLED TRIAL IN GHANA

November 2022

Research Team:

Nina Cunha

Wael Moussa

Marcia Davidson

Betty Temeng Mensah-Bonsu

Carina Omoeva

SUBMITTED BY:
FHI Partners LLC www.fhipartners.org
(a wholly-owned subsidiary of FHI 360)

359 Blackwell Street, Suite 200
Durham, NC 27701 USA
(919) 544-7040

CONTENTS

1. EXECUTIVE SUMMARY

Background: This study was funded by RTI International, under the Gates Foundation-funded Science of Teaching activity. The motivation for this study was to deepen our knowledge of remediation programs designed as part of larger interventions. More specifically, the main objective was to learn more about the remediation component being implemented in Ghana as a component of the comprehensive core reading program, Transition to English Plus (T2E Plus) program funded by USAID. The remediation component was designed utilizing elements from the Pratham Teaching at the Right Level (TaRL) approach, including the use of the ASER assessment tool to identify learning levels, and age-appropriate activities and games. The activities were aligned with each of the learning levels and designed to provide students with the core knowledge and skills essential for learning to read at each learning level. The remediation program was scheduled for a time of day in which core subjects were not taught, in order to complement the core reading program instruction by providing targeted students with additional support. Linking remediation to an evidence-based core program is a model that has not been reported in the literature. It is not a stand-alone remediation program, but a program that is well integrated into the national government approved evidence-based reading program.

Study summary: Using a cluster-randomized trial and a mixed-methods process evaluation, this study aims to evaluate the impact of remediation on reading outcomes of early grade students in government schools in Ghana. The English language reading remediation intervention was offered as an add-on activity within the context of the T2E Plus program. T2E Plus was implemented in Ghana from July 1st, 2021, to March 31st, 2023. The time frame of the study is January through August 2022.

Endline results: This report presents reading outcomes as measured by the Early Grade Reading Assessment (EGRA), at baseline and endline for 11 subtasks in Grades 1, 2 and 3. In addition, this report examines results from the data collected as part of the monitoring evaluation of the comprehensive USAID reading intervention for early grades in Ghana, which includes remediation practices from head teachers and teacher surveys, and classroom observation data. Finally, this report presents results of qualitative interviews with teachers and headteachers.

Sampling frame and sample: All 5425 schools across the 16 regions in Ghana participating in T2E Plus received the remediation component, except the control schools selected for this study. To facilitate logistics, we selected two regions of the country as the population for this study. The population for this study includes 307 schools and 26,524 students[1] from grades 1-3 in seven districts[2] in the Ashanti and Eastern regions in Ghana. The goal was to sample 180 schools (90 treatment and 90 control) and 2700 students from grades 1-3 (1350 treatment and 1350 control) participating in the early grade reading intervention in Ghana. The baseline sample includes a total of 2819 students (1425 from the control group and 1394 from the treatment group) from 180 schools (90 treatment and 90 control), and the endline sample includes 2793 students (1395 from the control group and 1398 from the treatment group) from the same 180 schools. We use a repeated cross section design, where students sampled at baseline and endline are not the same,

[1] 8376 students from grade 1, 8854 from grade 2, and 9294 from grade 3.

[2] The seven districts participating in this study are: Asokore Mampong Municipal, Bosomtwe, Kumasi Metro, Sekyere Afram Plains, Kwahu South, Kwahu West Municipal, and New Juaben South.

Duration: The total intervention duration was 8 months, from January 13th to August 18th, 2022. The EGRA baseline data collection took place between January 24th and February 4th of 2022, and the EGRA endline data collection took place between July 25th and August 5th.

Primary objectives: Evaluate the impact of the remediation program and its components on pupil learning outcomes to better understand the features that might provide the most effective remedial support to ensure that all children find learning success. We investigate heterogeneous effects by gender, SES, and pre-test score. Qualitative interviews with teachers help us better understand the mechanism of the effects.

Research questions

- What is the causal effect of the remediation on student achievement, in the context of an early grade reading intervention?
- What are the equity implications of the remediation? Are the effects equal by different groups, such as gender, SES, and baseline achievement levels?
- What are the mechanisms of the effect of remediation, and what are teachers' perspectives and experiences with remediation?

Measures: English and Ghanaian Language of Instruction (GLOI) EGRA administered to students at two time points (baseline and endline). In addition, we also have access to teachers and head teachers' interviews, as well as classroom observations administered by the FHI 360 team as part of the monitoring and evaluation of the reading program. As part of these interviews, teachers and headteachers are asked about remediation practices. We also conducted qualitative interviews at endline with 24 teachers (12 from control and 12 from treatment) and 6 headteachers (3 from control and 3 from treatment) to better understand the mechanism of the intervention.

Summary of findings:

- Given the severe problems we faced with compliance (75% of treatment teachers and 50% of the control teachers implemented remediation), we fail to detect any effect of remediation on EGRA outcomes.
- Descriptive statistics from the teacher and headteacher survey self-reporting on remediation practices indicate that teachers from the treatment group did dedicate more time to remediation activities when compared to the control group.
- The control teachers were performing most of the classroom general activities reported on the lesson observation more often than the treatment group, specially calling on pupils on the back, calling on boys and girls equally, and referencing the lesson plan often.
- The qualitative interviews with teachers and headteachers showed teachers are overwhelmed and that finding appropriate time to conduct remediation sessions has been a challenge for most teachers. They also feel they are working extra hours to implement remediation and are not being compensated for it. Many of the teachers declared they had to stop other activities they were doing to conduct remediation, such as physical education or cultural activities, or use time from other classes. Teachers were instructed to conduct remediation before or after school or during school break. There was no "free" period available during the school-day for remediation to be implemented.
- Based on the lessons learned from this study, we have five main recommendations for practitioners and researchers planning to implement remediation programs.

1. Ensure that teachers have sufficient time and incentives for remediation activities. It is key to guarantee teachers have a dedicated time to implement remediation as part of their daily routine, making sure they are not overwhelmed by the additional tasks and activities they are guided to perform.
2. Provide teachers with clear guidance and consistent supports for remediation. Teachers could benefit tremendously from refresher trainings or continuous monitoring and feedback.
3. Improve access to resources/materials (particularly for at-risk students/schools).
4. Establish remediation as a priority in the system. Partnering with the government and the ministry of education is essential to make sure effective measures will be in place for a successful implementation.
5. Generate rigorous, evidence-based research on remediation in LMICs. It is important to generate more evidence-based research on the effectiveness of remediation practices.

2. INTRODUCTION AND BACKGROUND

In 2013, a national Early Grade Reading Assessment was conducted in Ghana in both English and Ghanaian official local languages. Results showed that by the end of grade 2, a majority of public-school students were not able to read with comprehension in any Ghanaian language or in English. More than half of the sample of students were not able to read a single word and those that were able to read a few words were not understanding what they read (RTI, 2016). There is ample evidence that improving foundational reading and numeracy skills contributes substantially to economic development and quality of life (Hanushek & Woessman, 2012). An international focus on improving reading achievement remains a priority, with early grade reading interventions implemented at scale in several countries yielding generally positive, although mixed results in terms of improving reading skills (Graham & Kelly, 2018).

To address this need and support the country's education system, in 2016 USAID launched a national classroom-based early grade reading program, designed to improve reading instruction in the 11 official local Ghanaian languages in kindergarten through grade 2 (The USAID Ghana Partnership for Learning project). The Partnership for Learning intervention was implemented in public school classrooms in 7,200 schools in the 10 regions throughout Ghana. The reading intervention was a 60-minute daily reading program and was implemented in partnership with the Ministry of Education. Following a mid-term external impact evaluation that provided data on student learning after one-year of implementation in kindergarten and grade one, results indicated significant positive impact on student learning. However, most of the learning gains were achieved by the top 20% of students in the program and a substantial proportion of students were still unable to read any words in their local language (USAID 2018). The project implementing team examined formative data being collected twice per term and developed a plan to revise several components of the intervention with the goal of increasing student reading skills.

As part of this plan, remedial sessions held before or after school were introduced as the main new components of the intervention. The initial implementation of *Learning's* remediation program in 2019 was well received by district and regional educational directors and the Ghana Education Service. During the implementation of the Ghanaian languages reading program from 2017-2019 it was not possible to determine whether or how the supplemental remediation program contributed to the significant increases in reading achievement scores and the reduction in zero scores on the final impact evaluation.

Therefore, this study was launched in early 2022 to provide an opportunity to examine the impact that the remediation program has on the reading skills of participating struggling learners who are also participating in a core comprehensive reading program.

As context for this study, the *Learning* project implemented an 18-month Transition to English Plus program in 5425 schools across the 16 regions in Ghana including the implementation of the Ghanaian languages program, the remediation lessons for English and Ghanaian languages, and the Transition to English program. Using a cluster-randomized trial and a mixed-methods process evaluation, this study sought to evaluate the impact of a remediation intervention for English within the context of a larger reading intervention for early grades in Ghana. Treatment schools received the English remediation intervention for 12 months (3 terms), from January 2022 to December 2022, while the control schools only received the English remediation intervention between September and December of 2022 (1 term). The control schools received the English remediation component after the conclusion of this study (once endline data collection was already concluded) The primary goal of this study was thus to evaluate the impact of the English remediation program and its components on pupil learning outcomes, and to better understand the features that might provide the most effective remedial support to ensure that all children find learning success.

The study also examines heterogeneous effects by gender, SES, and pre-test score. Quantitative interviews with teachers and headteachers inform us of the frequency with which remediation is being implemented and provide a general sense of how challenging it might be for teachers to hold remediation sessions. Classroom observation data show how teacher practices differ for schools in the treatment and the control group. Qualitative interviews with teachers provide a better understanding of the mechanisms of the effects as well as teachers' perspectives and experiences with remediation. Specifically, they indicate the challenges teachers are facing when implementing the remediation sessions, how important they think the sessions are, if they believe the sessions are helping students to learn, if they perceive the remediation lessons are replacing the other core lessons, if sessions are changing their instruction in other ways, if they perceive the remediation as additional work or a burden, and if they perceive ASER implementation as a valuable tool to assess students.

The rest of the report is organized as follows: Section 3 describes the intervention and the setting in which it took place. Section 4 presents the research design. Section 5 shows pre-treatment balance using baseline data. Section 6 presents statistics on compliance followed by statistics from the Fidelity of Implementation (FOI) data collected as part of the larger reading intervention. Section 7 describes statistics on reading outcomes. Section 8 provides regression results showing the impact of the intervention on reading outcomes. Section 9 summarizes the findings of the qualitative interviews with teachers and headteachers. Section 10 concludes.

3. INTERVENTION

The Ghana *Learning* Transition to English Plus program, was an 18-month bilingual reading program with subject periods in which children learned to read in an official Ghanaian language of instruction and in English. In the *Learning* remediation 'Catch-up' program, the remedial sessions were a supplement to the core reading instruction taught to all students. Students who struggle, based upon teacher administered ASER data, joined sessions that focused on the foundational skills that the struggling students needed to learn next so that they could 'catch up' and return to the classroom with the skills necessary to be successful in the core reading program. The additional instructional support was designed to ensure that

students caught up to peers by targeting specific skills they needed and teaching these skills using highly engaging and effective learning activities. The remediation component of the program was based upon principles developed by researchers at Research for Improving Systems of Education (RISE) (Hwa et al., 2020) and included the following key points:

• *Principle 1: Set clear learning goals that are coherent with children's current learning levels. In most developing country contexts, particularly for primary school-age children, this means focusing on foundational literacy and numeracy.*

In the Ghana *Learning* program, learning goals were aligned with Ghana national curriculum standards and reflected the current scientific research from cognitive science and education on how children learn to read.

• *Principle 2: Make instruction coherent with children's current learning levels and targeted learning progress. Different approaches act on different components of instruction, including national curriculum standards, the content taught in the classroom, teaching and learning materials, and others. Many approaches bring multiple instructional components into coherence with each other.*

Learning used a modified Pratham Teaching at the Right Level (TaRL) approach in which teachers taught a set of activities in small groups selected according to their ASER learning level scores. Remedial instruction reinforced foundational skills taught in the comprehensive reading program and when students in the remediation program acquired the skills they needed to progress in the full reading program, they returned to the classroom. Each term all students were assessed to determine whether 1) students who returned were making adequate progress in the comprehensive classroom reading program, and 2) there were new students who met the learning standards for previous levels and were struggling with higher level skills and needed to be targeted for remediation. At the end of Term 3, teachers met with colleagues across grade levels to share the ASER data and student progress with colleagues in grades 1-3 to discuss plans for remediation sessions for the students who remained at risk at the end of the academic year, and any suggestions to improve learning impact of the remediation strategies at the beginning of the next academic school year.

• *Principle 3: Provide effective and coherent support to teachers. This is often delivered through practical, ongoing coaching rather than traditional teacher training.*

All teachers in the *Learning* program received 3 days of intensive professional development at the beginning of the academic year by Ghana Education Service (GES) national trainers. They also received a set of resources provided to ensure they could be effective in teaching students who were performing at different learning levels. Twice each academic term, trained regional GES coaches observed and co-taught with teachers, observed the remediation sessions, attended Professional Learning Community (PLC) meetings, and met with head teachers to provide coaching support.

• *Principle 4: Find contextually appropriate ways to implement the preceding principles. (pp. 2-4)*

In the Ghana *Learning* program, the GES, and the Ghana Ministry of Education (MOE) were partners in the co-development and implementation of the comprehensive reading program and the remediation program. The school instructional support officers monitored program implementation progress twice each academic term. A dashboard captured the monitoring data on fidelity of implementation, frequency

of coaching, absenteeism, and other quality indicators. This summary could be accessed to view data at the school, district, regional, and national levels and was updated after each monitoring visit.

The Ghana *Learning* remediation component

The first six months of the project prioritized learning how to read in GLOI for students in the last grade of kindergarten (KG2) and the first two grades of primary school (Basic 1 and Basic 2). During the last term of the school year (October-December 2021), GLOI reading lessons were taught daily as part of the program (30-mintute lessons) and remediation was only conducted for GLOI reading.

English reading classes were introduced as part of T2E Plus in the first term of 2022 to grades 1, 2 and 3 students (the same cohort of students who advanced to the next grade from 2021). Daily lessons taught in English for 60 minutes and in GLOI for 30 minutes were complemented by 30-minute remediation sessions implemented from 1 to 5 days a week in both subjects for the whole school year of 2022. The same classroom teachers taught GLOI lessons during the Ghanaian language subject period and English lessons during the English subject period.

All teachers in grades 1-3 participated in an initial 3-day training session that includes 5 hours of detailed instruction on how to conduct remediation sessions including administering and scoring the ASER reading assessment. All participating teachers received a Resource Packet during professional development training sessions that included instructions on how to conduct the remediation sessions, printable forms for scoring, recording, grouping, and monitoring the progress of students, instructions for teaching all activities by learning level, and models for conducting the weekly Professional Learning Community sessions.

During professional development training, headteachers received guidance on how to develop a school action plan to develop a schedule and provide materials to classroom teachers to conduct the remediation sessions. Teachers then developed a classroom-based action plan that was reviewed and updated each term that provided a blueprint for implementing the remediation program, including a form to document the number of learners in each remediation group.

The remediation component of the reading intervention consisted of 30-minute sessions conducted 3 to 5 times a week outside of the daily reading period and conducted by the classroom teacher or other trained teacher to the group of remedial students. The program recommended teachers to allocate time for remediation before or after school, or during an open period during the school day. Each student in grades 1-3 participated in a teacher administered ASER that was conducted at the beginning and the end of Term 1, and at the end of terms 2-3. The ASER assesses foundational reading skills by testing a child's ability to name letters or letter sounds, read words, short sentences, and a simple story (at approximately a grade 2 difficulty level). Children were grouped according to the highest score they obtained with Level 1 scores indicating non-readers and Level 6 consisting of both paragraph reading and answering comprehension questions.[3] Teachers were instructed to combine students from levels 1 and 2 in the same

[3] Level 1 scores indicate non-readers (few to no letter names/sounds); Level 2 scores indicate proficiency at letter names/sounds level and ready to learn to decode words; Level 3 scores indicate proficiency at a beginning level of reading words and students are ready to learn to read connected text (sentences); Level 4 scores reflect skill in simple sentence reading, where students are ready to learn to read more complex paragraphs; The Level 5 score indicates the student can read a simple paragraph and is ready to learn to respond to comprehension questions

remediation group, and students from levels 3 and 4 in another remediation group. They were also instructed to not deliver remediation to students from levels 5 and 6.

The remediation program consisted of game-like activities for each ASER learning level. These activities were adapted from the Pratham Teaching at the Right Level (TaRL) program (Banerjee et al., 2016) and were designed to be engaging and enjoyable for both students and teachers. For example, one activity at Level 1 involved a flash card game with small groups competing to quickly find the card with the letter that the teacher named. The activities were inexpensive and required minimal preparation for teachers. Each leveled set of activities addressed key foundational skills that are important in the process of acquiring reading proficiency. Radio lessons designed to build reading skills in English included worksheets that were aligned with the core reading intervention and were intended to be used along with the leveled activities from the TaRL program. The radio lessons were an additional option among the remediation activities and do not replace the targeted activities in the remediation guide.

The remediation program was fully aligned with the core reading program and the ASER levels broadly followed the developmental sequence of skills taught in the core reading program. All remedial activities contributed to building the requisite foundational skills necessary in learning to read. Table 1 summarizes the main components of the remediation intervention.

Table 1: Summary of remediation program components

Core English reading classes	Daily 60 minutes classes during school hours.
Remediation sessions	Additional 30 minutes instructions delivered outside of the regular school hours (before or after school, or during school breaks).
Frequency of remediation sessions	3-5 days a week recommended.
Grouping ASER for remediation sessions	Group levels 1 and 2; 3 and 4; and 5 and 6.
Training schedule and materials	3-day training for the larger reading intervention with 5 hours dedicated to remediation and ASER implementation. Teachers received a Resource Packet including instruction on how to conduct remediation. Control schools were dismissed from the 5 hours section on Remediation and had the remediation content excluded from the Resource Packet.
Additional resources	• Professional Learning Community (PLC) conducted as part of the larger reading intervention and included follow-up and discussion on remediation topics • Radio lessons that could be used during the remediation sessions

about what was read; and Level 6 (added by the project technical team) consists of both paragraph reading and answering comprehension questions.

	• District level WhatsApp platforms provided additional support for implementing the remediation • Regional school instructional support officers visited schools twice each term as part of the reading program and conducted follow-up and support for remediation

4. RESEARCH DESIGN

In this section we first present the outcomes, the sampling frame, the tools, and the timeline for the study, and we then describe the evaluation design.

Outcomes, sampling, tools, and timeline

We define three periods for data collection and analysis in this study, taking place in the year of 2022. Baseline data collection took place between January 24th and February 4th, and EGRA data was collected. Midline took place between February 25th and April 7th and Fidelity of Implementation (FOI) data was collected. Endline data collection took place between June 13 and August 5th, with FOI data collection taking place between June 13th and July 2nd, and EGRA data collection and qualitative interviews with teachers and headteachers taking place between July 25th and August 5th.

For this study we use the English EGRA as the primary outcome to measure student achievement, and the GLOI EGRA as our secondary outcome.[4] We analyze the impact of the English remediation program on each EGRA sub-task, as well as on the probability of zero scores for each task. EGRA data were collected at baseline between January 24th and February 4th 2022, before the English remediation program began, and endline data collection was collected between July 25th and August 5th 2022 at the end of the 3rd term. The school year started on January 18th and ended on December 22nd of 2022.

All 5425 schools across the 16 regions in Ghana participating in T2E Plus received the remediation component, except the control schools selected for this study. To facilitate logistics, we selected two regions of the country as the population for this study. The population for this study includes 307 schools and 26,524 students[5] from grades 1-3 in seven districts[6] in the Ashanti and Eastern regions in Ghana. The goal was to sample 180 schools (90 treatment and 90 control) and 2700 students from grades 1-3 (1350 treatment and 1350 control) participating in the early grade reading intervention in Ghana. The baseline sample includes a total of 2819 students (1425 from the control group and 1394 from the treatment group) from 180 schools (90 treatment and 90 control), and the endline sample includes 2793 students (1395 from the control group and 1398 from the treatment group) from the same 180 schools. We use a repeated cross section design, where students sampled at baseline and endline are not the same.

[4] For more details on the EGRA assessment, visit https://shared.rti.org/content/early-grade-reading-assessment-egra-toolkit-second-edition.

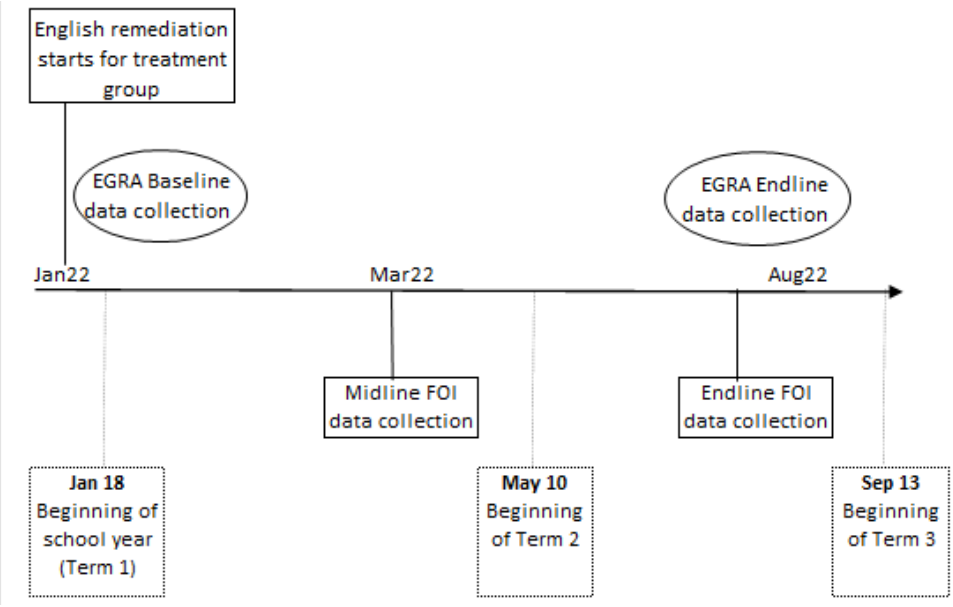
[5] 8376 students from grade 1, 8854 from grade 2, and 9294 from grade 3.

[6] The seven districts participating in this study are: Asokore Mampong Municipal, Bosomtwe, Kumasi Metro, Sekyere Afram Plains, Kwahu South, Kwahu West Municipal, and New Juaben South.

All the 90 treatment and 90 control schools received an initial remediation training to implement GLOI remediation in August/September 2021 and were instructed to implement GLOI remediation between September and December 2021. 3 months later, at the beginning of the school year in January 2022, only treatment schools received remediation training to implement remediation in English and control schools were instructed to stop delivering GLOI remediation. The sample for this study was drawn in January 2022, after GLOI remediation. In the ideal design, control schools would not have been trained in GLOI remediation and would not have implemented any remediation before the beginning of the study. However, this was the best design possible given the circumstances.

In this study we evaluate the impact of implementing the English remediation intervention for 8 months, between January 2022 and August 2022, for the first two terms of the school year. For the remainder of the year, treatment continued implementing remediation and it was also rolled out to control schools. Figure 1 below shows the timeline for data collection and study implementation.

Figure 1: Timeline



As part of the Fidelity of Implementation (FOI) data collection of the larger reading intervention, teacher and headteacher interviews, as well as classroom observations were implemented at midline (between February 25th and April 7th) and endline (between June 13th and July 2nd). Through these interviews, we obtained data on frequency of remediation and remediation practices, as well as teacher practices in the classroom coming from the classroom observations. One teacher per grade was interviewed and observed at midline, and one teacher from grade 2 and one teacher from grade 3 were interviewed and observed at endline.

Finally, we also conducted qualitative data collection with teachers and headteachers, where 24 teachers (12 from each treatment arm) and 6 headteachers (3 from each treatment arm) were interviewed at the endline (between July 25th and August 5th 2022). The goal of these interviewers was to obtain information from teachers that would allow us to better explore the mechanism of the effects and to deepen our understanding of the possible enabling factors, as described in more details above.

Evaluation design

In randomized control trials with imperfect compliance, when the treatment assignment differs from the treatment delivered, the effects of random assignment are called Intention-to-treat effects (ITT) (Angrist & Pischke 2014). The ITT approach includes all the subjects who were initially randomized according to randomized treatment assignment, ignoring noncompliance. To investigate the effects of the English remediation program on student performance we estimate intent-to-treat effects (ITT), an estimate of the impact of being offered a chance to participate in the experiment. We use a parsimonious set of controls to add in precision.

The ITT effect is estimated from Equation 1 below,

$$y_{igs} = \beta_0 + \beta_1 \bar{y}_{gs,t-1} + \boldsymbol{X'}_{igs}\,\beta_2 + \beta_3 Region_s + \delta_0 Treat_s + \varepsilon_{igs} \qquad [1]$$

Where y_{igs} is the dependent variable (EGRA subtask) for student *i,* in grade *g*, and school *s*; $\bar{y}_{igs,t-1}$ is the school average of the dependent variable at baseline; $\boldsymbol{X'}_{igs}$ represents a vector of students characteristics composed by gender and SES; $Region$ is a dummy variable for the region; $Treat_s$ is an indicator for whether the student was in a school that was offered participation in the intervention; and ε_{igs} is the error term, clustered at the school level. We estimate the regressions separately for each grade.

While the ITT provides an unbiased causal estimate of the treatment effect, it is often a diluted effect because of non-compliance issues and can provide an underestimate of the true effect (Angrist, 2006). Therefore, we also provide results of an Instrumental Variables (IV) approach in the Appendix, to assess the degree to which differential compliance affected program results.[7] We also conduct separate

[7] While treatment assignment is random, compliance is not. Therefore, to assess the degree to which differential compliance affected program results, we use an Instrumental Variables (IV) model to estimate the impact of the program on those schools that implemented remediation. The IV model allows us to capture the causal effect of treatment on the treated despite the nonrandom compliance decisions made by participants in experiments. Using the randomly assigned intent to treat as an instrumental variable for treatment delivered eliminates this source of selection bias (Angrist & Pischke 2014).

The IV estimation uses those randomly assigned into the program (that is, offered participation) as an instrument to predict remediation implementation. The IV estimate is conducted in a two-stage least-squares (2SLS) setup as initially used to adjust partial compliance in experiments by Angrist, Imbens, and Rubin (1996). In the first stage regression, we predict the degree of full engagement in the program from the random assignment. In the second stage, we regress our outcome variables on the predicted full engagement that we found in the first stage. The assumption is that schools implementing remediation satisfies the exclusion restriction that the treatment assignment only impacts outcome through the treatment itself, which in this case is true, given the random nature of treatment assignment. The regression model is shown by Equation (2),

$$y_{igs} = \beta_0 + \beta_1 \bar{y}_{gs,t-1} + \boldsymbol{X'}_{igs}\,\beta_2 + \delta_0 Remed_{gs} + \beta_3 Region_s + \nu_{igs} \qquad [2]$$

where $Remed_{gs}$ is a dummy indicating if teachers in grade *g* and school *s* implemented remediation. The associated first-stage relationship using $Treat_s$ as an instrument is equation 3:

$$Remed_{gs} = \alpha_0 + \alpha_1 \bar{y}_{gs,t-1} + \boldsymbol{X'}_{igs}\,\alpha_2 + \alpha Region_s + \pi Treat_s + \mu_{igs} \qquad [3]$$

This leads to the 2SLS estimation of Equation (4), where $\widehat{Remed}_{gs}$ is estimated in equation (3),

regressions to analyze heterogeneity of the impact of remediation, i.e., differential impacts on different subgroups. Using the Instrumental Variable model, we test for differential remediation impacts by gender, socioeconomic status (SES)[8], and pre-treatment school EGRA performance on ORF subtask[9].

5. PRE-TREATMENT BALANCE

In this section we present the pre-treatment balance of pupil characteristics and EGRA scores by treatment status using baseline data. Table 2 presents mean differences in pupil characteristics and EGRA scores by treatment status at baseline. All the 30 variables analyzed are balanced across treatment and control groups. These results show that the sample is balanced on pupils' observed characteristics.

Table 2. Mean differences in pupil characteristics and EGRA scores by treatment status

	Control (N=1425)	Treatment (N=1394)	Diff
Student Survey			
Girl (%)	50.11	49.35	-0.75
Age	8.22	8.25	0.04
High SES (%)	23.86	20.88	-2.98
Someone reads at home (%)	49.47	49	-0.48
Someone helps with homework (%)	68.21	70.44	2.23
Books at home (%)	46.32	44.69	-1.62
Student received remediation in GLOI (%)	52.14	54.66	2.52
Timed Scores			
Letter Sound Identification GLOI	17.42	17.75	0.33
Letter Name Identification English	30.59	30.93	0.34
Non-word reading GLOI	3.25	3.04	-0.21
Non-word reading English	3.96	4.03	0.07
Oral reading fluency GLOI	4.68	5.28	0.6
Oral reading fluency English	7.81	8.81	0.99
% Correct			
Oral vocabulary English	64.81	63.4	-1.41
Reading comprehension GLOI	1.56	1.71	0.15
Reading comprehension English	1.6	1.92	0.32

$$y_{igs} = \beta_0 + \beta_1 \bar{y}_{gs,t-1} + \boldsymbol{X'}_{igs}\,\beta_2 + \delta_0 \widehat{Remed}_{gs} + \alpha Region_s + \nu_{igs} \qquad [4]$$

See Appendix for more details.

[8] Students' SES was calculated as a composite based on responses to survey items about the availability of items in student households, including radio, cellphone, telephone, television, bicycle, car, electricity, and light at home to read when it is dark. The presence of more of these items indicated a higher level of SES. The index was calculated using factor analysis and then split into a binary variable where students in the top quartile of the index—the highest 25 percent of index scores—were labeled as having a "high" SES.

[9] We define "Low ORF" as a binary variable that assumes value 1 for schools in the bottom quartile of the English ORF distribution at the school level at baseline, and value 0 for the other three quartiles.

Listening comprehension GLOI	33.29	33.02	-0.26
Listening comprehension English	13.59	13.92	0.33
Zero Scores (%)			
Letter Sound Identification GLOI	20.55	19.53	-1.02
Letter Name Identification English	7.37	7.98	0.61
Non-word reading GLOI	68.08	70.86	2.78
Non-word reading English	63.84	64.88	1.04
Oral reading fluency GLOI	64.79	65.27	0.49
Oral reading fluency English	46.79	48.98	2.19
Oral vocabulary English	1.33	1	-0.33
Reading comprehension GLOI	95.09	94.26	-0.83
Reading comprehension English	96.56	96.56	0
Listening comprehension GLOI	39.79	40.75	0.96
Listening comprehension English	74.11	75.54	1.43

Note: * p<0.1 ** p<0.05 *** p<0.01.

6. COMPLIANCE AND DESCRIPTIVE STATICS USING FIDELITY OF IMPLEMENTATION (FOI) DATA

In this section we describe compliance of treatment and control groups, using EGRA endline and FOI data, to understand the percentage of each group that implemented remediation during the period of the study. We begin by showing compliance using the EGRA endline data. We then present statistics from the FOI data collected as part of the larger reading intervention. We first describe remediation practices from the headteacher survey, we then present statistics on remediation practices from the teacher survey, and we conclude showing teacher practices statistics from the lesson observation survey. We present t-statistics comparing the differences between the treatment and the control groups in the Appendix. Throughout this report, statistically significance comparing treatment and control are shown in the figures, where * indicates a p-value lower than 0.1, ** indicates a p-value lower than 0.05, and *** indicates a p-value lower than 0.01.

Compliance

During the EGRA endline data collection, the enumerators asked the teacher of each grade sampled if they implemented English remediation during the school year. 75.4% (1054) of the treatment students and 50.1% of the control students (706) were in classrooms where teachers implemented remediation. The 25.3% percent difference between treatment and control is statistically significant. In summary, there was a substantial level of non-compliance, and we account for that using the instrumental variable approach in the regression analysis, as described earlier.

We believe compliance was not perfect for three main reasons. First, as we show below, many teachers from the control group responded they attended the training on remediation, when they should not have attended it. Therefore, we believe the field team did not closely follow the instructions of excluding the control teachers from training. Second, many teachers also responded they receive the training materials on remediation, in addition to attending the training. While the team printed different materials for treatment and control groups, we believe the field team did not follow instructions on how to deliver the

materials to both groups. Third, as remediation was already embedded in the larger reading program, it was a challenge to stop teachers from doing something they were already used to doing. Teachers from both the treatment and control groups were trained on how to implement remediation in GLOI and were instructed to implement it in the term preceding the intervention.

Descriptive statistics using fidelity of implementation (FOI) data

As noted above, alongside outcome measurement at baseline and endline, the study draws on the program Fidelity of Implementation (FOI) data from the broader T2E Plus reading intervention within which the study was situated. This included teacher and headteacher interviews, as well as classroom observations implemented in March and July of 2022. We call the first round of interviews the "midline" as it falls roughly in the middle of the study (February 25 – April 7, 2022), and the second interview round the "endline" (June 13 – July 2). Teacher interviews provide information on the frequency of remediation and remediation practices, as well as teacher classroom practices coming from the classroom observations. One teacher per grade was interviewed and observed at midline, and one teacher from grade 2 and one teacher from grade 3 were interviewed and observed at endline. Because Grade 1 teachers were not interviewed and observed at endline, we only show statistics for the Grade 2 and Grade 3 teachers when describing the teacher data.

At midline, out of the total sample of 180 schools, headteacher data were received for 168 schools (85 treatment and 83 control), teacher data for 166 schools (85 treatment and 81 control), and classroom observation data for 160 teachers (81 treatment and 79 control). At endline, headteacher and teacher data were received for 173 schools (88 treatment and 85 control); classroom observation data for 170 schools (88 treatment and 82 control). It is important to emphasize that teachers interviewed as part of the FOI data collection might not be the same teachers interviewed as part of the EGRA data collection, as some schools have multiple teachers per grade.

Descriptive statistics on remediation practices using the teacher and headteacher data at midline and endline

Figure *2* below shows remediation practices from the headteacher survey at midline and endline, and Table A. *1* in the Appendix shows t-test statistics for the differences between treatment and control at each time period, for all the headteacher statistics showed in this section. **Even though teachers from the control group were not supposed to be trained on the English remediation program, 64% of headteachers at midline and 60% of headteachers at endline responded that teachers did receive the training**, against 59.9% of headteachers at midline and 54.3% of headteachers at endline for the treatment group. Differences between treatment and control are not significant, which emphasizes the concern that remediation was reported to occur in at least as many control schools as treatment schools

Note that all teachers and headteachers participated in the 3-day developing training delivered as part of the larger reading intervention, but the control teachers and headteachers were released earlier so they would not attend the training session on remediation, and they also had the remediation content excluded from the Resource Packet. Therefore, when headteachers are asked "Have Grade 1-3 teachers in your school received training in January 2022 on how to conduct remediation in English", it might be that headteachers from the control group are responding yes thinking about the general 3-day training they attended.

Figure *2* also shows that both at midline and endline, **a higher percentage of headteachers from the treatment group responded they discussed the best times to implement remediation with teachers, and that ASER was used to group students.** Finally, **a higher percentage of headteachers from the treatment group declared they observed a remediation section at midline when compared to the control**. At endline, the proportion is higher for the control group.

Figure 2: Remediation practices from headteacher survey at midline and endline

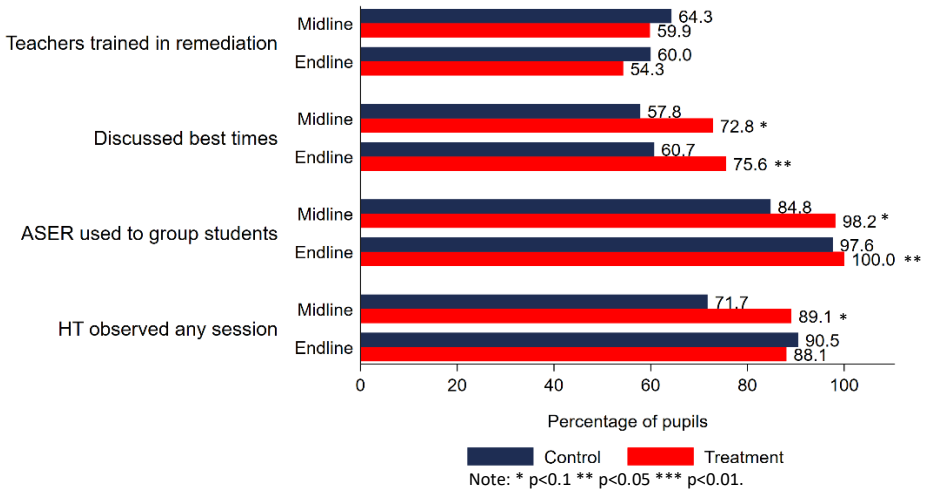
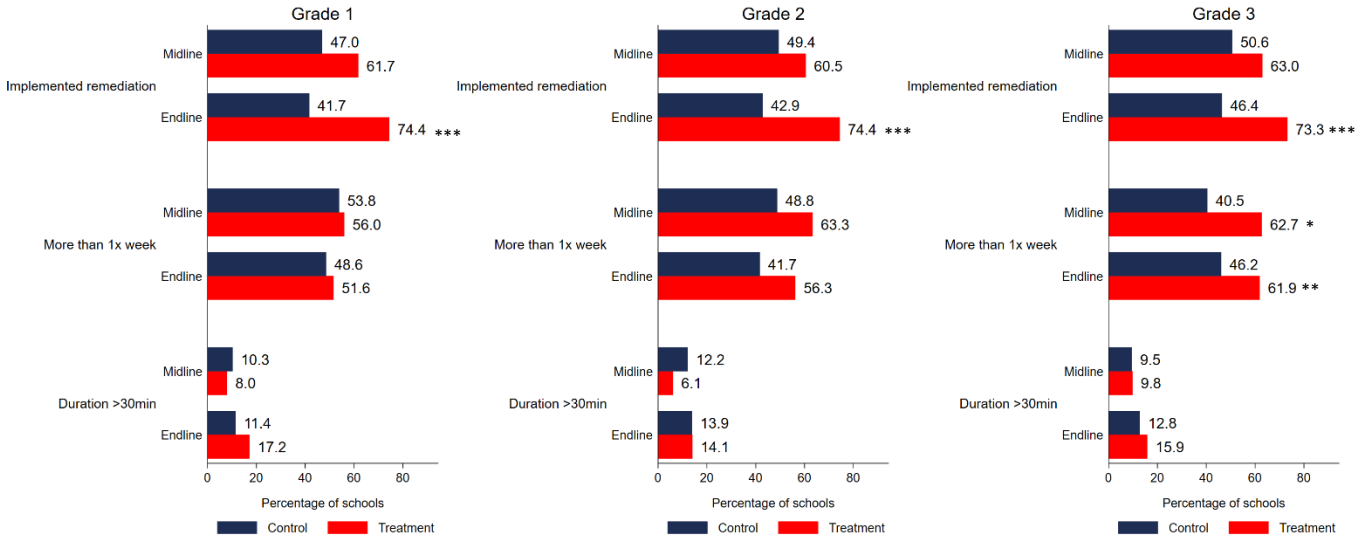


Figure *3* shows additional remediation practices from the headteacher survey by grade. For all grades, a **higher percentage of headteachers from the treatment group said the teachers are implementing remediation at both periods, and the differences are only statistically significant at endline**. At midline, around 47%-50.6% of headteachers from the control group said the school is implementing remediation, when compared to 60.5%-63% from the treatment group. This difference increases at endline: 41.7%-46.4 for the control and 73.3%-74.4% for the treatment group. The scenario is very similar when compared to the data collected during EGRA administration, as just presented in the compliance section above, confirming the imperfect compliance.

On average, **around 50% of the teachers are implementing remediation sessions more than one time per week, as opposed to once per week, and the percentage is higher for the treatment group**, especially for grades 2 and 3. When it comes to the average duration of the remediation sections, only 6% to 15.9% of the headteachers say the average duration is higher than 30 minutes, as opposed to 30 minutes or less.

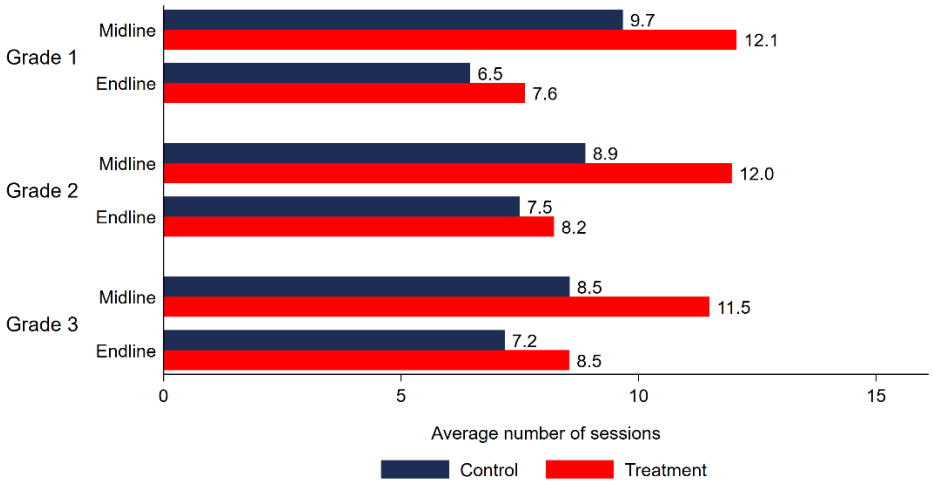
Figure 3: Remediation practices from headteacher survey, by grade



Note: * p<0.1 ** p<0.05 *** p<0.01.

Figure *4* shows the average number of remediation sessions held by teachers in each grade, according to the headteacher, at midline and endline. For all the time periods**, the treatment teachers implemented a higher number of sections when compared to the control teachers, and the average number of sections decreased from midline to endline, for all teachers.** At midline, control teachers implemented an average of 9 sections in the term, when compared to 12 sections for the treatment teachers. At endline, these numbers drop to 7 and 8 to the control and treatment groups, respectively. None of the differences between treatment and control are significant.

Figure 4: Average number of remediation sessions from the headteacher survey, by grade



Note: * p<0.1 ** p<0.05 *** p<0.01.

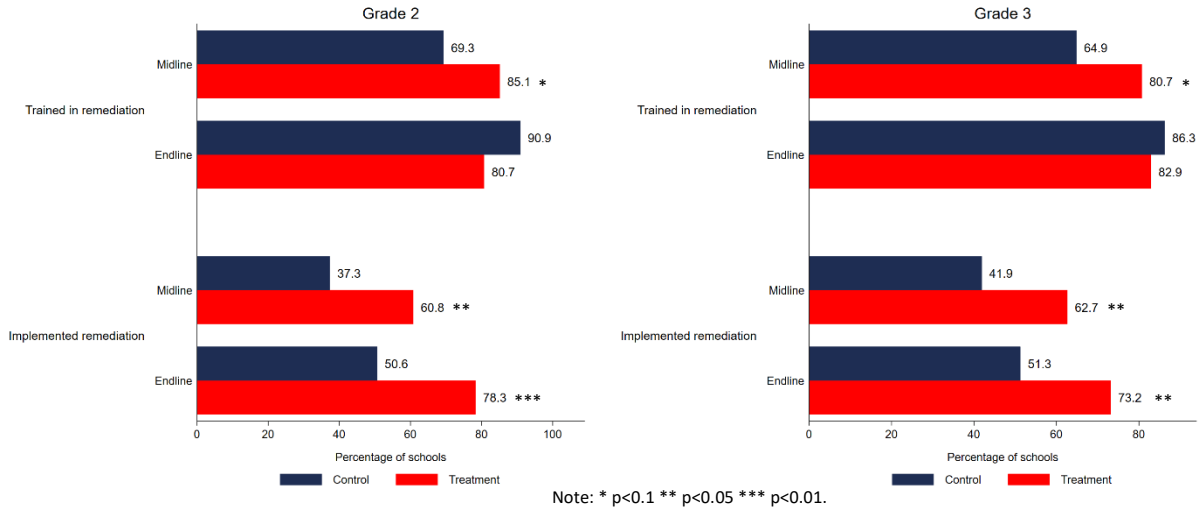
Descriptive statistics on remediation practices using the teacher data at midline and endline

Figure 5 *shows statistics on training participation and remediation implementation for teachers at midline and endline, and*

Table A. *2* in the Appendix shows t-test statistics for the differences between treatment and control at each time period, for all the teacher' statistics showed in this section**. At midline, 69.3% and 64.9% of control teachers from grades 2 and 3, respectively, confirmed they were trained in remediation in January 2022**. When asked the same question at endline, these percentages increased to 90.9% and 86.3% for control teachers from grades 2 and 3, respectively. These numbers shows that either the question was not clear to respondents, or they are responding yes thinking about the general 3-day training they received in January 2022. Responses from midline might be more reliable, given its proximity to the training when compared to endline. There was no new training after January 2022 and respondents were asked the exact same question at midline and endline. Statistics are more consistent for the treatment group and varied from 80.7%-85.1%.

Statistics on remediation implementation are similar to the scenario shown by the compliance analysis and the headteacher survey and confirm imperfect compliance. At midline, 37.3% and 41.9% of control teachers from grades 2 and 3, respectively, informed they implemented remediation in the past term and at endline these numbers increase to 50.6% and 51.3% for control teachers from grades 2 and 3, respectively. For the treatment teachers, when comparing midline and endline, the percentage increases from 60.8% to 78.3% for grade 2, and from 62.7% to 73.2% for grade 3.

Figure 5: Remediation training and implementation from the teacher survey, by grade

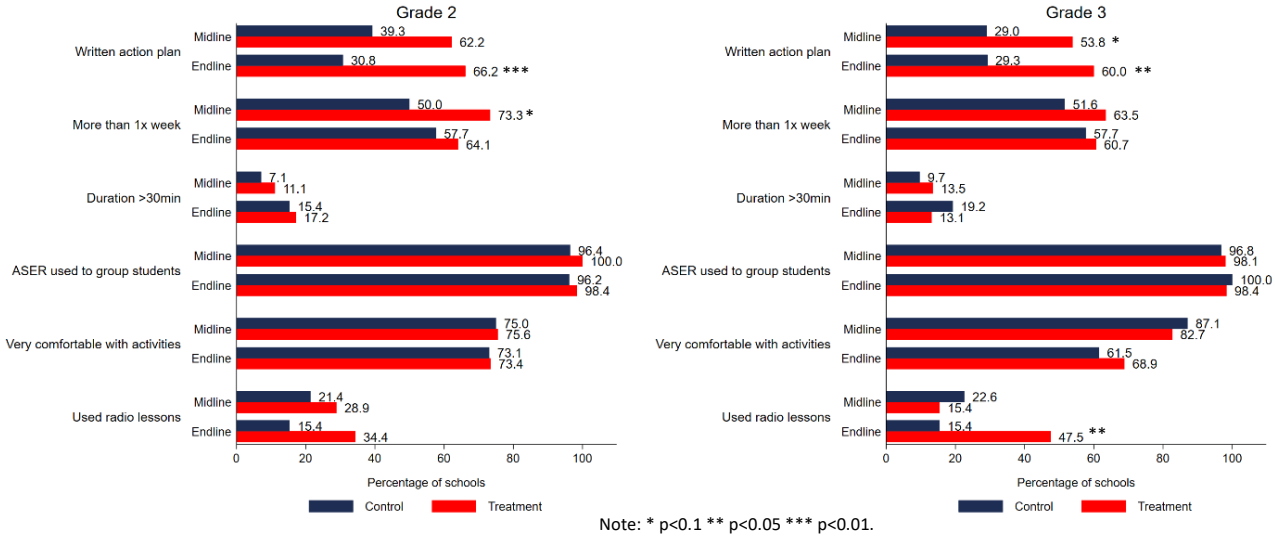


Note: * p<0.1 ** p<0.05 *** p<0.01.

Figure *6* shows statistics on remediation practices from the teacher survey by grade and treatment status. **A higher percentage of the treatment teachers reported engaging in remediation activities**: have a written action plan for remediation), conduct remediation more than once per week, duration of the remediation sections last more than 30 minutes (except for grade 3 at endline, where the frequency is higher for control teachers), and used radio lessons during the remediation section (except for grade 3 at midline, where the frequency is higher for control teachers). Responses are similar for both groups for using ASER to group students to the remediation section (slightly higher for the treatment group except for grade 3 at endline), and for being comfortable with the remediation activities (slightly higher for the

treatment group except for grade 3 at midline). Almost all teachers are using ASER to group students and a majority of teachers feel very comfortable conducting the suggested remediation activities.

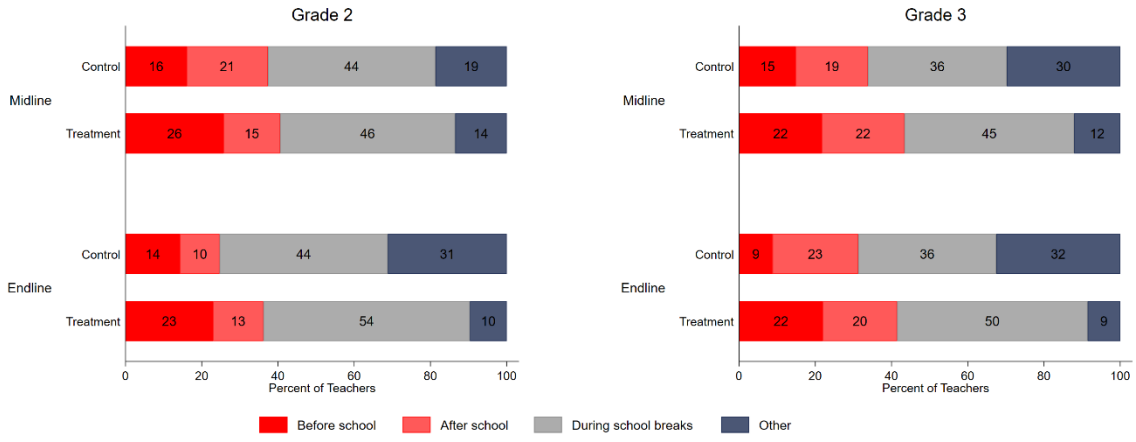
Figure 6: Remediation practices from teacher survey, by grade



Note: * p<0.1 ** p<0.05 *** p<0.01.

Figure *7* **shows school breaks are the most suitable time to implement remediation for teachers of both grades and both treatment arms**. The preference is similar for before or after school, with before school being preferred by the treatment group when compared to after school. Around 30% of the control group (19% for grade 2 midline) have preference for other time, no specified. While the treatment group received specific guidance on the time to implement remediation, the control group did not receive any guidance, which might help explain the larger proportion of "other" for the control group.

Figure 7: Most suitable time to implement remediation activities from the teacher survey, by grade

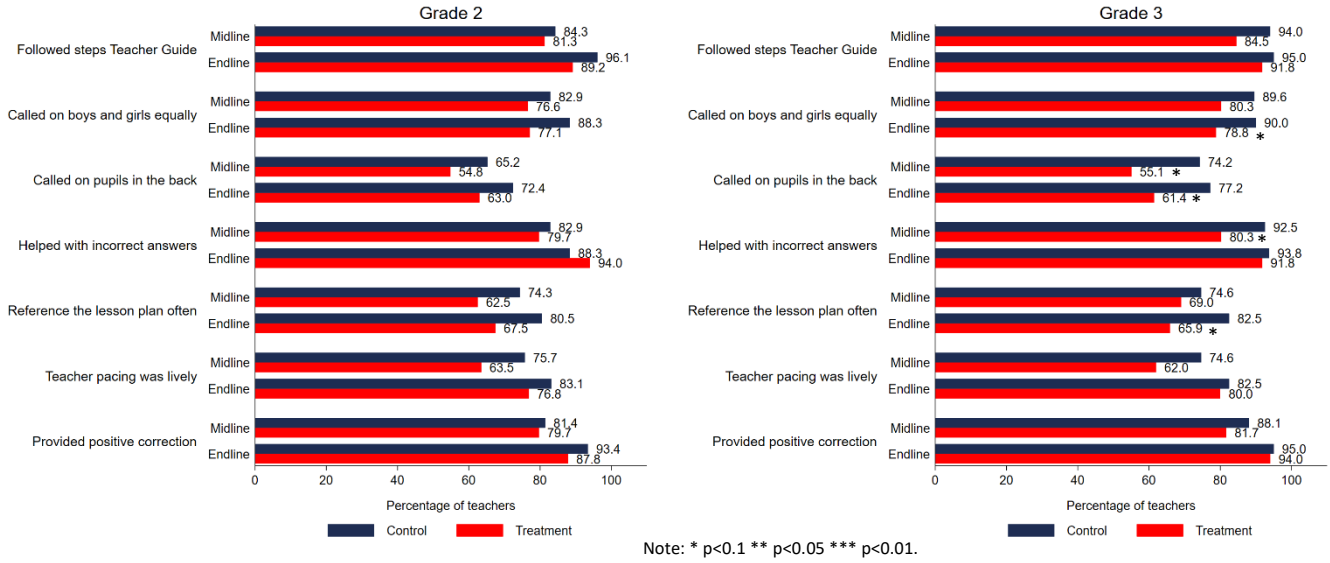


Descriptive statistics on teachers' classroom practices using the classroom observation data at midline and endline

Figure *8* shows statistics on teacher practices from the lesson observation survey by grade at midline and endline and Table A. *3* in the Appendix show t-statics for the difference between treatment and control. **It is surprising that the control group is performing all the activities listed below more often than treatment teachers at both time periods, except helping with incorrect answers at endline for grade 2.** At midline, the discrepancies are higher and statistically significant for calling on pupils in the back and helping with incorrect answers for Grade 3 teachers. At endline, the discrepancies are higher for calling on boys and girls equally, calling on pupils in the back, and referencing the lesson plan often. The differences are statistically significant for these three activities in Grade 3 at endline.

There are two possible explanations for these results. One, this might indicate that teachers from the treatment group have less time to dedicate to the classroom activities because they are busy or overwhelmed with the remediation sessions. Alternatively, teachers may be performing these activities less often in the regular classroom because they now have additional time to focus on these activities during the remediation session. As we will show in more detail when describing the results from the qualitative interview in Section 9, the first explanation is more reasonable, and teachers were indeed overwhelmed by the remediation sessions. Several studies have demonstrated the negative effect of stress on teacher (Yunarti et al. 2020, Dankade et al. 2016, Hamid et al. 2015, Anandasayanan and Subramaniam 2013).

Figure 8: Teacher practices from lesson observation survey, by grade



Note: * p<0.1 ** p<0.05 *** p<0.01.

7. DESCRIPTIVE STATISTICS ON READING OUTCOMES

In this section we first provide information on the refresher training, endline data collection and endline sample. Next, we present descriptive statistics on EGRA scores at endline by grade and treatment status and statistics on compliance. We then present the regression analysis results. In this section, we focus on

the EGRA subtasks for English, which are our main outcomes. We present results for the GLOI subtasks in the Appendix.

Refresher training, endline data collection and sample

Endline data collection was conducted by JMK Consulting between July 25th and August 5th, 2022. JMK organized a 3-day refresher training for 34 enumerators commencing from Tuesday 21st July to Thursday 28th July 2022. One inter-rater reliability test was conducted during training and all participants obtained acceptable scores of 90 percent or higher for EGRA assessment. The average IRR score was 97.8% for Asante Twi and 98.7% for Akuapem Twi. JMK visited all 180 schools and sampled an average of 5 pupils per grade. A total of 2793 students were sampled, 1398 from the treatment group and 1395 from the control group. Out of the 2793 pupils, 928 are from grade 1 (466 treatment, 462 control), 926 from grade 2 (468 treatment, 458 control), 939 from grade 3 (464 treatment, 475 control), as shown below.

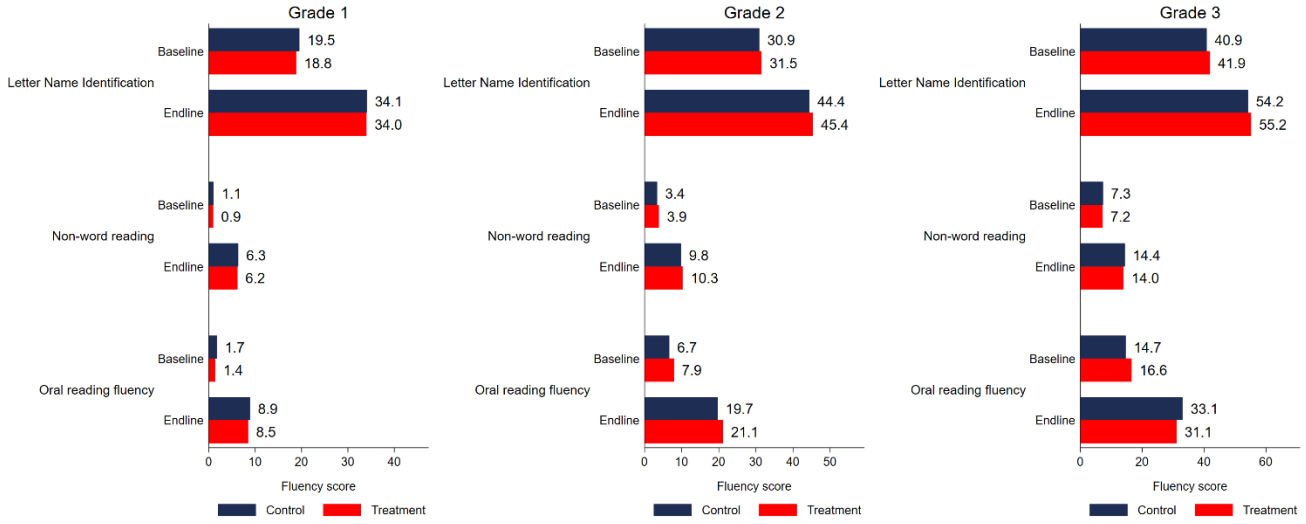
Table 3: Endline sample by grade

Grade	Control	Treatment	Total
1	462	466	928
2	458	468	926
3	475	464	939
Total	1,395	1,398	2,793

Descriptive statistics EGRA

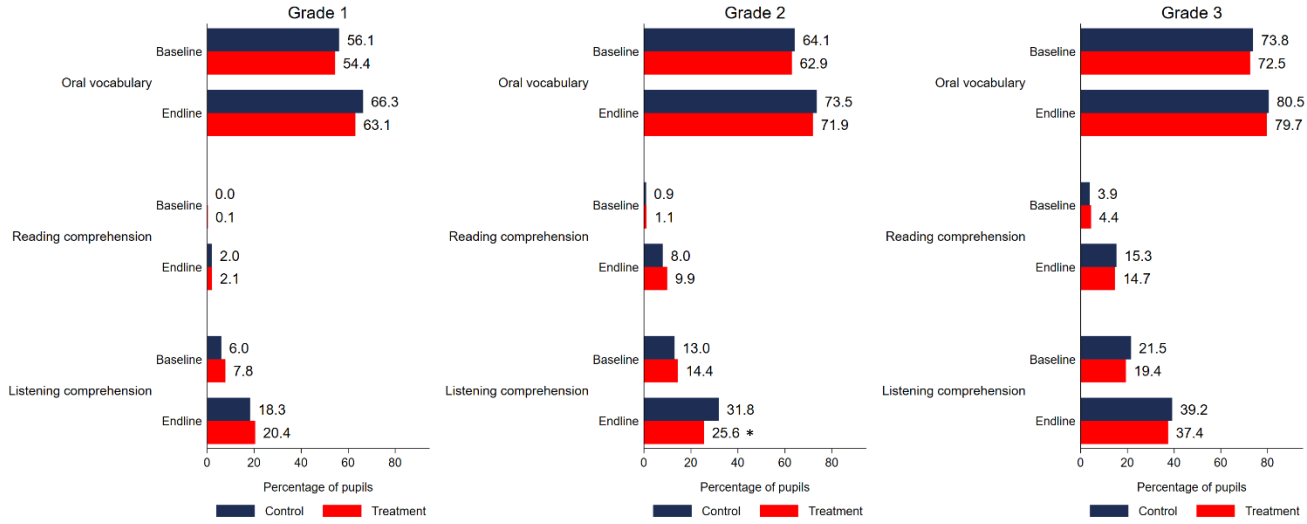
Figure 9 and Figure 10 shows the average fluency score on timed subtasks and the average percent correct items on untimed subtasks, respectively, by grade and time. Figure 11 shows the proportion of students who obtain a zero-score on English subtasks by time and grade. Overall, the average of all the subtasks increased from baseline to endline for both the treatment and the control group, and the proportion of students who obtain a zero-score decreased on all subtasks for both groups. The EGRA scores are similar for the treatment and control groups at endline for all subtasks. Table 4 shows t-statistics comparing the average endline scores for the English subtask between the treatment and control groups. There are no statistically significant differences in the EGRA English score for grade 1 students. For grade 2, the treatment group scores 6.24 percentage points (p.p.) lower than the control group on the English listening comprehension subtask, and the proportion of students obtaining a zero score on the same subtask is 8.4 p.p. higher for the treatment group. For grade 3, students in the control group score 1.67 p.p. lower in the English letter name subtask. Figure A. 1, Figure A. 2, Figure A. 3, and Table A. 4 in the Appendix show the equivalent analysis for GLOI. There are no statistically significant differences between the treatment and control averages at endline for the GLOI subtasks.

Figure 9: Average fluency scores on English timed subtasks, by grade and time



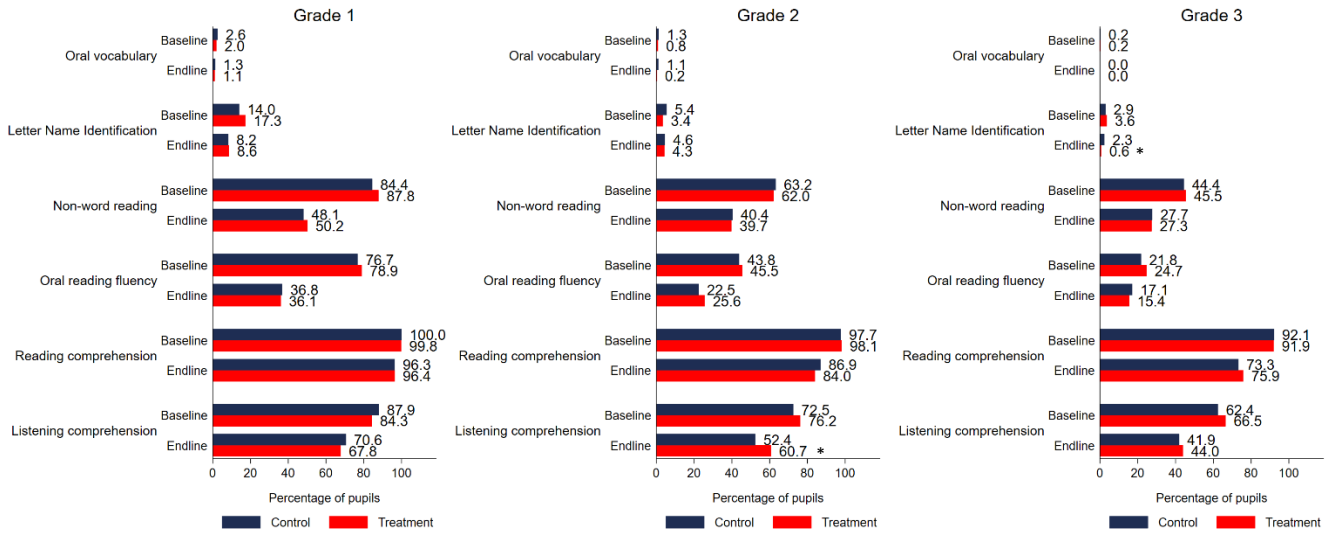
Note: * p<0.1 ** p<0.05 *** p<0.01.

Figure 10: Average percent correct on English untimed subtasks, by grade and time



Note: * p<0.1 ** p<0.05 *** p<0.01.

Figure 11: Proportion of students who obtain a zero score on English subtasks, by grade and time



Note: * p<0.1 ** p<0.05 *** p<0.01.

Table 4: EGRA descriptive statistics for English subtasks by grade and treatment status

	GRADE 1			GRADE 2			GRADE3		
	Cont.	**Treat.**	**Diff.**	**Cont.**	**Treat.**	**Diff.**	**Cont.**	**Treat.**	**Diff.**
Timed Scores									
Letter Name Identification	34.08	34.02	-0.06	44.44	45.4	0.96	54.21	55.19	0.98
Non-word reading	6.35	6.17	-0.17	9.82	10.27	0.45	14.37	13.97	-0.4
Oral reading fluency	8.89	8.49	-0.4	19.69	21.13	1.45	33.06	31.12	-1.93
% Correct									
Oral vocabulary	66.31	63.06	-3.25	73.53	71.88	-1.65	80.53	79.69	-0.84
Reading comprehension	1.99	2.06	0.07	7.99	9.87	1.88	15.28	14.66	-0.63
Listening comprehension	18.25	20.39	2.13	31.8	25.57	**-6.24***	39.16	37.36	-1.8
Zero Scores (%)									
Letter Name Identification	8.23	8.58	0.36	4.59	4.27	-0.31	2.32	0.65	**-1.67***
Non-word reading	48.05	50.21	2.16	40.39	39.74	-0.65	27.7	27.33	-0.36
Oral reading fluency	36.8	36.05	-0.75	22.49	25.59	3.1	17.12	15.4	-1.72
Oral vocabulary	1.3	1.07	-0.23	1.09	0.21	-0.88	0	0	0
Reading comprehension	96.32	96.35	0.03	86.9	83.97	-2.93	73.26	75.86	2.6
Listening comprehension	70.56	67.81	-2.75	52.4	60.68	**8.28***	41.89	43.97	2.07

Note: * p<0.1 ** p<0.05 *** p<0.01.

8. IMPACT ON READING OUTCOMES

In this section we first present the Intent-to-Treat effects, estimated using the Ordinary Least Squares (OLS) approach, for the timed and untimed EGRA English subtasks, aggregating the three grades. Next, we present the effects by grade. Results for EGRA English Zero Scores and EGRA GLOI are presented in the

Appendix. We present results using the IV approach in the Appendix, as well as heterogeneous effect by gender, socioeconomic status (SES), and pre-treatment school EGRA performance on ORF subtask using the IV approach.

Table *5* shows regression results using the ITT approach for the English EGRA timed (letter-name identification, non-word reading, and oral reading fluency) and untimed subtasks (oral vocabulary, reading comprehension, and listening comprehension), aggregating the three grades. Results for untimed subtasks are shown as standard deviation (z-score), while results for the untimed subtasks are shown as percentage. None of the results are statistically significant.

Table 5: Treatment effects on EGRA timed and untimed English subtasks – ITT estimates

	Letter Name Identification (z-score)	Non-word reading (z-score)	Oral reading fluency (z-score)	Oral vocabulary (%)	Reading comprehension (%)	Listening comprehension (%)
	(1)	(2)	(3)	(4)	(5)	(6)
Effect	0.02	-0.01	-0.03	-0.80	0.03	-2.03
	(0.06)	(0.05)	(0.05)	(1.36)	(1.42)	(2.65)
N	2787	2783	2776	2788	2788	2788

Note: Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01. Each cell shows effect of a different regression. EGRA timed subtasks were normalized relative to the distribution of the control group for each grade, such that the mean and standard deviations of the control group at each grade is zero and one, respectively. Results for the untimed subtasks showed in percentage. Standard error clustered at the school level showed in parenthesis. Controls for region, grade, gender, and SES were included in all the regressions, as well as control for the school average EGRA score at baseline. Sample size is different for each timed subtask because EGRA timed subtasks with an absolute value higher than 4 standard deviations were coded to missing.

Table *6* shows regression results using the ITT approach for the English EGRA timed and untimed subtasks by grade. The first panel shows results for each subtask for grade 1, the second panel for grade 2, and the third panel for grade 3. Each cell shows results of a different regression. Out of the 18 regression results, only listening comprehension for grade 2 is statistically significant, which is not different from what we would expect by chance. This result is negative by 6.65 percentage points (p.p.). While most of the coefficients' magnitude is close to zero or considerably low, grade 3 result for oral reading fluency is negative and with magnitude of 0.12 standard deviation (std.), even though it is not statistically significant. The direction of 67% of the results is negative, even though the magnitudes are small and the coefficients are not statistically significant.

Table 6: Treatment effects on EGRA timed and untimed English subtasks by grade – ITT estimates

	Letter Name Identification (z-score)	Non-word reading (z-score)	Oral reading fluency (z-score)	Oral vocabulary (%)	Reading comprehension (%)	Listening comprehension (%)
	(1)	(2)	(3)	(4)	(5)	(6)
Effect Grade 1	0.02	0.00	-0.00	-2.12	-0.57	1.17
	(0.08)	(0.05)	(0.04)	(2.07)	(0.74)	(3.35)
N	923	923	919	923	923	923
Effect Grade 2	0.03	-0.01	-0.00	-0.22	1.40	**-6.65***
	(0.09)	(0.07)	(0.06)	(1.75)	(1.77)	**(3.65)**
N	926	926	923	926	926	**926**
Effect Grade 3	0.02	-0.03	-0.12	-0.07	-1.24	-0.60
	(0.08)	(0.08)	(0.08)	(1.57)	(2.57)	(3.43)
N	938	934	934	939	939	939

Note: Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01. Each cell shows effect of a different regression. EGRA timed subtasks were normalized relative to the distribution of the control group for each grade, such that the mean and standard deviation of the control group at each grade is zero and one, respectively. Results for the untimed subtasks showed in percentage. Standard error clustered at the school level showed in parenthesis. Controls for region, gender, and SES were included in all the regressions, as well as control for the school average EGRA score at baseline. Sample size is different for each timed subtask because EGRA timed subtasks with an absolute value higher than 4 standard deviations were coded to missing.

Results using the IV approach are similar, with no consistent significant results. No significant results were found for GLOI EGRA scores for timed and untimed subtasks (Appendix Table A. *7*) and zero scores (Appendix Table A. *8*). Overall, remediation had no effects on EGRA outcomes, with an indication of possible negative results. We did not find any consistent evidence of heterogenous effect by gender, SES, and pre-treatment EGRA performance (Table A. *9*, Table A. *10*, and Table A. *11* in the Appendix).

9. QUALITATIVE INTERVIEWS WITH TEACHERS AND HEADTEACHERS

During EGRA endline data collection, enumerators conducted 30 qualitative interviews with teachers and headteachers from 24 schools. Schools were randomly selected from the study sample of 180 schools. A total of 24 teachers (12 from each treatment arm) and 6 headteachers (3 from each treatment arm) were interviewed. The average interview time was 47 minutes. The average interview time for the treatment group (61 minutes) was higher than for the control group (33 minutes) as more remediation questions were designed to the treatment group. Enumerators completed a debriefing spreadsheet after each interview, where they highlighted the main themes that emerged and emphasized any challenges or interesting facts from the interview. All interviews were recorded and then transcribed.

Responses to the qualitative interviews were analyzed through thematic content analysis. Analysts examined the transcripts for patterns in themes and convergences and divergences in experiences. The qualitative themes should be understood as descriptive rather than generalizable.

As shown by Table *7*, out of the 12 teachers from the treatment group, 10 affirmed they were trained on how to conduct remediation, and 9 responded to be implementing remediation in the school. As for the control group teachers, 5 out of 12 said they were trained on how to conduct remediation and 3 indicated they were conducting remediation. The three headteachers from the treatment group confirmed they received the training on remediation, and two informed remediation was being implemented in the school. Two of the headteachers from the control group stated they were trained on remediation, and one asserted to be implementing remediation in the school. Out of the 30 respondents, 20 informed they were trained on remediation and 15 confirmed the school is implementing remediation. These numbers confirm the imperfect treatment compliance showed in the previous sections.

Table 7: Remediation training and implementation from the qualitative interviews

	Teacher		Headteacher		Total
	Control	Treatment	Control	Treatment	
Trained in remediation	5	10	2	3	20
Implementing remediation	3	9	1	2	15

Four main themes emerged from the interviews: 1) challenge to find time to implement remediation; 2) teachers are organizing remediation in different ways; 3) lack of resources in the school and in the community; 4) lack of incentives to teachers. In addition, many interesting ideas emerged when respondents were asked what suggestions they would give to the Ghana Education Service to improve learning conditions in the school. We will expand on each of these themes below. In addition to the four themes that emerged from the qualitative interviews and the suggestions to GES, teachers do believe remediation is helping students, even though it's been a challenge to find appropriate time to conduct it.

Challenge to find time to implement remediation

The challenge to find appropriate time to implement remediation was a common theme in the interviews. Teachers were advised to implement remediation outside of regular class hours, such as before or after school or during school breaks. 18 of all respondents emphasized having this challenge, including 13 out of the 15 respondents implementing remediation, some of the respondents that are not implementing remediation due to this challenge, and one respondent not trained in remediation.

Respondents complained it is difficult to have students arrive earlier in the morning or to stay late after school. Some students have to go home after school with their siblings studying in the same school or have Arabic classes in Islamic schools. School breaks are also challenging as students are hungry or tired. Some teachers complained about the lack of support from parents that would not accommodate the remediation time in the pupils' schedule. In two of the schools, the teacher interviewed was the only teacher in the school, teaching all the grades. They said that finding time to implement remediation is impossible.

Ten out of the 15 respondents implementing remediation stated they had to give up other activities they were implementing to conduct remediation. Some teachers ended up using time from the timetable that were originally assigned to other subjects, while others had to stop extra-curricular activities such as sport,

art, club meetings, or others. Many teachers suggested that remediation should be part of the timetable and that the number of subjects could be reduced, to make time for remediation.

Teachers are organizing remediation in different ways

An interesting finding that emerged from the interviews is the variation on how teachers are organizing the remediation sessions. While most teachers are using ASER to group the students and have more than one group according to their levels, they found different ways to teach remediation to different groups of students. Some teachers split the group into different sessions and teach them on different days of the week. Others keep all groups in the same remediation sessions but split them into smaller groups and have them work on different activities simultaneously. Two of the teachers mentioned they were coordinating the remediation sessions with teachers from different grades, to aggregate students in similar levels. There are some teachers that simply call students one by one to deliver the remediation content according to their level. And some of the teachers will conduct remediation during the regular English classes.

Another interesting fact is that while some teachers are only assigning the struggling students to remediation, others are assigning all their students to remediation. One of the teachers even suggested adding one additional level to ASER, so the higher achieving students could also practice remediation activities assigned to their proper level, which shows that some might have used the activities as extra differentiated instruction, instead of focusing on struggling students

Lack of resources in the school and in the community

During the interviews, 70% of respondents (21) mentioned having challenges with the lack of resources in the school or in the community. The most cited challenge was with the lack of learning materials and resources in the school, such as reading books, pens, pencils, erasers, and tables. Some teachers mentioned the families did not have the resources to buy materials for their pupils and rely on the school to provide them. In addition, many teachers mentioned pupils did not have money to buy food during the school break and students are hungry during classes. Some of the teachers complained not being part of the school feeding program. Ten respondents mentioned they had used their own money to buy either food or learning materials for the students.

Teachers from rural areas mentioned the challenge of reaching the school and the lack of funds for transportation or accommodation. The commute for some teachers is very long, which makes it even harder to allocate time for remediation before or after school hours. Some teachers suggested additional incentives should be given to rural teachers, such as accommodation near the school or transportation.

Lack of incentives to teachers

The lack of incentives to perform remediation was the most frequently cited challenge during the teacher interviews. 93% (28) of the respondents mentioned teachers should receive more incentives and 60% (18) specifically mentioned financial incentives. The lack of incentives was mentioned not only in the context of remediation, but also as a general problem faced by all teachers regardless of remediation implementation. In the context of remediation, many teachers stated that they must work extra hours to be able to implement remediation and that they expected to receive additional monetary compensation.

Other types of incentives flagged by respondents were having access to basic learning materials, such as pencils, books, and erasers; incentives from parents and the community; appreciation from headteachers and the government; additional incentives to retain teachers in the rural areas; as well as more learning opportunities, such as workshops or refresher trainings; and better salaries.

Suggestions to the Ghana Education Service to improve learning conditions in the school

When asked about what suggestions they would give to the Ghana Education Service (GES) to improve learning conditions in the school, respondents touched on a variety of subjects, some of which were briefly mentioned above.

Some of the suggestions mentioned by respondents are related to remediation, such as: allocate time for remediation in the timetable; provide additional monetary incentive for teachers to implement remediation; make remediation required so that students are required to attend; increase family awareness on the importance of remediation; require headteachers to get involved in remediation; provide refresher trainings on remediation; allocate more time for the initial training on remediation.

In addition, respondents mentioned other recommendations such as learning activities for teachers, such as workshops; availability of basic resources in the school, such as exercise books, pencils, pen, and erasers; make transportation available for teachers living far away from the school, as well as accommodation for teachers in rural areas; additional incentive to retain rural teachers; and show appreciation for teachers.

Summary

In addition to the four themes that emerged from the qualitative interviews and the suggestions to GES, teachers do believe remediation is helping students, even though it's been a challenge to find appropriate time to conduct it. Most of the teachers believe ASER is well aligned with the curriculum and is an easy tool to implement. They also reported that the activities suggested in the training materials were easy to understand and execute. Finally, teachers believe it is easier to teach students in smaller groups, as in the remediation sessions, when compared to the regular class size, as they have more opportunities to engage with each student and focus on their need.

10. DISCUSSION

The different statistics presented in this study showed treatment compliance was imperfect. A high proportion of the control group declared they received training on the English remediation program and that they implemented English remediation in their schools. While we encountered a severe problem with compliance, the teachers who were part of the remediation training were more likely to implement remediation when compared to the control group. According to the data collected during the endline EGRA data collection, 75% of the treatment teachers implemented remediation, when compared to 50% from the control teachers.

The higher participation from the treatment group is also reflected on the frequency of remediation practices as reported by the headteachers and the teachers. For instance, a higher percentage of the headteachers from the treatment group discussed the best times to implement remediation with their teachers, when compared to the control group; and a greater number of treatment teachers had written an action plan for remediation, implemented remediation more than once a week, and used radio lessons

during remediation, when compared to control teachers. Therefore, teachers from the treatment group dedicated a higher portion of their time to remediation activities when compared to the control teachers.

While the higher participation of treatment group on remediation activities may lead us to expect higher EGRA performance when compared to the control group, statistics from the lesson observation might lead us to expect otherwise. The control teachers were performing most of the activities reported on the lesson observation more often than the treatment group, especially calling on pupils on the back, calling on boys and girls equally, and referencing the lesson plan often.

The qualitative interviews with teachers and headteachers showed teachers are overwhelmed and that finding appropriate time to conduct remediation sessions has been a challenge for most teachers. They also feel they are working extra hours to implement remediation and are not being compensated by it. Many of the teachers declared they had to stop other activities they were doing to conduct remediation, such as physical education or cultural activities, or use time from other classes.

The regression analysis did not show any positive results for the effect of the English remediation program and, if anything, showed signs of negative effects. Connecting the different pieces of analysis presented in this study, it is very likely that teachers from the treatment group felt more overwhelmed by the remediation activities when compared to the control group and had less time to dedicate to the regular classroom activities. There is also a possibility that the positive effects of the remediation activities have been cancelled out by the negative effects of the lesson observation activities. Note that the EGRA assessment was implemented with five students randomly selected from the whole classroom, independent of their participation in remediation activities. An effective remediation intervention should be able to improve learning for the struggling students without detriment to the learning of the other students in the classroom.

From the main themes discussed in this study, it is evident that teachers would benefit tremendously from having the remediation be a part of their regular schedule (the timetable) or have an official time assigned to it during school hours. Alternatively, if teachers are expected to conduct remediation activities outside of school hours, additional compensation could be provided, as well as adequate conditions to guarantee teachers and students can be present. In addition, teachers may also benefit from receiving more detailed guidance on how to structure the remediation sessions in practice and how to organize the sessions for the different levels of students. Finally, for remediation to be effectively implemented, some of the basic struggles faced by teachers and schools need to be addressed, such as the lack of learning materials, lack of feeding program to students, lack of support to rural teachers, and lack of incentives to teachers, including monetary incentives.

Based on the lessons learned from this study, we have five main recommendations for practitioners and researchers planning to implement remediation programs.

1. ***Ensure that teachers have sufficient time and incentives for remediation activities***. It is key to guarantee teachers have a dedicated time to implement remediation as part of their daily routine, making sure they are not overwhelmed by the additional tasks and activities they are guided to perform. Alternatively, additional compensation and incentives could be provided if teachers are expected to work extra hours.
2. ***Provide teachers with clear guidance and consistent supports for remediation.*** Given the large variation in how teachers are implementing remediation in practice, it is essential to provide

detailed instructions and materials to ensure implementation is occurring as desired. Teachers could benefit tremendously from refresher trainings or continuous monitoring and feedback.

3. ***Improve access to resources/materials (particularly for at-risk students/schools).*** We showed most teachers reported facing challenges due to the lack of resources and materials in the school. It is well known that the most disadvantaged schools are the ones needing the most support and it is vital to assure teachers have the necessary resources to successfully implement remediation.
4. ***Establish remediation as a priority in the system.*** Partnering with the government and the ministry of education is essential to make sure effective measures will be in place for a successful implementation. For instance, teachers provided interesting insights on measures that would increase their ability to efficiently implement remediation, such as making remediation required, increasing family awareness on the importance of remediation, and requiring involvement from headteachers and school personnel.
5. ***Generate rigorous, evidence-based research on remediation in LMICs.*** It is important to generate more evidence-based research on the effectiveness of remediation practices. As shown in this study, even well-designed impact evaluations might fail to detect results when facing problems during the implementation phase. Future randomized evaluations should ensure that compliance is adhered to determine if the remediation is effective for improving student outcomes.

REFERENCES

Anandasayanan, S. (2013). Effect of stress on teachers' performance with special reference to Jaffna district schools. *Available at SSRN 2386067*.

Angrist, J. D. (2006). Instrumental variables methods in experimental criminological research: what, why and how. Journal of Experimental Criminology, 2(1), 23-44.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. Journal of the American statistical Association, 91(434), 444-455.

Angrist, J. D., & Pischke, J. S. (2014). Mastering 'metrics: The path from cause to effect. Princeton university press.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukherji, S., Shotland, M., & Walton, M. (2016). Mainstreaming an effective intervention: Evidence from randomized evaluations of "Teaching at the Right Level" in India. *National Bureau of Economic Research*, w22746.

Dankade, U., Bello, H., & Deba, A. A. (2016). ANALYSIS OF JOB STRESS AFFECTING THE PERFORMANCE OF SECONDARY SCHOOLS'VOCATIONAL TECHNICAL TEACHERS IN NORTH EAST, NIGERIA. *Journal of Technical Education and Training*, *8*(1).

Easterlin, R. (1981). Why isn't the whole world developed? *The Journal of Economic History,* 41 *(*1), 1 – 17.

Graham, J, & Kelly, S. (2018). How effective are early grade reading interventions? A review of the evidence. *World Bank Policy Research Working Paper,* 8292*.*

Hamid, Z., Bisschoff, C., & Botha, C. (2015). Work stress and its influence on the work performance of Swaziland teachers in the Swaziland school education system. *Journal of Psychology*, *6*(2), 142-157.

Hanushek, E., & Woessman, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth, 17,* 267–32.

Hwa, Y., Kaffenberger, M. and Silberstein, J. 2020. Aligning Levels of Instruction with Goals and the Needs of Students (ALIGNS): Varied Approaches, Common Principles. RISE Insight Series. 2020/022.

RTI International & Ghana Education Service (2016). *Ghana 2015: Early Grade Reading Assessment and Early Grade Mathematics Assessment*. Eddata II: USAID/Ghana. https://www.sapghana.com/data/documents/2015-EGRA-EGMA_22Nov2016_FINAL.pdf

Gross, J. (2006). *The long term costs of literacy difficulties*: KPMG Foundation.

USAID (2018). *Ghana early grade reading program impact evaluation: Midline report.* Social Impact, Inc.

Yunarti, B. S., Asaloei, S. I., Wula, P., & Werang, B. R. (2020). Stress and performance of elementary school teachers of Southern Papua: A survey approach. *Universal Journal of Educational Research*, *8*(3), 924-930.

APPENDIX

Table A. 1: T-test for remediation practices from the headteacher survey by time

	Midline			Endline		
	Cont.	**Treat.**	**Diff.**	**Cont.**	**Treat.**	**Diff.**
Teachers trained in remediation	64.31	59.85	-4.46	62.13	56.99	-5.14
Discussed best times	57.83	72.84	15.01*	59.28	74.25	14.97**
ASER used to group students	84.78	98.18	13.40*	90.91	99.18	8.27**
HT observed any session	71.74	89.09	17.35*	80.68	88.52	7.84
Grade 1						
Implemented remediation	46.99	61.73	14.74	44.31	68.26	23.95***
More than 1x week	53.85	56.00	2.15	51.35	53.51	2.16
Duration >30min	10.26	8.00	-2.26	10.81	13.16	2.35
# of sections in the term	9.67	12.06	2.39	8.15	9.56	1.41
Grade 2						
Implemented remediation	49.40	60.49	11.10	46.11	67.66	21.56***
More than 1x week	48.78	63.27	14.48	45.45	59.29	13.84
Duration >30min	12.20	6.12	-6.07	12.99	10.62	-2.37
# of sections in the term	8.88	11.96	3.08	8.23	9.84	1.61
Grade 3						
Implemented remediation	50.60	62.96	12.36	48.50	68.26	19.76***
More than 1x week	40.48	62.75	22.27*	43.21	62.28	19.07**
Duration >30min	9.52	9.80	0.28	11.11	13.16	2.05
# of sections in the term	8.55	11.49	2.94	7.89	9.86	1.97

Note: * p<0.1 ** p<0.05 *** p<0.01.

Table A. 2: T-test for remediation practices from the teacher survey by time

	Midline			Endline		
	Cont.	Treat.	Diff.	Cont.	Treat.	Diff.
Grade 2						
Trained in remediation	69.33	85.14	15.80*	90.91	80.72	-10.19
Implemented remediation	37.33	60.81	23.48**	50.65	78.31	27.66***
Written action plan	39.29	62.22	22.94	30.77	66.15	35.38***
More than 1x week	50.00	73.33	23.33*	57.69	64.06	6.37
Duration >30min	7.14	11.11	3.97	15.38	17.19	1.80
ASER used to group students	96.43	100.00	3.57	96.15	98.44	2.28
Very comfortable with activities	75.00	75.56	0.56	73.08	73.44	0.36
Used radio lessons	21.43	28.89	7.46	15.38	34.38	18.99
Grade 3						
Trained in remediation	64.86	80.72	15.86*	86.25	82.93	-3.32
Implemented remediation	41.89	62.65	20.76**	51.25	73.17	21.92**
Written action plan	29.03	53.85	24.81*	29.27	60.00	30.73**
More than 1x week	51.61	63.46	11.85	57.69	60.66	2.96
Duration >30min	9.68	13.46	3.78	19.23	13.11	-6.12
ASER used to group students	96.77	98.08	1.30	100.00	98.36	-1.64
Very comfortable with activities	87.10	82.69	-4.40	61.54	68.85	7.31
Used radio lessons	22.58	15.38	-7.20	15.38	47.54	32.16**

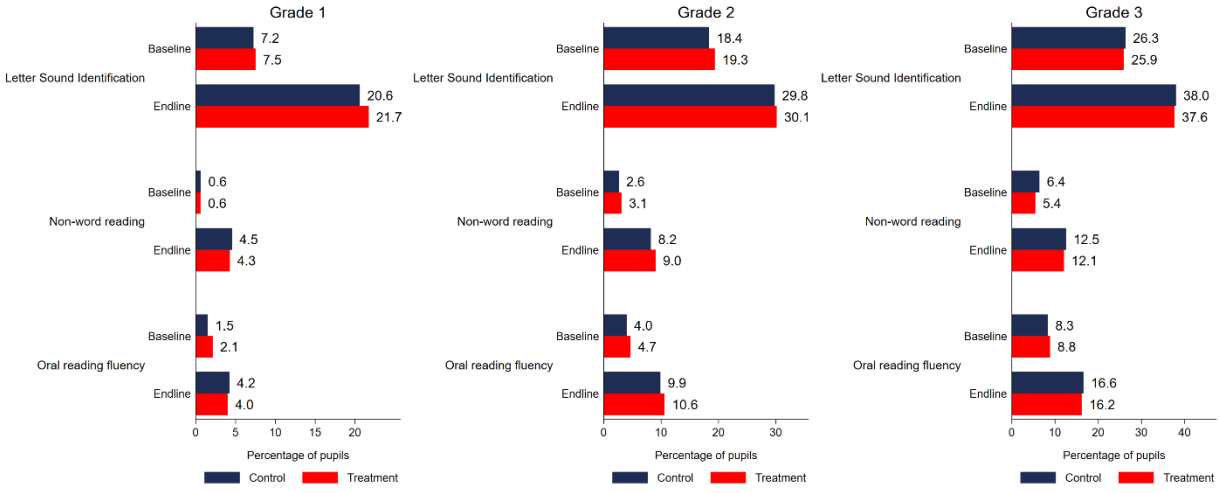
Note: * p<0.1 ** p<0.05 *** p<0.01.

Table A. 3: T-test for teacher practices from the lesson observation survey by treatment status at endline by grade

	Midline			Endline		
	Cont.	**Treat.**	**Diff.**	**Cont.**	**Treat.**	**Diff.**
Grade 2						
Followed steps Teacher Guide	84.29	81.25	-3.04	96.10	89.16	-6.95
Called on boys and girls equally	82.86	76.56	-6.29	88.31	77.11	-11.20
Called on pupils in the back	65.22	54.84	-10.38	72.37	62.96	-9.41
Helped with incorrect answers	82.86	79.69	-3.17	88.31	93.98	5.66
Reference the lesson plan often	74.29	62.50	-11.79	80.52	67.47	-13.05
Teacher pacing was lively	75.71	63.49	-12.22	83.12	76.83	-6.29
Provided positive correction	81.43	79.69	-1.74	93.42	87.80	-5.62
Grade 3						
Followed steps Teacher Guide	94.03	84.51	-9.52	95.00	91.76	-3.24
Called on boys and girls equally	89.55	80.28	-9.27	90.00	78.82	-11.18*
Called on pupils in the back	74.24	55.07	-19.17*	77.22	61.45	-15.77*
Helped with incorrect answers	92.54	80.28	-12.26*	93.75	91.76	-1.99
Reference the lesson plan often	74.63	69.01	-5.61	82.50	65.88	-16.62*
Teacher pacing was lively	74.63	61.97	-12.66	82.50	80.00	-2.50
Provided positive correction	88.06	81.69	-6.37	95.00	94.05	-0.95

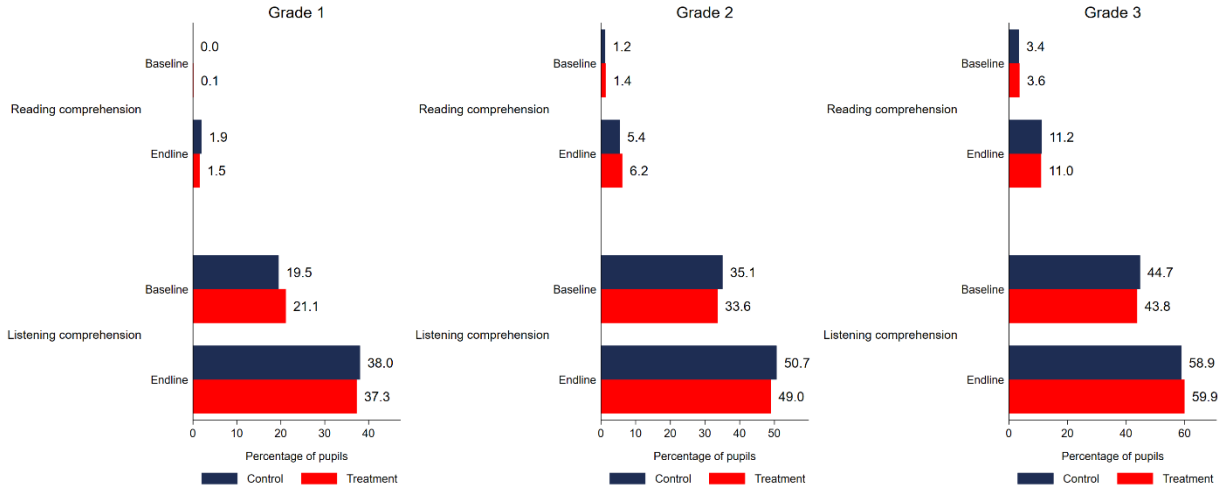
Note: * p<0.1 ** p<0.05 *** p<0.01.

Figure A. 1. Average fluency scores on GLOI timed subtasks, by grade and time



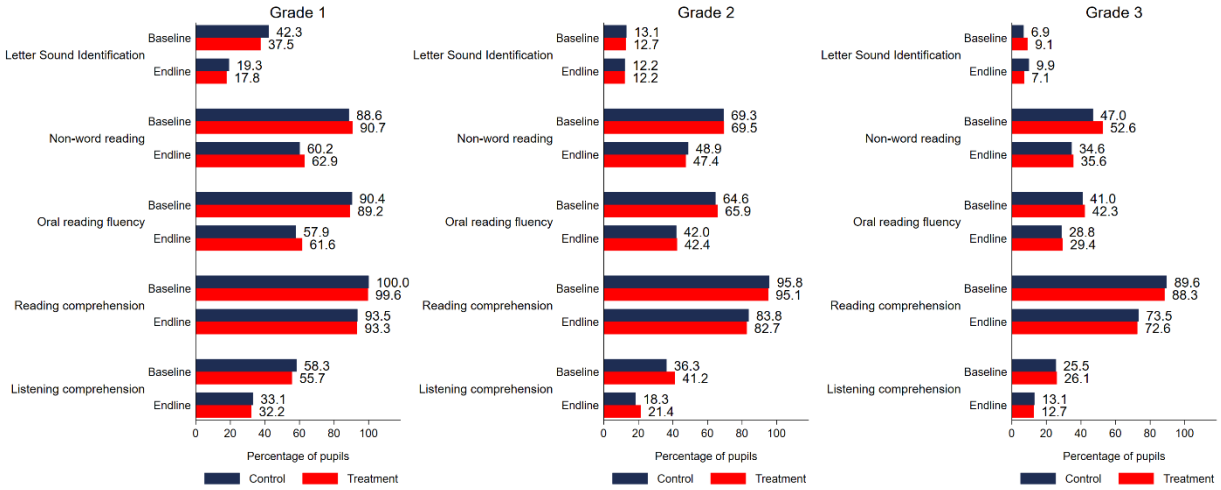
Note: * p<0.1 ** p<0.05 *** p<0.01.

Figure A. 2. Average percent correct on GLOI untimed subtasks, by grade and time



Note: * p<0.1 ** p<0.05 *** p<0.01.

Figure A. 3. Proportion of students who obtain a zero score on GLOI subtasks, by grade and time



Note: * p<0.1 ** p<0.05 *** p<0.01.

Table A. 4. EGRA descriptive statistics for GLOI subtasks by grade and treatment status

	GRADE 1			GRADE 2			GRADE3		
	Cont.	Treat.	Diff.	Cont.	Treat.	Diff.	Cont.	Treat.	Diff.
Timed Scores									
Letter Sound Identification	20.6	21.72	1.12	29.78	30.14	0.36	37.97	37.63	-0.34
Non-word reading	4.55	4.27	-0.28	8.19	9.02	0.83	12.53	12.07	-0.46
Oral reading fluency	4.23	4.01	-0.22	9.85	10.61	0.76	16.57	16.17	-0.4
% Correct									
Reading comprehension	1.95	1.55	-0.4	5.41	6.15	0.74	11.16	10.99	-0.17
Listening comprehension	38.02	37.27	-0.76	50.66	49	-1.65	58.88	59.91	1.04
Zero Scores (%)									
Letter Sound Identification	19.26	17.81	-1.45	12.23	12.18	-0.05	9.89	7.11	-2.78
Non-word reading	60.17	62.88	2.7	48.91	47.44	-1.47	34.6	35.59	0.99
Oral reading fluency	57.92	61.59	3.67	42.01	42.4	0.39	28.81	29.41	0.6
Reading comprehension	93.51	93.35	-0.16	83.84	82.69	-1.15	73.47	72.63	-0.84
Listening comprehension	33.12	32.19	-0.93	18.34	21.37	3.03	13.05	12.72	-0.34

Note: * p<0.1 ** p<0.05 *** p<0.01.

Table A. 5: Treatment effects on EGRA timed and untimed English subtasks by grade – 2SLS estimates

	Letter Name Identification (z-score)	Non-word reading (z-score)	Oral reading fluency (z-score)	Oral vocabulary (%)	Reading comprehension (%)	Listening comprehension (%)
	(1)	(2)	(3)	(4)	(5)	(6)
Effect Grade 1	0.09	0.01	-0.01	-8.21	-2.23	4.81
	(0.31)	(0.21)	(0.16)	(8.31)	(3.00)	(13.70)
N	923	923	919	923	923	923
Effect Grade 2	0.12	-0.04	-0.01	-0.94	5.95	-28.15
	(0.38)	(0.30)	(0.25)	(7.29)	(7.68)	(17.37)
N	926	926	923	926	926	926
Effect Grade 3	0.08	-0.11	-0.47	-0.30	-4.97	-2.38
	(0.31)	(0.32)	(0.34)	(6.29)	(10.36)	(13.50)
N	938	934	934	939	939	939

Note: Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01. Each cell shows effect of a different regression. EGRA timed subtasks were normalized relative to the distribution of the control group for each grade, such that the mean and standard deviation of the control group at each grade is zero and one, respectively. Results for the untimed subtasks showed in percentage. Standard error clustered at the school level showed in parenthesis. Controls for region, gender, and SES were included in all the regressions, as well as control for the school average EGRA score at baseline. Sample size is different for each timed subtask because EGRA timed subtasks with an absolute value higher than 4 standard deviations were coded to missing. The first stage results vary between 0.23-0.26 and are statistically significant at 1% level in all models. F-tests for all the models are strong and statistically significant (F value > 10 and p-value<0.0007).

Table A. 6: Treatment effects on EGRA zero scores for the English subtasks by grade – 2SLS estimates

	Oral vocabulary (%)	Letter Name Identification (%)	Non-word reading (%)	Oral reading fluency (%)	Reading comprehension (%)	Listening comprehension (%)
	(1)	(2)	(3)	(4)	(5)	(6)
Effect Grade 1	-1.55	-1.10	-1.95	-5.23	2.47	-7.26
	(2.83)	(9.98)	(17.35)	(18.56)	(5.20)	(18.35)
N	923	923	923	919	923	923
Effect Grade 2	-3.48	0.82	-1.21	6.58	-14.30	26.89
	(2.26)	(6.49)	(16.97)	(16.58)	(12.40)	(21.23)
N	926	926	926	923	926	926
Effect Grade 3		**-6.74***	-2.29	-12.75	11.22	3.80
		(3.76)	(13.94)	(12.76)	(14.51)	(16.46)
N		**938**	934	934	939	939

Note: Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01. Each cell shows effect of a different regression. Results showed in percentage. Standard error clustered at the school level showed in parenthesis. Controls for region, gender, and SES were included in all the regressions, as well as control for the school average EGRA score at baseline. Sample size is different for each timed subtask because EGRA timed subtasks with an absolute value higher than 4 standard deviations were coded to missing. There are no results for the Oral Vocabulary subtask for Grade 3 because there are no zero scores in this subtask for Grade 3 students at endline. The first stage results vary between 0.23-0.26 and are statistically significant at 1% level in all models. F-tests for all the models are strong and statistically significant (F value > 10 and p-value<0.0007).

Table A. 7: Treatment effects on EGRA timed and untimed GLOI subtasks by grade – 2SLS estimates

	Letter Sound Identification GLOI (z-score)	Non-word reading GLOI (z-score)	Oral reading fluency GLOI (z-score)	Reading comprehension GLOI (%)	Listening comprehension GLOI (%)
	(1)	(2)	(3)	(4)	(5)
Effect Grade 1	0.20	-0.07	-0.09	-1.69	-3.90
	(0.29)	(0.21)	(0.16)	(2.32)	(11.94)
N	923	923	912	923	923
Effect Grade 2	-0.08	0.07	0.07	2.16	-2.98
	(0.33)	(0.31)	(0.30)	(4.96)	(10.98)
N	926	926	924	926	926
Effect Grade 3	-0.02	0.13	-0.16	-0.89	6.28
	(0.31)	(0.30)	(0.35)	(7.03)	(10.93)
N	939	932	931	939	939

Note: Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01. Each cell shows effect of a different regression using the IV approach. EGRA timed subtasks were normalized relative to the distribution of the control group for each grade, such that the mean and standard deviation of the control group at each grade is zero and one, respectively. Results for the untimed subtasks showed in percentage. Standard error clustered at the school level showed in parenthesis. Controls for region, gender, and SES were included in all the regressions, as well as control for the school average EGRA score at baseline. Sample size is different for each timed subtask because EGRA timed subtasks with an absolute value higher than 4 standard deviations were coded to missing. The first stage results vary between 0.24-0.26 and are statistically significant at 1% level in all models. F-tests for all the models are strong and statistically significant (F value > 10 and p-value<0.0007).

Table A. 8: Treatment effects on EGRA zero scores for the GLOI subtasks by grade – 2SLS estimates

	Letter Sound Identification GLOI (%)	Non-word reading GLOI (%)	Oral reading fluency GLOI (%)	Reading comprehension GLOI (%)	Listening comprehension GLOI (%)
	(1)	(2)	(3)	(4)	(5)
Effect Grade 1	0.01	4.36	18.65	-0.15	-1.28
	(15.41)	(16.80)	(18.93)	(7.98)	(16.87)
N	923	923	912	923	923
Effect Grade 2	0.90	-8.23	-4.19	-0.96	3.27
	(12.48)	(18.40)	(17.59)	(12.51)	(12.69)
N	926	926	924	926	926
Effect Grade 3	-15.51	-6.29	0.13	-0.52	-2.30
	(10.14)	(14.64)	(15.44)	(13.52)	(10.68)
N	939	932	931	939	939

Note: Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01. Each cell shows effect of a different regression using the IV approach. Results showed in percentage. Standard error clustered at the school level showed in parenthesis. Controls for region, gender, and SES were included in all the regressions, as well as control for the school average EGRA score at baseline. Sample size is different for each timed subtask because EGRA timed subtasks with an absolute value higher than 4 standard deviations were coded to missing. The first stage results vary between 0.24-0.26 and are statistically significant at 1% level in all models. F-tests for all the models are strong and statistically significant (F value > 10 and p-value<0.0007).

Table A. 9: Heterogeneous treatment effects by gender on EGRA timed and untimed English subtasks– 2SLS estimates

	Letter Name Identification (z-score)	Non-word reading (z-score)	Oral reading fluency (z-score)	Oral vocabulary (%)	Reading comprehension (%)	Listening comprehension (%)
	(1)	(2)	(3)	(4)	(5)	(6)
Effect Girls Grade 1	0.04	-0.1	-0.08	-11.8	-4	-5.34
Effect Boys Grade 1	0.11	0.22	0.13	-1.42	1.78	28.73
Difference Grade 1	-0.07	-0.32	-0.21	-10.38	-5.78	-34.07
	(0.50)	(0.41)	(0.24)	(11.65)	(6.87)	(23.45)
N	923	923	919	923	923	923
Effect Girls Grade 2	0.62	0.27	0.22	10.11	10.84	-45.4
Effect Boys Grade 2	-0.14	-0.23	-0.14	-8.18	3.43	-10.66
Difference Grade 2	0.76	0.50	0.36	**18.29***	7.41	-34.74
	(0.54)	(0.43)	(0.36)	**(10.80)**	(12.94)	(26.21)
N	926	926	923	**926**	926	926
Effect Girls Grade 3	0.24	-0.25	-0.52	-1.52	5.38	10.67
Effect Boys Grade 3	-0.21	-0.01	-0.48	-1.30	-19.98	-20.81
Difference Grade 3	0.45	-0.24	-0.04	-0.22	**25.36***	31.48
	(0.50)	(0.51)	(0.49)	(8.64)	**(14.62)**	(21.53)
N	938	934	934	939	939	939

Note: Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01. Each cell shows effect of a different regression using the IV approach. EGRA timed subtasks were normalized relative to the distribution of the control group for each grade, such that the mean and standard deviation of the control group at each grade is zero and one, respectively. Results for the untimed subtasks shown in percentage. Standard error clustered at the school level shownin parenthesis. Controls for region, gender, and SES were included in all the regressions, as well as control for the school average EGRA score at baseline. Sample size is different for each timed subtask because EGRA timed subtasks with an absolute value higher than 4 standard deviations were coded to missing.

Table A. 10: Heterogeneous treatment effects by SES on EGRA timed and untimed English subtasks– 2SLS estimates

	Letter Name Identification (z-score)	Non-word reading (z-score)	Oral reading fluency (z-score)	Oral vocabulary (%)	Reading comprehension (%)	Listening comprehension (%)
	(1)	(2)	(3)	(4)	(5)	(6)
Effect Low SES Grade 1	0.46	-0.02	-0.06	-1.43	0.98	8.68
Effect High SES Grade 1	-0.13	0.01	0.00	-12.25	-3.88	3.22
Difference Grade 1	0.59	-0.03	-0.06	10.82	4.86	5.46
	(0.52)	(0.31)	(0.23)	(12.75)	(4.94)	(24.85)
N	923	923	919	923	923	923
Effect Low SES Grade 2	0.59	-0.13	-0.22	-1.39	-6.64	-41.5
Effect High SES Grade 2	-0.06	0.08	0.18	1.03	15.40	-16.50
Difference Grade 2	0.65	-0.21	-0.40	-2.42	**-22.04***	-25.00
	(0.56)	(0.42)	(0.34)	(9.81)	**(12.48)**	(25.10)
N	926	926	923	926	**926**	926
Effect Low SES Grade 3	0.61	0.38	-0.1	2.45	-10.86	5.98
Effect High SES Grade 3	-0.22	-0.37	-0.69	-3.14	-3.74	-7.32
Difference Grade 3	0.83	0.75	0.59	5.59	-7.12	13.30
	(0.60)	(0.58)	(0.55)	(11.92)	(16.60)	(25.10)
N	938	934	934	939	939	939

Note: Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01. Each cell shows effect of a different regression using the IV approach. EGRA timed subtasks were normalized relative to the distribution of the control group for each grade, such that the mean and standard deviation of the control group at each grade is zero and one, respectively. Results for the untimed subtasks showed in percentage. Standard error clustered at the school level showed in parenthesis. Controls for region, gender, and SES were included in all the regressions, as well as control for the school average EGRA score at baseline. Sample size is different for each timed subtask because EGRA timed subtasks with an absolute value higher than 4 standard deviations were coded to missing. Students' SES was calculated as a composite based on responses to survey items about the availability of items in student households, including radio, cellphone, telephone, television, bicycle, car, electricity, and light at home to read when it is dark. The presence of more of these items indicated a higher level of SES. The index was calculated using factor analysis and then split into a binary variable where students in the bottom quartile of the index—the lowest 25 percent of index scores—were labeled as having a "low" SES.

Table A. 11: Heterogeneous treatment effects by pre-treatment school ORF score on EGRA timed and untimed English subtasks– 2SLS estimates

	Letter Name Identification (z-score)	Non-word reading (z-score)	Oral reading fluency (z-score)	Oral vocabulary (%)	Reading comprehension (%)	Listening comprehension (%)
	(1)	(2)	(3)	(4)	(5)	(6)
Effect Low ORF Grade 1	0.15	0.01	-0.03	-10.34	0.17	12.25
Effect High ORF Grade 1	0.03	-0.01	-0.00	-7.37	-3.93	1.86
Difference Grade 1	0.12	0.02	-0.03	-2.97	4.10	10.39
	(0.56)	(0.37)	(0.26)	(15.03)	(5.25)	(23.83)
N	919	919	919	919	919	919
Effect Low ORF Grade 2	0.21	0	0.08	3.72	1.37	-29.18
Effect High ORF Grade 2	0.25	0.07	0.02	-1.73	11.40	-22.02
Difference Grade 2	-0.04	-0.07	0.06	5.45	-10.03	-7.16
	(0.68)	(0.53)	(0.42)	(13.41)	(13.04)	(32.55)
N	926	926	923	926	926	926
Effect Low ORF Grade 3	0.84	0.12	0.09	12.67	-10.21	13.03
Effect High ORF Grade 3	-0.10	-0.17	-0.59	-4.68	-3.00	-6.50
Difference Grade 3	0.94	0.29	0.68	17.35	-7.21	19.53
	(1.03)	(0.71)	(0.72)	(23.70)	(19.34)	(41.06)
N	938	934	934	939	939	939

Note: Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01. Each cell shows effect of a different regression using the IV approach. EGRA timed subtasks were normalized relative to the distribution of the control group for each grade, such that the mean and standard deviation of the control group at each grade is zero and one, respectively. Results for the untimed subtasks showed in percentage. Standard error clustered at the school level showed in parenthesis. Controls for region, gender, SES, books at home, and parental engagement (reads at home and help with homework) were included in all the regressions, as well as control for the school average EGRA score at baseline. Sample size is different for each timed subtask because EGRA timed subtasks with an absolute value higher than 4 standard deviations were coded to missing. "Low ORF" is a binary variable that assumes value 1 for schools in the bottom quartile of the English ORF distribution at the school level at baseline, and value 0 for the other three quartiles.