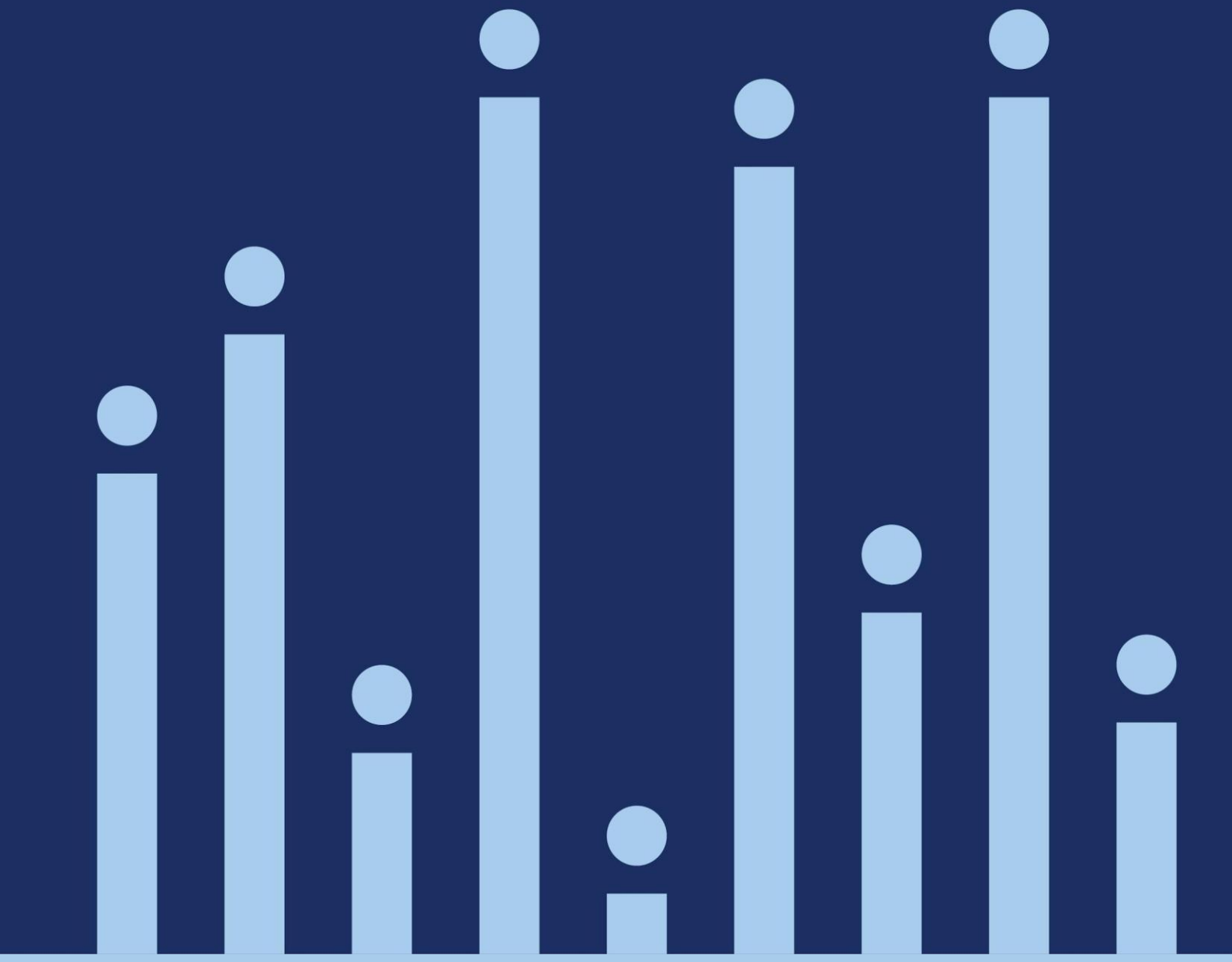


Luminos Program Impact Evaluation: Longitudinal Outcomes from a Randomized Controlled Trial of Accelerated Learning for Out-of-School Children in Liberia

October 31st, 2024

IDinsight

Authors:

Kenna Mokobi: **kenna.mokobi@idinsight.org**

Maadhav Kumar: **maadhav.kumar@idinsight.org**

Mico Rudasingwa: **mico.rudasingwa@idinsight.org**

Jeffery McManus: **jeffery.mcmanus@idinsight.org**

Acknowledgements

We would like to thank Luminos, Q&A Inc. and the Liberian Ministry of Education for their commitment to using data and evidence to improve learning outcomes for children in Liberia; Luminos for coordinating and implementing the accelerated learning program's curriculum in schools, Q&A for the support provided to collect the data and evidence and the Liberian Ministry of Education for being in support of the initiative and guidance in gaining ethical clearance. We'd also like to thank the implementing partners Restoring Our Children's Hope (ROCH) and Liberia Institute for the Promotion of Academic Excellence (LIPACE) for assisting Luminos by implementing the program, providing support for our team on the ground and sharing data for the evaluation.

This study is made possible by the support of the American people through the United States Agency for International Development (USAID). The contents are the responsibility of the Luminos Fund and do not necessarily reflect the views of USAID or the United States Government.

Acronyms and Glossary

EGRA – Early Grade Reading Assessment

EGMA – Early Grade Math Assessment

GSC – Government School Children

IRB – Institutional Review Board

LIPACE - Liberia Institute for the Promotion of Academic Excellence

OOSC – Out-of-School Children

ROCH - Restoring Our Children's Hope

RCT – Randomized Controlled Trial

Contents

Table of Contents

1	**Executive Summary**	**4**
2	**Introduction**	**6**
2.1	Background	6
2.2	Evaluation Process	7
3	**Evaluation Methodology**	**9**
3.1	Research Questions and Outcomes	9
3.2	Randomization	11
3.3	Sample	13
3.4	Data Collection and Attrition	14
4	**Evaluation Results**	**18**
4.1	Enrollment outcomes	18
4.2	EGRA	22
4.3	EGMA	27
4.4	GSC benchmark	30
4.5	Teacher survey	32
5	**Discussion**	**33**
5.1	Limitations	34
5.2	Recommendations:	36
Appendix A: Instruments and Results Tables		**39**
Appendix B: Analytical Modelling		**69**

1 Executive Summary

This report describes the second endline results of an impact evaluation of the Luminos program in Liberia. The second endline was conducted in May-June 2024, which was one year after children in treatment communities graduated from the program.

The Luminos program is an accelerated learning program that teaches children basic reading and numeric skills and supports their socioemotional development. Luminos focuses on children who have never been to school or have been out of school for several years. The Luminos program aims to help out-of-school children (OOSC) rapidly acquire foundational literacy and numeracy skills so that they are prepared to enroll in government schools and effectively engage with the curriculum.

During the first endline, we assessed the literacy, numeracy, and socio-emotional skills of 1,502 out-of-school children (OOSC) across 49 treatment communities that received the Luminos Liberia program in the 2022-23 school year and 49 control communities that did not receive the program. Additionally, we collected data on the same measures for 348 government schoolchildren (GSC) to serve as a benchmark for learning gains. Results from the first endline of the RCT showed large, significant gains across all EGRA and EGMA subtasks (IDinsight, 2023).

For this study, we reassessed 1,372 OOSC and 340 GSC from the original RCT one year after the end of the Luminos program. We recorded the enrollment status of these children to measure the program's effect on enrollment rates and determine if the program was successful in helping children to transition to mainstream school. Additionally, we administered another round of EGRA and EGMA assessments to determine if learning gains persisted one year after the program and to compare treatment effect sizes to the first endline's result.

This study found that 75% of OOSC in treatment communities reported that they were enrolled in government or private schools, indicating that the Luminos program led to a 15 percentage point increase in enrollment rates among OOSC. Large numbers of control OOSC also enrolled between the first and second endline: enrollment rates in the control group went from 32% at the first endline to 60% at the second endline. Although we cannot identify all

the reasons for the large increase in enrollment rates in the control group, we believe that control group enrollment in government schools is likely responsible for comparatively larger growth in learning outcomes in the control group.

The study also found that the Luminos program leads to large improvements in reading and numeracy for OOSC that persist for at least a year after children graduate. Compared to the first endline, effect sizes are 10-15% smaller this year for most EGRA and EGMA subtasks, and effect sizes declined further for the most basic subtasks (letter recognition and number recognition). The slight decline in treatment effect sizes is primarily driven by larger growth in control group scores over the past year compared to the treatment group, likely due to high enrollment rates of control children by the end of year 2. We continue to find that treatment effects on reading subtasks are similar in size for younger vs older children and children who were previously enrolled in school vs dropouts. When comparing boys vs girls, we find that treatment effects are similar across most subtasks, though boys outperform girls across some subtasks.

When we compare government school children's and treatment OOSC's scores, we find that treatment OOSC continue to have higher endline averages for most EGRA subtasks, though GSC growth in EGMA subtasks outpaced treatment OOSC growth between the first and second endline. These results were corroborated by teachers, who reported that Luminos graduates generally compared favorably with other students who did not graduate from the Luminos program, particularly with regards to reading ability and attendance.

The positive and encouraging results from the second endline suggest that the program helps Luminos graduates to enroll and succeed in mainstream schools. Our results point to a few tweaks that Luminos could implement to strengthen post-program support for graduates and further improve long-term learning outcomes, including:

- Ensuring that transition support is reaching link schools and considering options for making this support more flexible so that it follows children who move to other communities
- Continuing to carefully track program graduates and, if possible, monitoring post-program attendance in addition to enrollment
- Providing communication or sensitization with government schoolteachers on the grade placement criteria used by the Implementing Partners (IPs) and Luminos

2 Introduction

IDinsight is a research organization that provides social sector leaders with evidence to improve their programs. The Luminos Fund is an international NGO that helps out-of-school children (OOSC) to catch up to grade level and reintegrate into local government schools. Since 2022, IDinsight has been supporting Luminos in Liberia in collecting evidence on the impact of their program to inform implementation and scale-up decisions. Between 2022 and 2023, IDinsight conducted a randomized controlled trial of the Luminos program to evaluate the effect of the program. We found that the program has a large, positive effect on treatment children's reading and numeracy skills compared to the control group. This report presents the results of the second endline, which involves following up with children in the original RCT one year after treatment children graduated from the program. The goal of this evaluation is to observe what happens to the enrollment outcomes and learning trajectories of OOSC who do or do not go through the Luminos program.

2.1 Background

The Luminos Liberia program (previously known as Second Chance) is a 10-month program that provides intensive, structured lessons on foundational reading and numeracy skills to OOSC. Relative to government schools, the Luminos program uses an approach that is more adaptive to children's learning levels and consists of diverse interaction and play-based learning. Luminos also uses a phonics-based approach to reading instruction. Within each community, Luminos recruits facilitators from the community to teach in Luminos classrooms. Teachers receive 25 days per year of intensive training, as well as additional coaching. Both students and facilitators are intensively monitored throughout the school year by Luminos' supervisors. Luminos does not charge fees, and free daily lunches are provided by Mary's Meals as part of the national school feeding program. Luminos also provides transition support for Luminos graduates by giving 5-10 USD per child to link schools that enroll graduates to cover any fees associated with a child's enrollment (the exact amount of transition support varies by location).

In Liberia, the Luminos program operates in an environment with one of the lowest enrollment rates in the world. The estimated out-of-school rate of primary-age children was 31% in 2023 (UIS, 2023). Liberia also has one of the highest rates of overage children in school. Two in three students in Liberia

are overage for their grade, with an overage enrollment rate for boys of 70% (UNESCO, 2024). Many students in school drop out without acquiring basic skills such as reading and simple arithmetic. In the World Bank's Liberia Human Capital Assessment 2020 report, a child born in Liberia was expected to complete 4.2 years of schooling by the time they turn 18. Adjusting this figure for the quality of instruction, using learning-adjusted years of school (LAYS), a child in Liberia receives the equivalent of 2.2 years of quality schooling, one of the lowest values in the world (World Bank, 2021). Operating in this challenging environment, the Luminos program aims to help OOSC rapidly acquire foundational literacy and numeracy skills so that children are prepared to (re-)enroll in government schools and effectively engage with the curriculum.

2.2 Evaluation Process

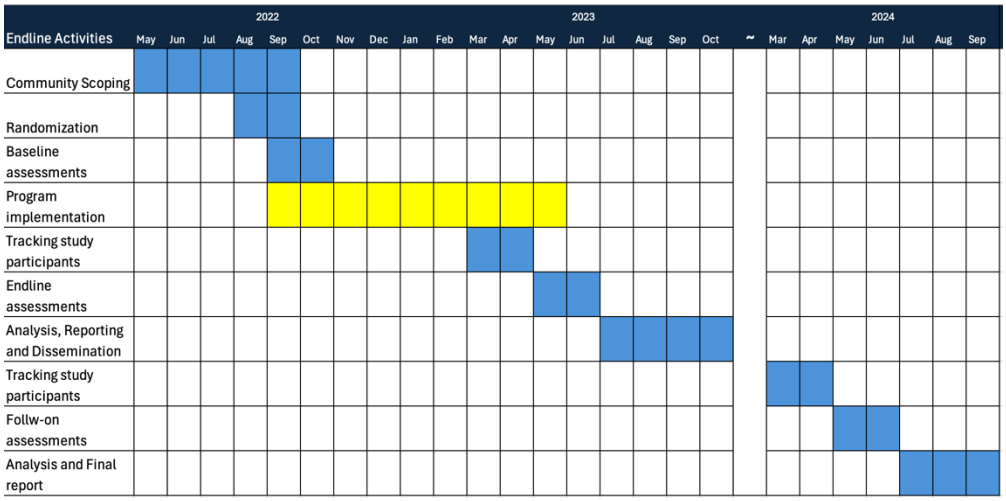
IDinsight partnered with Luminos to conduct a randomized controlled trial (RCT) to assess the impact of the Luminos program on children's learning. The RCT consisted of 3 rounds of data collection: a baseline, first endline immediately after program implementation, and a second endline one year after the end of the program. The sample consisted of a 'treatment group' of 49 communities that received the Luminos program in the 2022-23 school year, and a 'control group' of 49 comparable communities that did not receive the Luminos program.[1] At each round, we also collected data on a sample of government school children from every community to serve as a benchmark for treatment effects.

Luminos and their implementing partners in Liberia, Restoring our Children's Hope (ROCH) and Liberia Institute for the Promotion of Academic Excellence (LIPACE), conducted scoping of suitable communities for the RCT and created lists of eligible children in all communities between May and September 2022. Following this, IDinsight randomized communities into control and treatment groups and conducted baseline data collection between August and October 2022. The program was implemented between September 2022 and June 2023, while IDinsight and Q&A (the survey firm) conducted regular follow-ups, focusing mainly on control communities, to facilitate tracking students during the first endline. IDinsight conducted the first endline

[1] The initial sample consisted of 50 treatment and 50 control communities. However, due to the challenges described in Section 3, we were not able to collect data from one treatment community (Samie Town) and one control community (Gohnzoedua), resulting in an effective sample size of 49 treatment and 49 control communities.

data collection at the end of the program and analyzed the differences between baseline and endline between May 2023 and October 2023. Ahead of the second endline, IDinsight and Q&A conducted regular follow-ups during the 2023/24 academic year to collect location information and updated contact information for children in our study areas. Data for the second endline was collected between May and June 2024 and analysis was conducted between July and September 2024. Figure 1 summarizes the evaluation timeline for all three rounds of data collection.

Figure 1: Timeline of program and evaluation activities



At the first endline, we found large and statistically significant effects of the program on children's reading and numeracy. OOSC in treatment communities were able to read 29 words per minute compared to 7 words per minute for control OOSC on the passage reading subtask. OOSC in treatment communities correctly answered twice as many addition questions and twice as many subtraction questions than control. Effects were similar in size across subgroups such as girls and boys, younger and older children, and children who started with lower baseline learning levels and higher baseline learning levels.

The remainder of this report focuses on the results of the second endline.

3 Evaluation Methodology

To estimate the impact of the Luminos program, IDinsight conducted a randomized controlled trial (RCT) in 98 communities across 5 Liberian counties, namely, Bomi, Grand Cape Mount, Margibi, Montserrado and Bong. We estimated program impact by comparing outcomes for OOSC in treatment communities, where the Luminos program was implemented in the 2022-23 academic year, to children in control communities, where the program was not implemented. We summarize the research questions, randomization, sampling and data processing below.

3.1 Research Questions and Outcomes

Our primary research questions for the second endline were:

A. What is the causal impact of the Luminos program on enrollment rates for out-of-school children in Liberia?

B. Does the causal impact of the Luminos program on learning outcomes for out-of-school children in Liberia persist one year after the Luminos program?

The secondary research questions for the evaluation were:

A. How do treatment effects of the Luminos program vary by subgroups of interest, such as by gender, age, implementing partner, and previous schooling?
B. How do learning gains compare for OOSC in the Luminos program and government school children in the same communities?
C. What are teachers' perceptions of the differences in attendance, performance and attitudes of Luminos program graduates and non-Luminos graduates in treatment communities?

We used the Early Grade Reading Assessment (EGRA) to assess foundational literacy skills, such as recognizing sounds, letters, words, and reading and listening comprehension. We used the Early Grade Mathematics Assessment (EGMA) to assess foundational numeracy skills, such as recognizing numbers, quantity discrimination and arithmetic. Children were assessed at baseline and

both endlines using both EGRA and EGMA. A detailed description of EGRA and EGMA subtasks and assessed competencies can be found in Table 1 below.

Table 1: Description of EGRA and EGMA subtasks

Subtask (Number of Items in Subtask)	Description of Subtask (Timed/Untimed)
EGRA	
Letter name identification (100)	Children attempt to read the names of letters in a 10x10 grid. (Timed-1 minute)
Phonemic awareness (10)	Children are told three words and must identify the word that starts with a different sound. The exercise is repeated 10 times with different words. (Untimed)
Familiar word reading (50)	Children are asked to read familiar/sight words out loud for one minute. The data from this subtask is used to calculate the number of correct words read per minute. (Timed-1 minute)
Non-word reading (50)	Children are asked to decode nonsense words by sounding out letters and their connections. (Timed-1 minute)
Passage reading (60)	Children are asked to read a short passage of 60 words on a topic that is familiar to them. (Timed-1 minute)
Reading comprehension (5)	If the child was able to read the previous passage, then the child is asked five questions about the passage. If the child could not read the previous passage, then this subtask is marked as zero. (Untimed)
Listening comprehension (3)	The enumerator reads a passage to the child and asks the child questions about the text. (Untimed)
EGMA	
Number identification (30)	The child is asked to identify various written numerals between 0 and 999. (Timed-1 minute)

Quantity discrimination (10)	The child compares two numbers and identifies which number is larger. (Untimed)
Missing number (10)	The child is asked to fill in the missing number when given two other numbers in the set. (Untimed)
Addition (15)	The child is asked to solve addition problems with one-digit and two-digit numbers. (Timed-1 minute)
Subtraction (15)	The child is asked to solve subtraction problems with one-digit and two-digit numbers. (Timed-1 minute)
Word problems (5)	The enumerator reads short scenarios featuring simple addition problems to the child. (Untimed)

In addition to our main sample, we sampled government school children in all treatment and control communities and assessed them to provide a benchmark of learning gains. OOSC and GSC are not strictly comparable groups, since OOSC and GSC come from fundamentally different types of households that were either able to send their children to school prior to the Luminos program or not. However, a comparison of learning gains offers suggestive evidence of how learning trajectories differ for children who were offered the Luminos program to children in government schools who did not receive the program.

For the second endline, we also measured teachers' perception of Luminos program graduates' literacy skills and numeracy skills, as well as other academic skills such as attendance, confidence, resilience, and motivation. The teacher survey questionnaire is included in appendix A1. In each of our treatment communities, we attempted to survey a teacher who reported that they were aware of the Luminos program and had taught a class with Luminos program graduates.

3.2 Randomization

The implementing partners (IPs), LIPACE and ROCH, identified 115 communities that met Luminos' eligibility criteria for the program. A community was considered eligible if it had support for the program from leaders and parents within the community, had at least 30 OOSC (i.e., enough to fill one class) who were 8-14 years old, was located a maximum of 3 hours

from the main road and had an available physical space to run the program.[2] The IPs identified eligible OOSCs in each of the 115 communities. Approximately half of these communities had more than 60 OOSC, making them eligible for two 30-child classes. One community had enough children to run three classes, while the remainder had enough children to run one class.

Luminos and IDinsight selected 100 of the 115 communities for the evaluation at baseline, and IDinsight randomized them to treatment and control groups. Due to agreements between Luminos and the IPs about the number of classes each IP had to operate in the 2022-23 academic year, the treatment group plus non-sampled communities had to include 55 anticipated classrooms for LIPACE and 50 anticipated classrooms for ROCH.

To balance the allocation to treatment and control across each IP and geographically, we stratified communities according to IP, anticipated number of classes per community, and county.[3] The final community sample consisted of ten strata (IP x # classes x county).[4] The probability of being treated varies across strata but is constant within each stratum. We include strata fixed effects and probability weights in our analysis to account for the randomization process. Figure 2 shows the location of treatment and control communities.

[2] For the Luminos program, children are considered to be out-of-school if they have not attended school in the last 2 years.

[3] We would have ideally stratified on baseline scores as well. However, due to the government school calendar and program calendar, there would not have been sufficient time for the IPs to prepare for program launch after baseline data collection.

[4] There were 2 IPs, communities had one, two or three expected classes, and there were 5 counties, though the counties mostly differed between IPs.

Figure 2: Location of treatment and control communities

3.3 Sample

For sampling OOSC, Luminos and its implementing partners shared a list of roughly 35 children per study community who appeared to be eligible for the program. We aimed to sample 20 OOSC per community. We stratified the random selection of OOSC using binary indicators for whether a child was above/below 10 years old, using age data collected during scoping, and whether the child was female. Given our power calculations, our target sample was 2,000 OOSC from 100 communities. However, during in-person verification of program eligibility at baseline, we discovered that many children were not eligible for the program due to their age or recent school enrollment status. Additionally, many children were involved in traditional school activities, which prevented our enumerators from conducting assessments with them.[5] After broadening the eligibility criteria to allow for

[5] Traditional schools, within Liberia often known as "bush schools", are temporary, informal schools in which children receive some introduction about and often initiation into the traditional cultures in an area. Sessions are

children who reported being in school within the last two years, the sample consisted of 1,745 OOSC from 98 communities. One treatment (Samie Town) and one control community (Gohnzoedua) had no OOSC available for assessment at baseline and were consequently dropped from the study.[6]

For the second endline, rather than only reassessing the children seen at the first endline (1,502 OOSC and 348 GSC), we sought to reassess all the children surveyed at baseline to maximize statistical power (1,745 OOSC and 445 GSC).

3.4 Data Collection and Attrition

Data collection for the second endline started on May 29th, 2024. We aimed to complete data collection before July, when the Ministry of Education planned to close schools. As advised by Q&A and Luminos, locating and assessing children immediately before and after the official end of school is challenging due to varying school closure dates and activities like parades and exams, which affect certain grades' attendance. Despite our data collection schedule, we still encountered significant challenges in locating students, which led to extended data collection and mop-ups. Q&A recruited 34 enumerators and 5 supervisors to administer learning assessments for 4 weeks following an enumerator training session led by IDinsight and Q&A.

During our second endline, we successfully assessed:

- 710 treatment OOSC out of the 902 assessed at baseline (21.3% attrition)
- 662 control OOSC out of the 843 assessed at baseline (21.5% attrition)
- 157 GSC in treatment communities out of the 211 assessed at baseline (25.6% attrition)
- 183 GSC in control communities out of the 234 assessed at baseline (21.8% attrition)

restricted to leaders and students and researchers are strongly discouraged from interrupting or entering traditional schools, making the children inaccessible.

[6] In the case of Samie Town, traditional schools were active at the onset of the program, making the children unavailable for at least several weeks. Implementors of the program did not operate a program and barred enumerators from entering the community in September. While enumerators were able to sample OOSC children in both communities, neither was ultimately included in our sample.

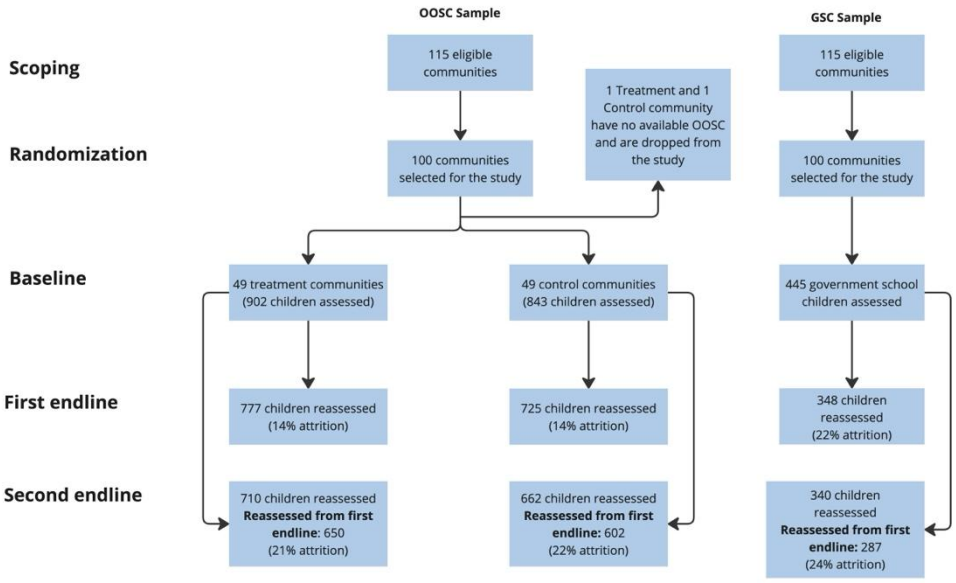


Figure 3 summarizes the sampling, randomization, and assessment process.

Figure 3: Sampling flow chart

As at the first endline, the differences in attrition between treatment and control communities at the second endline are not statistically significant for OOSC (p-value = 0.93) or GSC (p-value = 0.35).

We successfully reassessed a large percentage of the children we assessed in the first two rounds of the study. At the second endline we reassessed 650 OOSC (out of 777) in treatment communities and 602 OOSC (out of 725) in control communities who were assessed at the baseline and first endline. The rest of our second endline OOSC sample consists of 60 treatment OOSC (out of 118) and 60 control OOSC (out of 120) who were assessed at baseline but not at the first endline. Similarly, we successfully reassessed 137 GSC (out of 167) in treatment communities and 150 GSC (out of 181) in control communities who were assessed at the baseline and first endline. The rest of the GSC sample is composed of 20 GSC (out of 44) in treatment communities and 33 GSC (out of 53) in control communities who were assessed at baseline but not at the first endline.

Out of the 373 OOSC assessed at baseline that we were not able to reassess at the second endline, the most common reasons for not being reassessed were: (i) the child relocated and we could not find updated contact information; (ii) the child couldn't be reached due to lack of network in the area; (iii) contact information on file was no longer correct or couldn't be used to locate the child.

As shown in Table 2 below, 68.4% of OOSC we could not survey had relocated with insufficient contact information to locate them. Q&A reported that a frequent occurrence in these communities was that one individual in the community owned a cell phone and it was shared between community members. As a result, if the community member who owned this phone relocated or was not available on the date of the survey, enumerators did not have a phone number they could use to reach the families of OOSC.

Table 2: Reasons not surveyed (OOSC)

Reason respondent was not surveyed	Percent	N
Relocated with no contact info	68.4%	255
Contact can't be reached due to lack of network	11.5%	43
Contact information on file is incorrect	9.3%	35
Relocated to unreachable location, including other countries	4.6%	17
Child did not consent or assent to be assessed	2.7%	10
Other (child is sick, cultural activities prevent assessment, technical difficulties, etc.)	2.9%	11
Child is deceased	0.5%	2

Note: N = 373, the number of OOSC assessed at baseline that were not reassessed at the second endline

During data collection, Q&A field supervisors and IDinsight staff conducted spot checks of field teams. IDinsight also ran daily checks of the raw data to track attrition and survey quality. To locate children in our sample, Q&A field supervisors contacted community leaders in advance to try to inform parents that enumerators would be entering the community to conduct another round of learning assessments and request parents to exclude children from farming activities for that day. To minimize attrition, teams on the ground also conducted revisits to many of the communities with smaller teams to increase the number of sampled children who were assessed.

Table 3 reports the averages, across treatment and control groups, of baseline characteristics used for stratification and baseline scores on EGRA and EGMA subtasks. The balance table shows that across the three data collection rounds the sample generally maintains the same balance on baseline characteristics that was seen at baseline. Consistent with our baseline report (IDinsight, 2022), across all three rounds we find that the largest difference between treatment and control is whether a child reported being in school previously at baseline. We believe that this difference is largely explained by differences in how children in the two groups self-reported prior enrollment status. Baseline data was collected after we randomized and after IPs had started community mobilization in treatment communities, including messaging about student eligibility criteria. We think that children in treatment communities may have been more likely to misreport prior enrollment status if they (or their parents) worried that reporting prior enrollment, even if it was more than two years ago, would make them ineligible for the Luminos Program. EGRA and EGMA scores are generally well balanced across the three rounds with some small differences in scores that do not follow a consistent pattern.

Table 3: Baseline characteristics for children assessed at baseline, first endline, and second endline

	Baseline (N = 1,745)			First Endline (N = 1,502)			Second Endline (N = 1, 372)		
General Characteristics	**C**	**T**	**p-value**	**C**	**T**	**p-value**	**C**	**T**	**p-value**
Age (years)	10.19	10.41	0.02**	10.17	10.37	0.04**	10.20	10.32	0.24
Female (1 = female)	0.47	0.45	0.50	0.46	0.45	0.93	0.42	0.44	0.54
In school before (1 = yes)	0.48	0.33	0.00***	0.49	0.32	0.00***	0.50	0.32	0.00***
EGRA+EGMA	**C**	**T**	**p-value**	**C**	**T**	**p-value**	**C**	**T**	**p-value**
Reading orientation	1.62	1.74	0.07*	1.62	1.76	0.05**	1.69	1.76	0.29
Letter name id (100)	28.25	33.01	0.00***	28.72	34.14	0.00***	29.33	33.58	0.01***
Phonemics (10)	2.95	3.03	0.44	2.95	3.06	0.36	3.00	3.05	0.66
Familiar word reading (50)	2.11	1.72	0.16	2.08	1.76	0.27	2.25	1.74	0.11
Familiar word per minute	2.34	1.82	0.11	2.33	1.85	0.18	2.52	1.82	0.07*
Non-word reading (50)	0.25	0.04	0.01***	0.25	0.05	0.03**	0.32	0.05	0.01***
Passage reading	3.50	3.48	0.97	3.41	3.53	0.84	3.89	3.45	0.49
Passage score per minute	3.69	4.01	0.62	3.56	4.04	0.50	4.12	3.95	0.82
Reading comp (5)	0.19	0.16	0.38	0.18	0.16	0.41	0.20	0.15	0.19
Listening comp (3)	1.48	1.59	0.05*	1.49	1.58	0.13	1.49	1.58	0.11
Number id (30)	6.05	6.43	0.32	6.02	6.64	0.14	6.36	6.52	0.72
Quantity discrim (10)	2.39	2.79	0.01***	2.40	2.89	0.00***	2.47	2.81	0.04**

Missing number (10)	0.29	0.28	0.87	0.28	0.28	0.87	0.29	0.29	0.97
Addition (15)	2.12	2.53	0.02**	2.10	2.62	0.01***	2.17	2.55	0.05*
Subtraction (15)	1.08	1.25	0.13	1.06	1.28	0.06*	1.10	1.26	0.20
Word problems (4)	1.21	1.28	0.23	1.22	1.27	0.35	1.24	1.28	0.48

Following data collection, we processed and analyzed the data in Stata. We implemented the analytical models that we pre-specified for this RCT on the public AEA RCT registry.[7] We focus our analysis on 'intent-to-treat' (ITT) estimates, which reflect the impact of the program on the average child who is eligible to join the program in communities that were offered that program. Given the high enrollment rates at the first endline of treatment OOSC in the Luminos program and no non-compliance in the control group, we do not present treatment effect on the treated (ToT) estimates in this study. At the first endline, these estimates were very similar to ITT effects.

4 Evaluation Results

At the second endline, we find that the program led to a 15-percentage point greater increase in enrollment in government and private schools in the treatment group compared to the control group. This modest difference between treatment and control OOSC in enrollment rates is partially explained by the large increase in enrollment in the control group. Control group enrollment increased from 32% to 60% between first and second endline.

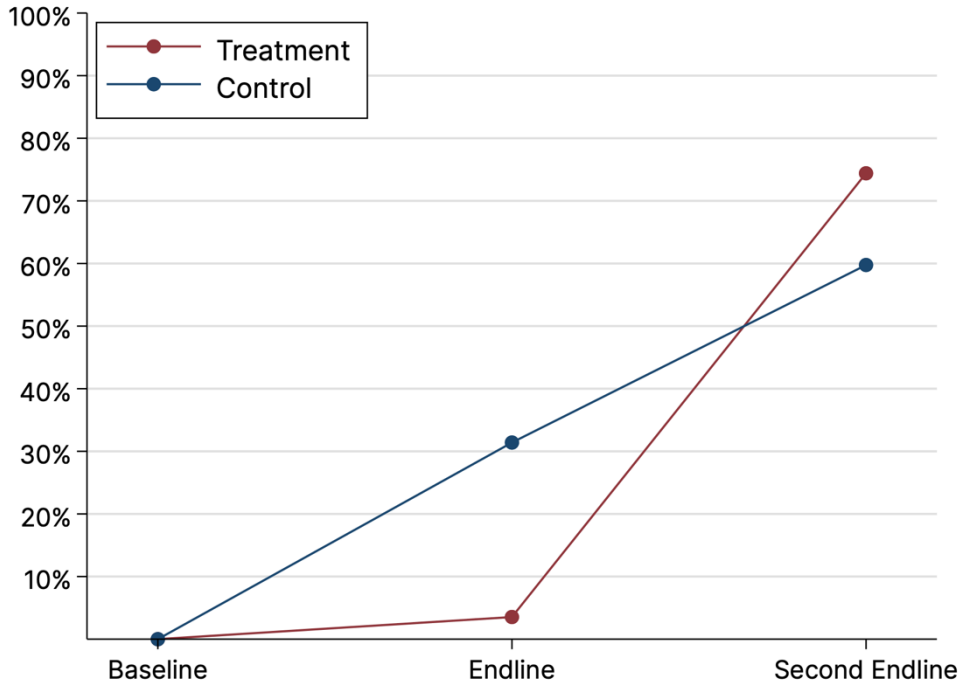
In terms of learning outcomes, the large effects from the first endline persist, but their overall magnitudes are slightly smaller, which is explained by faster learning growth in the control group compared to the first endline. GSC students continued to experience learning gains, with larger gains in the second endline compared to the first endline. The following section describes the key results from the second endline.

4.1 Enrollment outcomes

[7] https://www.socialscienceregistry.org/trials/10649

At the second endline, the enrollment rate of OOSC in government or private schools was 75% for treatment OOSC and 60% for control OOSC. When we control for baseline covariates, we find that the program led to a 15-percentage point increase in enrollment rates in the treatment group compared to the control group.

Figure 4: Enrollment in government or private school over time



Note: Enrollment at baseline has been coded as 0% for both groups, per the sampling frame of out-of-school children constructed by IPs. At baseline, we asked children to report if they were currently enrolled in school. However, we believe that this question was misinterpreted, and many treatment children reported they were enrolled in school when they meant they were enrolled in the Luminos program. We clarified the wording of the question in the first endline and second endline instrument.

One factor driving the modest change in enrollment rates is the large increase in enrollment rates in the control group. Between the first and second endline, control children's reported enrollment rate increased from around 32% to 60%. This result was unexpected, and the reasons for it remain unclear. We cannot rule out the possibility that additional factors may have driven the large increase in enrollment in the control group, such the presence of other education interventions that benefitted control group children or that OOSC who were not offered entry into the Luminos program preferred to enroll in government school rather than remain out-of-school.

While we are not able to determine with confidence what is driving the large increase in enrollment in the control group, we did ask children why they enrolled this year, and we report the results in Table 4. Around 71% of enrolled control children and 62% of enrolled treatment children reported that they enrolled because their family found the money for school fees. The next most reported reason for enrollment was that the child was motivated to enroll (24% of enrolled control children and 34% of enrolled treatment children).

Table 4: Reasons a child enrolled in government or private school

Reason the child enrolled	Control	Treatment
My family found the money for school fees	70.7%	62.2%
I was motivated to enroll/re-enroll	23.7%	34.3%
Parents/Relatives told me to attend	6.6%	3.2%
I passed the entrance exams	5.3%	11.0%
Luminos encouraged me to enroll or provided transition support for this school	0.3%*	2.8%
Completed the Luminos program / Luminos program no longer available	0.0%	1.7%
Other (specify)	6.8%	3.6%
Refused or don't know	1.5%	3.4%

Note: N = 930. The total number of children in the second endline sample who reported they were enrolled was 932. However, for two observations, this question was skipped due to entry error. Children could select more than one option, so percentages may not add up to 100%.

* One control group child reported that they enrolled in the Luminos program. We believe this could be explained by a control group child relocating to a community where Luminos was offering the program.

Additionally, we recorded if a child had enrolled in school earlier in the 2023/2024 school year but then dropped out. Across treatment and control, 77 OOSC children (T = 44, C. = 33) reported that they dropped out of school between the first and second endline. The most frequently reported reason for dropping out was that the child's family did not have the money for school fees (63% C, 61% T). Other reasons were reported less frequently, but included relocation (9% C, 9% T), family emergency (9% C, 2% T), and lack of a uniform (6% C, 2% T).

Our enrollment findings are based on children self-reporting enrollment status during data collection. To ensure that these reports were reliable, we attempted to confirm the enrollment status of all children in our OOSC sample with at least one faculty member. For control OOSC who reported enrollment, we were able to confirm enrollment with one faculty member for 90% and with two faculty members for 89%. For treatment OOSC who reported enrollment, we were able to confirm enrollment with one faculty member for 93% and with two faculty members for 92%. In no cases did a faculty member contradict the enrollment status reported by a child. For the remaining children whose status we could not confirm, we were unable to speak to a faculty member because school was not in session, the teacher was unavailable, or the child was surveyed at their home or another location far away from the link school.[8] Given the high rate of agreement between child self-reports and faculty reports, we believe that the child self-reports are reliable even when we were unable to speak with a faculty member.

In addition to our second endline data collection, Luminos followed up with program graduates between January to February 2024 to ask about their enrollment status. To compare records, we merged this data with the data in our sample. We successfully matched Luminos enrollment records with 678 out of the 710 treatment OOSC in our second endline sample.

Overall, we found broad agreement between our enrollment data and Luminos' enrollment records, with explanations for most discrepancies. When we compare Luminos' enrollment records with our data, we find that for 448 of the 678 treatment OOSC in our merged second endline sample (66%), Luminos' enrollment records are consistent with our records. For 165 of the 678 treatment OOSC (24%), Luminos recorded that the child was not enrolled in January/February 2024, but we found that the child was enrolled during our second endline data collection. This likely indicates that the child enrolled later in the year after Luminos' checks. For the remaining 65 of the 678 treatment OOSC in our merged second endline sample (10%), Luminos recorded that they were enrolled in school in January/February 2024, but our data found that they were not enrolled in government or private school. For 14 of these 65 children, the child reported being enrolled in school earlier in the year but dropped out. The most common reason listed for these children dropping out was a lack of funds to pay school fees. For the remaining children (51 children,

[8] Due to logistical and budgetary constraints, if a child was not surveyed close to a link school, the enumerator proceeded with the interview for teacher verification to be done manually at a later date as part of mop-ups. However, in some cases teacher verification was de-prioritized to prioritize mop-ups of OOSC.

or 8% of the merged second endline sample), our records appear to be contradict Luminos' data.[9]

4.2 EGRA

Treatment effects persist for treatment OOSC one year after their graduation from the program across all EGRA subtasks. Table 4 shows the average scores for the control group and treatment group OOSC at baseline and the second endline and the ITT estimates from the first and second endline. Compared to last year, ITT estimates are 10-15% smaller this year for most subtasks, though the decline is larger for the more basic subtasks such as letter recognition. Smaller treatment effect sizes are explained by relatively more growth in average scores seen in the control group. When we measure treatment effects by considering the varying difficulty of EGRA sub-tasks using item response theory (IRT), we find that treatment effects in SDs remain positive and significant at the 1% level, but they are also smaller when compared to the results of the first endline. The full IRT results can be found in the appendix (Table B1), along with a description of the methodology used

For example, at the second endline, treatment OOSC children were able to read 21.60 more WPM than control OOSC children, down slightly from the difference of 22.04 WPM during the first endline. However, in terms of learning levels, at the second endline treatment OOSC read an average of 34.30 WPM, up from 28.74 WPM at the first endline. The reason treatment effects stayed steady was that the growth in treatment OOSC scores was coupled with growth in control OOSC scores, with a first-endline average of 7.34 WPM and a second-endline average of 15.31 WPM for control OOSC.

Table 5: Raw EGRA scores at baseline (BL) and second endline (EL2), with treatment effects from first (EL1) and second endlines

EGRA	Control		Treatment		Treatment Effects	
Subtask	BL	EL2	BL	EL2	ITT EL1	ITT EL2
Letter name id (100)	28.94	52.38	32.03	80.47	37.12	23.38

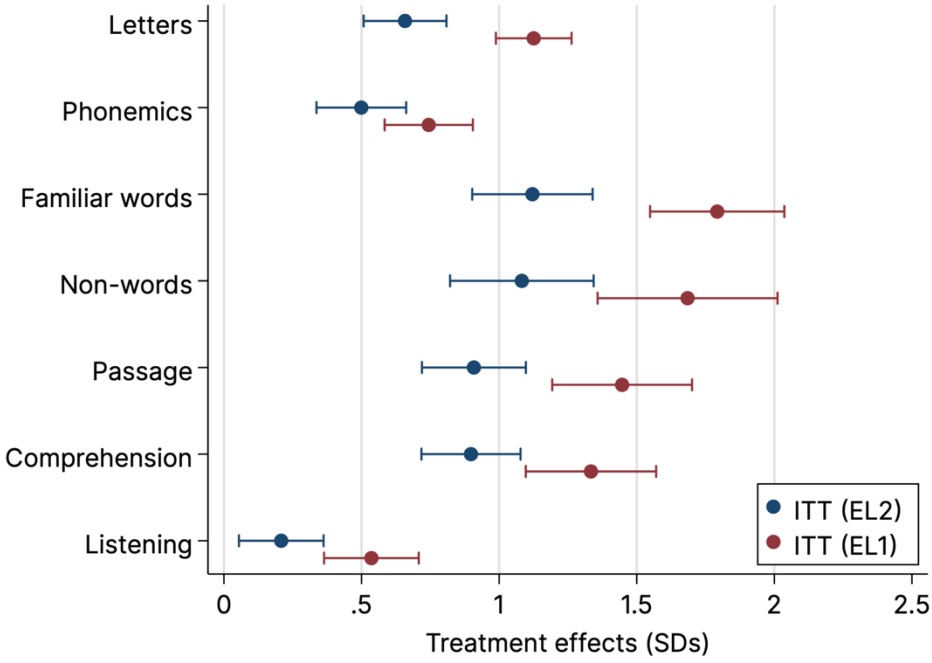
[9] For the 32 children in our treatment EL2 sample who we could not match with Luminos' records, 20 reported they were enrolled and 12 reported that they were not enrolled during our second endline data collection.

Phonemics (10)	3.00	4.36	3.00	5.42	1.67	1.17
Familiar word reading (50)	2.21	8.62	1.61	23.75	18.15	16.04
Familiar word per minute	2.04	10.41	1.58	26.68	18.56	17.18
Non-word reading (50)	0.33	1.56	0.05	7.93	7.90	7.10
Passage reading (60)	3.87	12.21	3.25	29.05	21.28	18.92
Passage per minute	3.89	15.31	3.21	34.30	22.04	21.60
Reading comp (5)	0.20	0.81	0.15	1.94	1.31	1.29
Listening comp (3)	1.48	1.79	1.55	2.05	0.57	0.23

Note: Baseline averages differ slightly from the baseline averages at the first endline due to different sample compositions between the first and second endline.

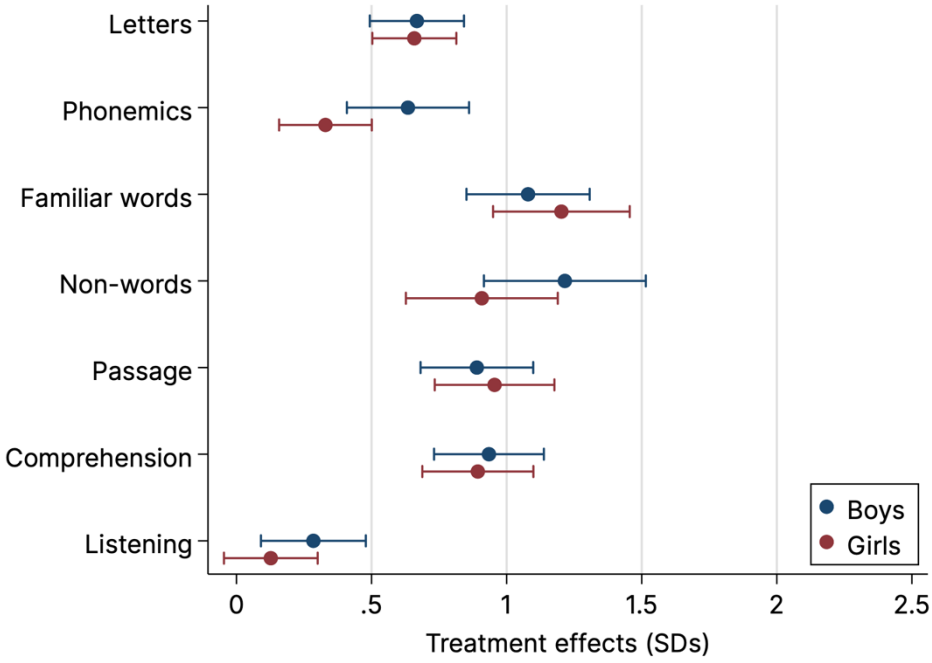
Figure 5 shows the standardized treatment effects for all EGRA subtasks along with 95% confidence intervals. Standardized effect sizes larger than 0.5 SD are generally considered large for education programs. Treatment effects at the first endline ranged from 0.5SD to 1.6SD, while at the second endline, they ranged from 0.21SD to 1.12SD. Treatment effects for all subtasks except listening comprehension were greater than 0.5SD, while effects for listening comprehension were more modest (0.21SD). Although standardized effect sizes have fallen comparably more than raw effect sizes, this is a result of higher growth and much higher variance in control scores in the second endline. For example, the raw ITT estimates for the passage reading task fell 2% from the first endline to the second endline (from 22.04 WPM to 21.60 WPM). But the ITT in terms of SDs fell 47% (from 1.12 SD to 0.76 SD). When we look at the standard deviation of scores for ORF by treatment group, we see that the standard deviation for ORF for the control group increased 45% from the first to second endline (from 19.36 to 28.06 WPM).

Figure 5: Standardized treatment effects per EGRA subtasks



When we investigate if there are any differences across subgroups, we continue to see that treatment effects on reading subtasks are similar in size for younger vs older children and children who were previously enrolled in school vs dropouts. When comparing boys vs girls, we find that treatment effects are similar across most subtasks (Figure 6). However, boys outperform girls across some subtasks. We find that boys can identify 0.67 more phonemes (p-value = 0.01) and 1.68 more non-words per minute (p-value = 0.04). The full sub-group analysis results are included in appendix A3 and appendix A4 (Gender-wise).

Figure 6: Standardized treatment effects per EGRA subtasks (by gender) at the second endline



Oral reading fluency deep dive:

In Figure 7a and 7b, we explore the program's effect on oral reading fluency (ORF) (or the number of words per minute that the child was able to read in the passage reading subtask) in more detail. Figure 5a shows the average scores at baseline, first endline, and the second endline. In Figure 5a, we see that both treatment and control OOSC see growth in their scores between the first and second endline. Control OOSC grow more between the first and second endlines than between baseline and the first endline, and they grow slightly more than treatment OOSC between the first and second endlines. However, the growth in both groups between the first and second endline is not as large as the growth seen in the treatment group between baseline and the first endline.

In Figure 7b, we distinguish between control and treatment OOSC who reported being enrolled in school or unenrolled at each endline (we consider all treatment children "unenrolled" in the first year). In general, enrolled OOSC, for both treatment and control, had much higher second endline average scores compared to the average first endline score. Treatment OOSC who did not enroll in the second endline experienced regression in scores. However, even with this regression, non-enrolled treatment children still had slightly higher scores at the end of the second year than enrolled control OOSC. Treatment children who enrolled in school in the second endline experienced similar growth in scores as control children who enrolled, though GSC growth was slightly larger than both groups.

Figure 7: Average ORF over time

Figure 7a:

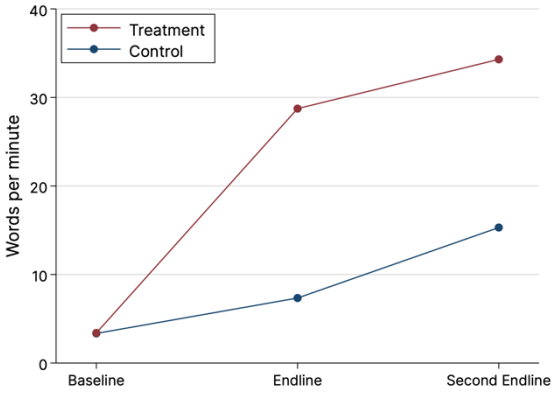
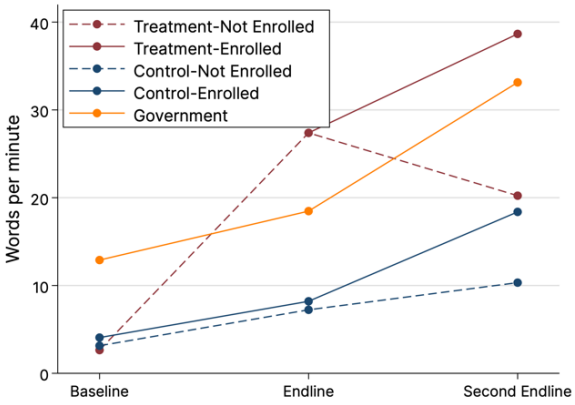


Figure 7b:



In Figure 8a and 8b, we show the distribution of treatment and control children's scores at the first endline and the second endline, respectively. The histograms show that the percentage of control OOSC who scored 0 WPM decreased by more than 15%. Although scores still follow a similar trend across both endlines for control OOSC, at the second endline control OOSC appear more spread out across the distribution. For treatment OOSC, the distribution of scores begins to take on a more pronounced "U" shape, with a smaller percentage of children scoring in the mid-range and a higher proportion in the left and right tails, corroborating our finding that for this subtask the treatment group appears to be split into those who continued to solidify their reading skills and those who saw losses in their reading ability.

Figure 8: Distribution of ORF scores

Figure 8a: First Endline

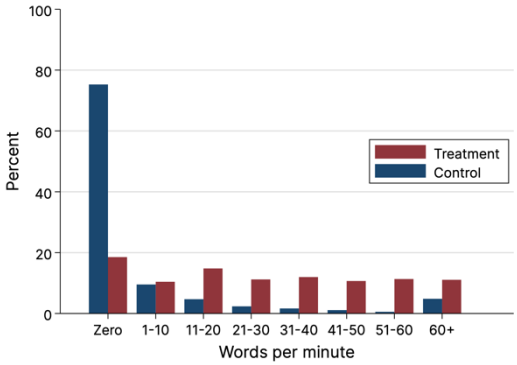
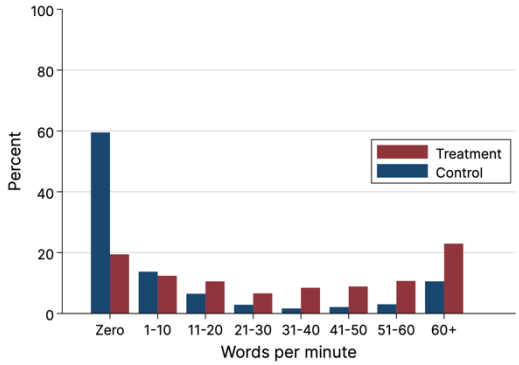


Figure 8b: Second Endline



4.3 EGMA

Similar to EGRA, treatment effects persisted across EGMA subtasks at the second endline. However, we observe marginally smaller gains at the second endline compared to the treatment effects at the first endline across all subtasks. At the second endline, treatment OOSC children were able to complete 2.22 more addition problems than control OOSC, down from 2.63 more correct addition problems at the first endline. In terms of raw averages, at the second endline treatment OOSC answered an average of 7.21 addition problems correctly, a slight increase over the first endline average of 6.18 addition problems. Control OOSC children saw a similar growth in performance, with the average number of correct addition problems increasing from 3.16 to 4.89 between the first and second endline. As with EGRA, when we account for the varying difficulty of EGMA sub-tasks using item response theory (IRT), we find that treatment effects in SDs remain positive and significant at the 1% level but are smaller when compared to the first endline results (Appendix B1) .

Table 6: Raw EGMA scores at baseline and second endline, with treatment effects from first and second endlines

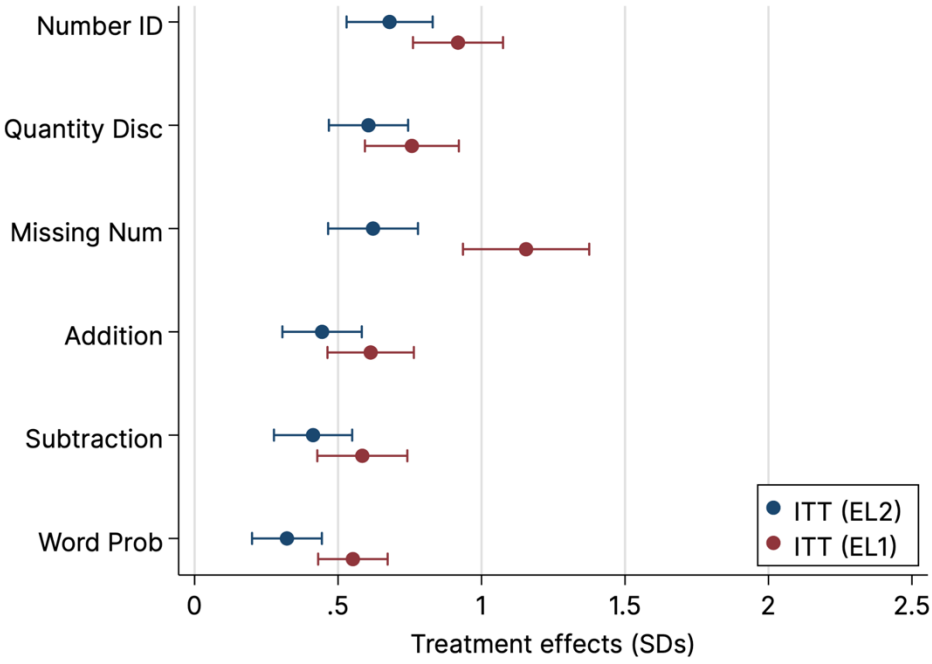
EGMA	Control		Treatment		Treatment Effects	
Subtask	BL	EL2	BL	EL2	ITT EL1	ITT EL2
Number id (50)	6.29	12.70	6.25	19.79	8.25	6.90
Quantity discrim (10)	2.46	4.52	2.75	6.69	2.47	2.06
Missing number (10)	0.28	1.36	0.28	2.52	1.84	1.38
Addition (15)	2.15	4.89	2.46	7.21	2.63	2.22
Subtraction (15)	1.07	3.07	1.22	4.70	1.88	1.66
Word problems (4)	1.23	2.13	1.26	2.60	0.75	0.48

Note: Baseline averages differ slightly from the baseline averages at the first endline due to different sample compositions between the first and second endline.

Figure 9 shows the standardized treatment effects per EGMA subtask along with 95% confidence intervals for the first and second endline. Similar to EGRA standardized effect sizes, EGMA standardized effect sizes show larger

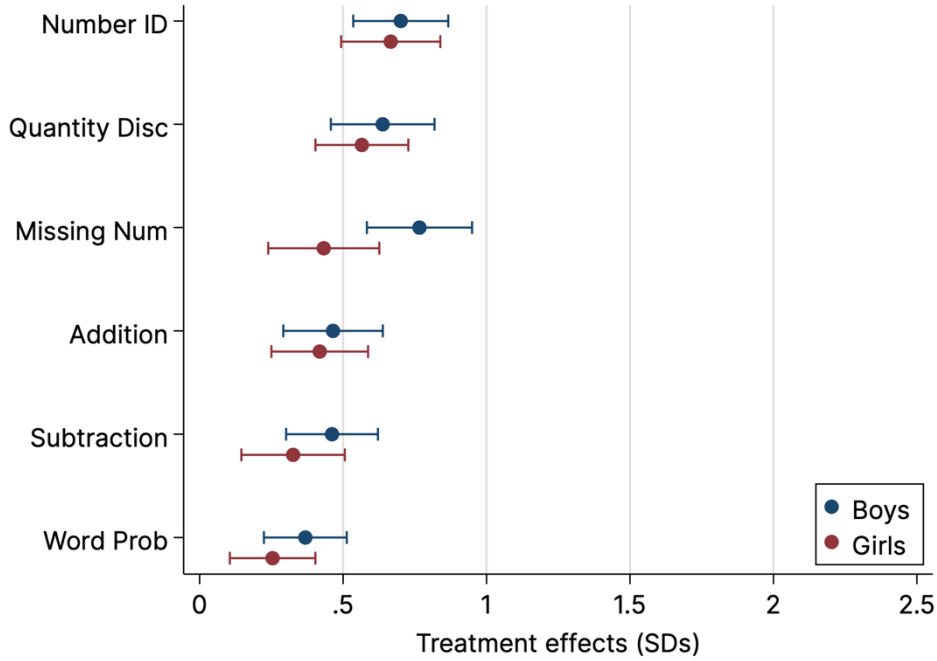
decreases when compared to raw averages. Treatment effects at the initial endline ranged from 0.55SD to 1.16SD, while at the second endline, they ranged from 0.32SD to .68SD. As with EGRA, this can partly be explained by the higher variance in scores seen in the control group. For the addition task, the control group standard deviation increased by 16% from the first to second endline (from 4.3 to 5.0).

Figure 9: Standardized effect sizes for EGMA subtasks



As with EGRA subtasks, we investigate if there are any differences across subgroups for EGMA subtasks. We continue to see that treatment effects on numeracy are similar across subgroups. In Figure 10, we present standardized treatment effects for EGMA subtasks across gender. We find that boys can identify 0.75 more missing numbers (p-value = 0.00), answer 0.84 more subtraction questions correctly (p-value = 0.02) and answer 0.92 more word problems (p-value = 0.06). Across all other subtasks, treatment effects are similar between boys and girls. The full sub-group analysis results are included in appendix A3 and appendix A4 (Gender-wise)

Figure 10: Standardized treatment effects per EGMA subtasks (by gender) at the second endline



Addition deep dive:

In Figures 11a and 11b we do a deep dive of the addition subtask task. In both figures we see similar trends as in the ORF deep dive. On average, treatment and control OOSC saw increases in the addition scores between the first endline and second endline, with control group children seeing slightly larger gains. When these trends are broken down by enrollment status, we see that unenrolled treatment OOSC regressed between the first and second endline and are at a similar level as enrolled control OOSC by the second endline. Enrolled treatment OOSC see a larger increase in scores over the first endline average. In the control group, enrolled OOSC see larger improvements in scores than non-enrolled control OOSC.

Figure 11: Average number of correct addition problems over time

Figure 11a:

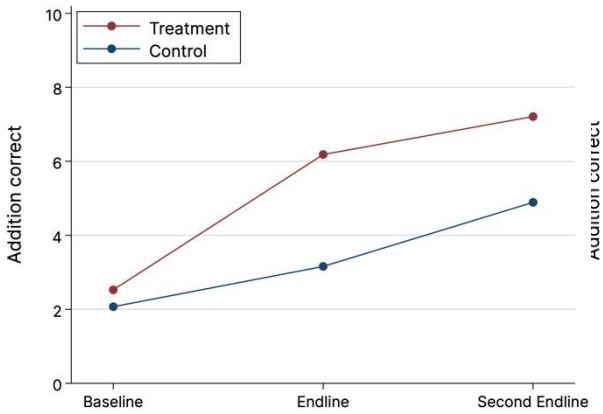
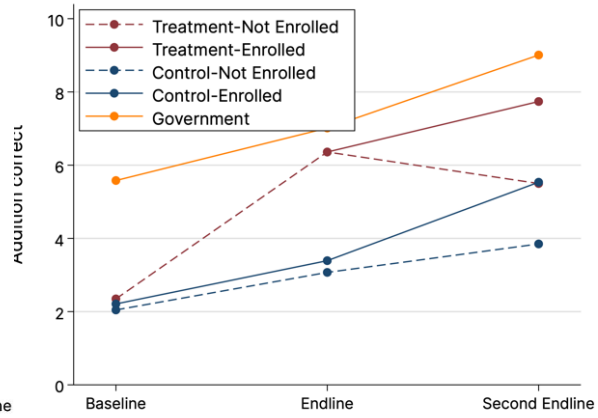


Figure 11b:



In Figures 12a and 12b, we plot the distribution of scores at the first and second endline, respectively. As in the ORF task, few control OOSC score zero points on the addition subtasks. The distribution of control OOSC scores is also more evenly distributed across the distribution of scores. For treatment OOSC, the patterns are less pronounced. We see a slight decrease in the number of treatment OOSC who scored zero points and slight increases in the bins for "9-10" and "13-14" points. However, we do not see large movements towards the right tail of the distribution for treatment OOSC.

Figure 12: Distribution of addition scores

Figure 12a: First Endline

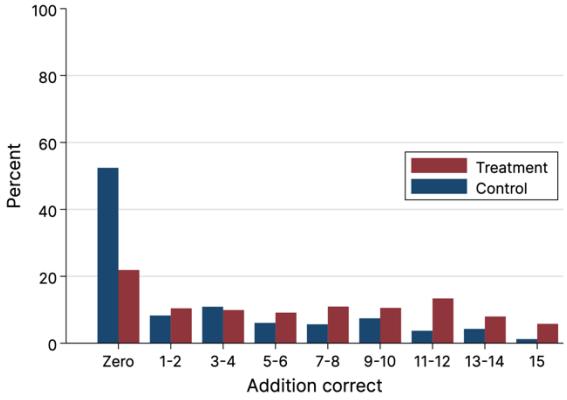
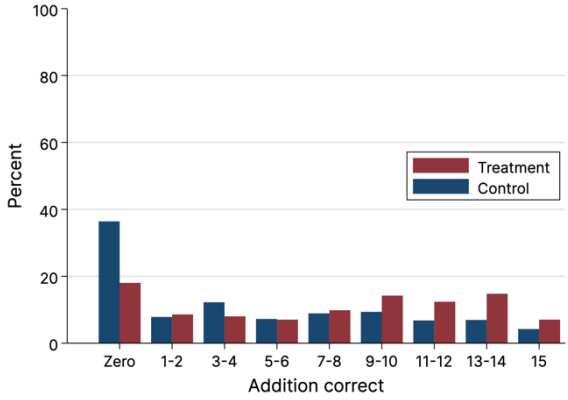


Figure 12b: Second Endline



4.4 GSC benchmark

In Tables 7 and 8, we present the baseline and second endline averages of the treatment group, control group and for GSC students for EGRA and EGMA, respectively. Overall, compared to the growth in scores between the baseline and the first endline, we observed similar growth in GSC scores between the first and second endline for most subtasks but larger growth for passage reading, addition and missing number tasks.

For passage reading, at the second endline GSC children were able to read 33.13 WPM in the oral reading fluency task, up from 18.47 WPM at the first endline. For the addition subtask, GSC children were able to answer 9.01

addition problems correctly at the second endline, up from 7.01 correct addition problems at the first endline. When we compare government school children's and treatment OOSC's scores, we see that treatment OOSC continue to have higher endline averages for most EGRA subtasks, except for letter name identification and listening comprehension where GSC were able to identify around one more letter name and 0.04 more correct listening comprehension questions. However, for EGMA, GSC growth in the second endline outpaced treatment OOSC growth. For example, at the second endline GSC were able to answer 9.01 addition problems correctly compared to treatment OOSC's average of 7.21 addition problems.

Table 7: EGRA GSC scores

EGRA	Control		Treatment		Treatment Effects		GSC	
Subtask	BL	EL2	BL	EL2	ITT EL1	ITT EL2	BL	EL2
Letter id 100)	28.94	52.38	32.03	80.47	37.12	23.38	63.52	81.16
Phonemics (10)	3.00	4.36	3.00	5.42	1.67	1.17	4.02	5.39
Familiar word (50)	2.21	8.62	1.61	23.75	18.15	16.04	7.18	19.16
Familiar word per min	2.04	10.41	1.58	26.68	18.56	17.18	7.07	22.71
Non-word reading (50)	0.33	1.56	0.05	7.93	7.90	7.10	0.86	3.69
Passage reading (60)	3.87	12.21	3.25	29.05	21.28	18.92	13.15	26.20
Passage per min	3.89	15.31	3.21	34.30	22.04	21.60	13.21	33.13
Reading comp (5)	0.20	0.81	0.15	1.94	1.31	1.29	0.73	1.79
Listening comp (3)	1.48	1.79	1.55	2.05	0.57	0.23	1.91	2.08

Note: Baseline averages differ slightly from the baseline averages at the first endline due to different sample compositions between the first and second endline.

Table 8: EGMA GSC scores

EGMA	Control		Treatment		Treatment Effects		GSC	
Subtask	BL	EL2	BL	EL2	ITT EL1	ITT EL2	BL	EL2
Number id (50)	6.29	12.70	6.25	19.79	8.25	6.90	14.63	21.49
Quantity discrim (10)	2.46	4.52	2.75	6.69	2.47	2.06	5.19	7.12
Missing number (10)	0.28	1.36	0.28	2.52	1.84	1.38	1.00	2.95

Addition (15)	2.15	4.89	2.46	7.21	2.63	2.22	5.65	9.01
Subtraction (15)	1.07	3.07	1.22	4.70	1.88	1.66	3.02	6.10
Word problems (5)	1.23	2.13	1.26	2.60	0.75	0.48	2.00	3.13

Note: Baseline averages differ slightly from the baseline averages at the first endline due to different sample compositions between the first and second endline.

4.5 Teacher survey

During the second endline data collection, we attempted to survey one teacher in each treatment community about their perceptions of Luminos program graduates in their classrooms. We successfully surveyed one teacher in 45 out of the 49 treatment communities in our sample.

Overall, teachers perceive that Luminos graduates compare favorably to non-Luminos peers, particularly on reading ability. When asked about Luminos program graduates' reading ability compared to their peers, 73% of the teachers surveyed said that Luminos program graduates had somewhat better or much better reading ability compared to their peers. Around 16% reported that Luminos graduates had similar reading ability to their peers, and 11% reported they had worse or much worse reading ability. For numeracy, 44% said that Luminos program graduates had somewhat better or much better numeracy skills than non-program graduates in their class. Around 27% said that Luminos graduates had similar numeracy skills to their peers, and 29% reported that Luminos graduates had worse or much worse numeracy skills.

Luminos graduates also compare favorably to non-Luminos students on other academic and social-emotional skills, including:

- Attendance: 62% of teachers said that Luminos program graduates had somewhat higher or much higher attendance, 27% said they had similar attendance, and 11% said they had worse or much worse attendance
- Expressiveness: 60% of teachers said that Luminos program graduates were more expressive, 24% said they were similarly expressive, and 16% said non-program graduates were more expressive.
- Confidence: 53% of teachers said that Luminos program graduates were more confident than their peers, 36% said that Luminos graduates had similar confidence levels, and 11% said they had lower confidence.

- Resilience: 51% of teachers said that Luminos graduates were more resilient than their peers, 38% said they had similar levels of resilience and 11% saying they had less resilience.
- Motivation: 49% of teachers said that Luminos program graduates were more motivated than their peers, 44% said they had similar levels of motivation, and 6% said they were less motivated.
- Cooperation: 45% of teachers said that Luminos program graduates were more cooperative than their peers, 44% said they had similar levels of cooperation, and 9% said they were less cooperative.

We also asked teachers if they had any additional comments on the differences between Luminos graduates and non-Luminos graduates, or about the Luminos program in their community. More than half of the teachers surveyed (60%) had positive things to say about the differences between Luminos graduates and non-Luminos graduates, and two-thirds (67%) had positive things to say about the Luminos program in the community. A number of the responses reiterated that Luminos graduates had better attendance and resilience than their peers. Other responses highlighted how teachers noticed other socio-emotional improvements. For example, one teacher reported that "*the [Luminos] program has impacted not only the children but the community because the children now are reading and able to do different things on their own.*" A teacher reported that Luminos graduates "*are always attentive in class*", while another reported that Luminos graduates are "*more focus[ed] than the other kids.*" This could point to the Luminos program contributing to improvements in several socio-emotional skills, such as independence and focus. A number of teachers also expressed that they were happy with what the Luminos program had done in their community and asked that Luminos return to their community to help other OOSC.

Some of the qualitative responses from teachers highlighted areas of improvement for the program, which we discuss in more detail in the next section. Some teachers highlighted that Luminos graduates in their community have struggled with the material in their assigned grade or that Luminos graduates have been dropping out of school due to a lack of school fees. In the next section, we discuss tweaks to the program that Luminos could make to address these concerns.

5 Discussion

The first endline of this RCT found that the Luminos program had large, positive, and statistically significant effects on OOSC's foundational reading and numeracy skills. The objective of this second endline was to determine whether the program's effects persist one year after the program and whether Luminos graduates successfully transition to government schools.

The second endline shows that the Luminos program has a modest but positive and statistically significant effect on enrollment rates. OOSC in treatment communities were 15 percentage points more likely to be enrolled in school one year after the program compared to control OOSC. However, we also observed that there was a large growth in enrollment in the control group between the first and second endline, from 32% to 60%. Treatment effects of the program on reading and numeracy persist one year after the program across all the foundational reading and numeracy skills assessed. However, the size of the treatment effects is 10-15% smaller at the second endline compared to first endline, which is largely explained by the growth seen in the control group and the benefits that appear to be associated with control OOSC's enrollment in school. Treatment OOSC who successfully enrolled in school had significantly higher reading and numeracy scores at the second endline compared to first endline, whereas treatment OOSC who did not enroll in school regressed in their reading and numeracy levels.

Our results did not find substantial heterogeneity at the subgroup level on gender, age, implementing partner, or previous schooling. There were some slight differences noted between boys and girls, where boys achieved higher results on most subtasks at the second endline, as seen at the first endline, with statistically significant differences for the phoneme task, non-familiar word reading task, missing number task and subtraction task.

5.1 Limitations

Attrition

Attrition continued to be a challenge during the second as at the first endline. Our teams encountered many of the same issues, including phone numbers being unreachable, the absence of phone network coverage in some areas, and the dispersion of sampled households. A particular challenge during the second endline was the itinerancy of our sample. Many OOSC relocated between the first and second endline, with some traveling to other counties in Liberia or to neighboring countries where we could not assess them. Other notable

challenges included unforeseen school closures, ongoing traditional school activities, and government school activities, such as the West African Senior Schools Certificate Examination (WASSCE) that limited our survey firm's ability to find certain children.

We anticipated many of these challenges and made efforts to minimize attrition. First, we sought to offset attrition from the first endline by attempting to resurvey all 1,745 OOSC assessed at baseline, as opposed to just the 1,502 OOSC children assessed at the first endline. Second, Q&A conducted monthly check-ins with control communities to update contact information, while Luminos tracked children in treatment communities. Third, Q&A made multiple revisits ("mop-ups") to communities to try and reach as many sampled children as possible. As a result of these efforts, we limited attrition to 21% of OOSC and 24% of GSC.

Although attrition rates were balanced between treatment and control children, as a robustness check, we ran our ITT analysis with inverse-probability weights (IPW) for attrition. For the IPW analysis, treatment effects stay positive and significant at the 1% level across all subtasks. Estimates of treatment effects are very close to the estimate without adjusting for attrition. For example, our IPW estimates indicate that treatment children could identify 21.69 more familiar words than control OOSC children. The full set of results are included in appendix A5.

Limited visibility into program implementation challenges

During data collection, we received reports from Q&A that some households refused consent to the second endline assessment because they did not receive transition support to cover school fees (see Table 2: Reasons not surveyed). In the teacher survey, a teacher reported that they "*hope [Luminos] will come [back] because most of the children in the community are not coming to school because of school fees*." Another teacher reported that "*most Luminos students are dropping out of school*." Despite these reported challenges from teachers, we observe in the data that most treatment children in these communities did enroll in school. However, these reports from government schoolteachers and Q&A might point to some implementation challenges with the transition support that Luminos aimed to provide all graduates. Luminos' implementing partners (ROCH and LIPACE) provided the transition support to the link school based on how many children the IPs confirmed were attending the link schools in each community. Luminos would provide the funds to the IPs for the transition support based on IP reports of how many children they confirmed were enrolled. If there were errors in attendance records or a child moved

between schools in between IP checks, this may have led to the transition support not reaching its intended recipients. We are unable to identify which children in our sample did or did not receive the transition support because transition support records are aggregated and not tied to individual children. However, we believe this may have impacted the size of the treatment effects on enrollment.

Luminos highlighted that one implementing partner experienced some notable challenges with providing the transition support. When we test for differences in treatment effects on enrollment by implementing partner, we see that the treatment effects for the IP with notable implementation challenges appear slightly smaller (~3%), but this difference in treatment effects is not statistically significant ($p = 0.77$). Therefore, it appears that the sustainability of treatment effects did not differ between IPs.

Another challenge that we heard about was that some children may have been placed into inappropriate grades. Currently, Luminos makes placement decisions by comparing students' literacy and numeracy results on placement tests to national standards. Additionally, many schools may still administer their own entrance exams before placing students in classes. However, in the teacher survey, one teacher claimed that: *"[Luminos] promote children to the class they are not fit for."* Another teacher requested that the Luminos program reevaluate how children are promoted. They claim that "*most of [the children] who were promoted are causing us, the other teachers in the government school, too much work and hard time to catch up because of the lack of reading and writing ability of these students."* This study did not seek to evaluate the effect of possible mismatches in skills between Luminos graduates and the classrooms they ultimately attend and how these came about. However, grade placement may impact enrollment rates because mismatches in skill level and grade placement could lead to drop-out due to failing class examinations or becoming discouraged due to overly challenging material.

5.2 Recommendations:

The Luminos program leads to large improvements in reading and numeracy for OOSC that persist for at least a year after children graduate. As Luminos seeks to scale the program, we have a few recommendations for how Luminos can build on this success to further increase the impact of the program:

- **Ensure that transition support is reaching link schools and possibly provide more resources to IPs to ensure the successful enrollment of program graduates into mainstream school:** OOSC in treatment communities who enrolled into school improved learning outcomes between the first year and second year, with similar growth in reading and numeracy ability as their GSC peers. OOSC who failed to enroll or dropped out regressed. Enrollment is key to future learning outcomes for OOSC.

  There are likely many factors that prevent some program graduates from enrolling in school, and some factors may be more tractable than others. The most common reason why OOSC reported not enrolling or dropping out between the first and second endline (reported by 58% of non-enrolled OOSC) was a lack of funds to pay school fees. This suggests that Luminos may be able to do more to ensure that link schools receive transition support for all program graduates. Perhaps there are also ways to make the transition support more flexible, so that schools other than the link school (such as private schools or government schools in other communities) can receive the stipend. We found that out of the OOSC children in our sample who were enrolled at the second endline, 22% were enrolled in private school. Additionally, as mentioned in earlier sections of this report, relocation was a common occurrence for the children in our sample. If this transition support follows the child even when they move to other communities, it may remove another barrier to enrollment.

  Some children reported not enrolling in school in order to support their families with income-generating activities. In other cases the link school closed down or did not have enough teaching staff. In such cases it may be more challenging for Luminos to help children to enroll. Perhaps Luminos could explore if these children could be referred to other, complementary programs in their area to support their learning or follow up with those children the following year to see if schooling options in their communities have improved.

- **Continue to carefully track program graduates:** As Luminos considers scaling up the program to more communities, we recommend that Luminos continue to carefully track program graduates to identify their outcomes after graduation. This monitoring could include following up with the parents and guardians of program graduates to determine if they have successfully received the transition support provided or identify any other implementation challenges as they arise. It could also include a more frequent collection of location and enrollment information, given how itinerant this population of children

appears to be, to target Luminos' post-graduation support better. This monitoring could include phone surveys with caregivers and/or school staff to determine what challenges children may be encountering and identify any other needs that may be keeping children from enrolling and staying in school.

If possible, it would be valuable to not only track enrollment but also attendance, since ultimately what matters is whether children are attending classes. If link schools have incomplete systems for recording attendance, perhaps Luminos could offer capacity building to help improve link school record-keeping.

- **Provide communication or sensitization with government schoolteachers on the criteria used to place Luminos graduates in their post-transition grade:** As discussed above, some teachers reported that they believe Luminos graduates might be being placed in grades above their current level. Currently, Luminos and IPs conduct assessments to determine a graduate's grade placement. Luminos or the IPs may consider sharing the criteria for determining a graduate's grade placement with government schoolteachers. This could help teachers become aware of the skills that were assessed to show the child's readiness for the grade. Luminos could also consider asking teachers for their input on the grade placement criteria or to provide ongoing feedback during the school year on how graduates are faring with the material at their assigned grade level.

Appendix A: Instruments and Results Tables

Appendix A1: Teacher Survey Questionnaire

[Administer consent protocols]

1. What is your name?
2. What is the name of your school?
3. What is your role in the school?
    a. Principal
    b. Teacher
    c. Teaching Assistant
    d. Other (please specify)

4. For how many years have you worked at the school?

5. Are you aware of the Luminos program?

   ☐ Yes

   ☐ No

[If yes to 5, proceed]

6. Over the 2023/2024 school year, have you taught a class with children who previously participated in the Luminos Program?

   ☐ Yes

   ☐ No

[If yes to 6, proceed]

7. How does the average Luminos program graduate compare with other students in your class on reading ability?

   a. Much poorer

b. Somewhat poorer

c. Similar performance

d. Somewhat better

e. Much better

8. How does the average Luminos program graduate compare with other students in your class on numeric ability?

a. Much poorer

b. Somewhat poorer

c. Similar performance

d. Somewhat better

e. Much better

9. How does the average Luminos program graduate compare with other students on other academic skills, besides reading and numeracy?

a. Much poorer

b. Somewhat poorer

c. Similar performance

d. Somewhat better

e. Much better

10. I would now like to ask you questions about the behavior of Luminos program graduates and other students [i.e., non-Luminos program graduates] in your class.

	Question	**Response Options**
1	When comparing the attendance rates of Luminos program graduates and other students [i.e. non-Luminos graduates], for the current school year, which of the following statements is most accurate for your class?	1. Luminos program graduates have had much higher attendance rates 2. Luminos program graduates have had somewhat higher attendance rates

		3. Luminos program graduates and non-Luminos students have had similar attendance rates 4. Non-Luminos students have had somewhat higher attendance rates 5. Non-Luminos students have had much higher attendance rates
2	When comparing the confidence of Luminos program graduates and other students [i.e. non-Luminos graduates], for the current school year which of the following statements is most accurate for your class?	1. Luminos program graduates are much more confident 2. Luminos program graduates are somewhat

		more confident 3. Luminos program graduates and non-Luminos students are similarly confident 4. Non-Luminos students are somewhat more confident 5. Non-Luminos students are much more confident
3	When comparing the resilience of Luminos program graduates and other students [i.e. non-Luminos graduates], for the current school year which of the following statements is most accurate for your class?	1. Luminos program graduates are much more resilient 2. Luminos program graduates are somewhat

		more resilient 3. Luminos program graduates and non-Luminos are similarly resilient 4. Non-Luminos students are somewhat more resilient 5. Non-Luminos students are much more resilient
4	When comparing the motivation of Luminos program graduates and other students [i.e. non-Luminos graduates], for the current school year which of the following statements is most accurate for your class?	1. Luminos program graduates are much more motivated 2. Luminos program graduates are somewhat

		more motivated 3. Luminos program graduates and non-Luminos students are similarly motivated 4. Non-Luminos students are somewhat more motivated 5. Non-Luminos students have are much more motivated
5	When comparing how outspoken Luminos program graduates and other students [i.e. non-Luminos graduates] are, for the current school year which of the following statements is most accurate for your class?	1. Luminos program graduates are much more outspoken

		2. Luminos program graduates are somewhat more outspoken 3. Luminos program graduates and non-Luminos students are similarly outspoken 4. Non-Luminos students are somewhat more outspoken 5. Non-Luminos students are much more outspoken
6	When comparing how cooperative Luminos program graduates and other students [i.e. non-Luminos graduates]	1. Luminos program graduates are much more

	are, for the current school year which of the following statements is most accurate for your class?	cooperative 2. Luminos program graduates are somewhat more cooperative 3. Luminos program graduates and non-Luminos students are similarly cooperative 4. Non-Luminos students are somewhat more cooperative 5. Non-Luminos students are much more cooperative

11. Do you have any other comments about how Luminos program students differ from their peers?

12. Do you have any other comments about the Luminos program in your community?

Appendix A3: Full Results Tables

EGRA	Control			Treatment			Treatment Effects			GSC		
Subtask	BL Avg	EL1 Avg (N = 725)	EL2 Avg (N = 662)	BL Avg	EL1 Avg (N = 777)	EL2 Avg (N = 710)	ITT EL1	ITT EL2	ITT EL2 (SDs)	BL Avg (N = 340)	EL1 Avg (N = 348)	EL2 Avg (N = 340)
Letter name identification (100)	28.94	38.01	52.38	32.03	80.84	80.47	37.12	23.38	0.66	63.52	72.99	81.16
Letter name identification per min	29.11	38.80	54.99	32.23	84.19	86.68	39.10	26.24	0.66	63.44	74.85	86.92
Phonemics (10)	3.00	3.46	4.36	3.00	5.13	5.42	1.67	1.17	0.50	4.02	4.60	5.39
Familiar word reading (50)	2.21	4.43	8.62	1.61	22.24	23.75	18.15	16.04	1.12	7.18	13.23	19.16
Familiar word reading per minute	2.04	4.94	10.41	1.58	23.12	26.68	18.56	17.18	0.90	7.07	14.25	22.71
Non-word reading (50)	0.33	0.85	1.56	0.05	8.19	7.93	7.90	7.10	1.08	0.86	2.59	3.69
Non-word reading per minute	0.13	0.92	1.59	0.05	8.03	7.83	7.67	6.95	1.11	0.37	2.62	3.70
Passage reading (60)	3.87	6.05	12.21	3.25	26.62	29.05	21.28	18.92	0.91	13.15	16.38	26.20

Oral reading fluency per minute	3.89	7.34	15.31	3.21	28.74	34.30	22.04	21.60	0.76	13.21	18.47	33.13
Reading comprehension (5)	0.20	0.39	0.81	0.15	1.63	1.94	1.31	1.29	0.90	0.73	1.24	1.79
Listening comprehension (3)	1.48	1.38	1.79	1.55	1.97	2.05	0.57	0.23	0.21	1.91	1.97	2.08

Note: Baseline averages differ slightly from the baseline averages at the first endline due to different sample compositions between the first and second endline.

EGRA Sub-group analysis	Control					Treatment					Treatment Effects		
Subtask	BL Avg	EL1 Avg	N EL1	EL2 Avg	N EL2	BL Avg	EL1 Avg	N EL1	EL2 Avg	N EL2	ITT EL1	ITT EL2	ITT EL2 (SDs)
Letter name identification (100)	28.94	38.01	725	52.38	662	32.03	80.84	777	80.47	710	37.12	23.38	0.66
Male	28.51	35.48	395	50.76	381	31.16	79.74	425	79.34	397	37.62	23.72	0.67
Female	29.51	41.02	330	54.57	281	33.16	82.16	352	81.94	313	36.64	23.38	0.66
Older (10 and above)	37.09	45.42	413	60.18	378	37.81	83.43	480	83.28	432	33.27	19.40	0.55
Younger (Below 10)	18.04	28.23	312	41.98	284	23.10	76.69	297	76.14	278	40.87	28.02	0.79
LIPACE	30.83	41.78	418	54.14	385	38.06	82.51	428	84.76	381	33.65	25.55	0.72
ROCH	26.65	33.31	307	50.26	277	26.42	79.04	349	76.49	329	40.93	20.55	0.58

Previous School (Yes)	41.44	49.58	352	62.38	329	46.09	83.39	251	86.38	227	27.75	17.94	0.50
Previous School (No)	16.99	27.32	373	42.84	333	25.51	79.65	526	77.73	483	43.87	27.00	0.76
Letter name identification per minute	29.11	38.80	725	54.99	662	32.23	84.19	777	86.68	710	39.10	26.24	0.66
Male	28.60	36.14	395	53.35	381	31.38	83.31	425	85.32	397	40.03	26.59	0.67
Female	29.78	41.95	330	57.19	281	33.32	85.26	352	88.44	313	38.21	26.43	0.67
Older (10 and above)	37.33	46.45	413	63.48	378	38.09	87.42	480	90.93	432	35.72	22.93	0.58
Younger (Below 10)	18.12	28.71	312	43.64	284	23.17	79.02	297	80.12	278	42.12	29.87	0.75
LIPACE	31.08	42.94	418	57.31	385	38.34	86.11	428	91.94	381	35.21	28.00	0.71
ROCH	26.72	33.63	307	52.19	277	26.55	82.13	349	81.80	329	43.34	23.86	0.60
Previous School (Yes)	41.69	50.75	352	65.79	329	46.20	87.92	251	94.79	227	30.55	22.08	0.56
Previous School (No)	17.08	27.75	373	44.67	333	25.74	82.46	526	82.92	483	45.56	29.18	0.74
Phonemics (10)	3.00	3.46	725	4.36	662	3.00	5.13	777	5.42	710	1.67	1.17	0.50
Male	2.92	3.42	395	4.20	381	3.07	5.24	425	5.55	397	1.64	1.49	0.63
Female	3.11	3.52	330	4.59	281	2.92	5.01	352	5.24	313	1.68	0.78	0.33
Older (10 and above)	3.41	3.70	413	4.62	378	3.17	5.42	480	5.68	432	1.78	1.19	0.51

Younger (Below 10)	2.46	3.15	312	4.03	284	2.75	4.68	297	5.00	278	1.39	1.02	0.44
LIPACE	2.89	3.59	418	4.29	385	3.14	5.12	428	5.84	381	1.53	1.61	0.69
ROCH	3.14	3.31	307	4.45	277	2.87	5.15	349	5.03	329	1.83	0.67	0.29
Previous School (Yes)	3.48	3.99	352	4.74	329	3.25	5.38	251	5.68	227	1.37	0.85	0.36
Previous School (No)	2.55	2.98	373	4.00	333	2.89	5.02	526	5.29	483	1.89	1.33	0.57
Familiar word reading (50)	2.21	4.43	725	8.62	662	1.61	22.24	777	23.75	710	18.15	16.04	1.12
Male	2.02	4.08	395	8.11	381	1.82	22.58	425	23.46	397	18.05	15.33	1.07
Female	2.47	4.84	330	9.31	281	1.33	21.82	352	24.11	313	18.20	17.09	1.19
Older (10 and above)	3.30	5.72	413	10.57	378	2.27	23.49	480	25.39	432	18.22	15.60	1.09
Younger (Below 10)	0.76	2.72	312	6.01	284	0.59	20.24	297	21.21	278	17.64	15.92	1.11
LIPACE	2.41	4.98	418	8.71	385	1.79	24.28	428	26.96	381	20.42	19.63	1.37
ROCH	1.96	3.73	307	8.52	277	1.44	20.05	349	20.75	329	15.25	11.68	0.82
Previous School (Yes)	3.41	6.03	352	10.20	329	3.04	24.44	251	27.21	227	18.56	16.53	1.16
Previous School (No)	1.06	2.94	373	7.11	333	0.94	21.22	526	22.14	483	17.90	15.32	1.07
Familiar word reading per minute	2.04	4.94	725	10.41	662	1.58	23.12	777	26.68	710	18.56	17.18	0.90

Male	1.95	4.59	395	9.88	381	1.77	23.66	425	26.24	397	18.79	16.20	0.85
Female	2.18	5.34	330	11.13	281	1.33	22.47	352	27.24	313	18.23	18.65	0.97
Older (10 and above)	3.01	6.28	413	12.70	378	2.21	24.41	480	28.71	432	18.43	16.84	0.88
Younger (Below 10)	0.75	3.16	312	7.35	284	0.61	21.07	297	23.54	278	18.35	16.78	0.88
LIPACE	2.18	5.77	418	10.70	385	1.70	25.03	428	30.19	381	20.61	21.03	1.10
ROCH	1.88	3.90	307	10.06	277	1.47	21.08	349	23.41	329	15.95	12.56	0.66
Previous School (Yes)	3.17	6.61	352	12.38	329	2.96	25.69	251	30.59	227	19.33	17.73	0.93
Previous School (No)	0.97	3.39	373	8.53	333	0.94	21.94	526	24.86	483	18.19	16.44	0.86
Non-word reading (50)	0.33	0.85	725	1.56	662	0.05	8.19	777	7.93	710	7.90	7.10	1.08
Male	0.22	0.94	395	1.55	381	0.06	9.30	425	8.76	397	8.81	8.01	1.22
Female	0.47	0.74	330	1.58	281	0.03	6.84	352	6.86	313	6.79	5.99	0.91
Older (10 and above)	0.49	0.80	413	2.06	378	0.07	8.43	480	8.45	432	8.09	7.13	1.09
Younger (Below 10)	0.10	0.92	312	0.89	284	0.01	7.80	297	7.13	278	7.45	6.62	1.01
LIPACE	0.30	0.96	418	1.85	385	0.05	8.29	428	8.53	381	8.31	7.14	1.09
ROCH	0.35	0.71	307	1.21	277	0.05	8.08	349	7.37	329	7.44	7.14	1.09

Previous School (Yes)	0.32	1.22	352	2.30	329	0.08	9.38	251	8.93	227	8.64	7.52	1.15
Previous School (No)	0.33	0.51	373	0.85	333	0.04	7.63	526	7.46	483	7.59	6.87	1.05
Non-word reading (50)	0.13	0.92	725	1.59	662	0.05	8.03	777	7.83	710	7.67	6.95	1.11
Male	0.10	1.01	395	1.58	381	0.06	9.03	425	8.66	397	8.55	7.84	1.25
Female	0.16	0.82	330	1.59	281	0.04	6.82	352	6.75	313	6.62	5.82	0.93
Older (10 and above)	0.20	0.84	413	2.09	378	0.07	8.20	480	8.29	432	7.84	6.90	1.10
Younger (Below 10)	0.03	1.03	312	0.91	284	0.01	7.76	297	7.11	278	7.17	6.52	1.04
LIPACE	0.08	1.08	418	1.90	385	0.06	8.18	428	8.48	381	7.98	6.98	1.11
ROCH	0.18	0.73	307	1.21	277	0.04	7.87	349	7.23	329	7.31	6.89	1.10
Previous School (Yes)	0.17	1.12	352	2.32	329	0.08	9.15	251	8.76	227	8.49	7.30	1.16
Previous School (No)	0.08	0.74	373	0.89	333	0.04	7.51	526	7.40	483	7.28	6.71	1.07
Passage reading	3.87	6.05	725	12.21	662	3.25	26.62	777	29.05	710	21.28	18.92	0.91
Male	3.16	5.40	395	11.60	381	3.72	26.88	425	28.80	397	21.04	18.36	0.88
Female	4.84	6.83	330	13.03	281	2.64	26.31	352	29.38	313	21.66	19.71	0.95
Older (10 and above)	5.91	7.54	413	14.69	378	4.65	28.76	480	31.27	432	21.56	18.74	0.90

Younger (Below 10)	1.15	4.09	312	8.90	284	1.08	23.22	297	25.63	278	19.85	18.00	0.86
LIPACE	3.50	7.43	418	11.99	385	3.52	30.45	428	34.22	381	24.11	24.30	1.17
ROCH	4.32	4.34	307	12.48	277	2.99	22.52	349	24.25	329	17.59	12.21	0.59
Previous School (Yes)	6.07	7.93	352	13.77	329	5.72	31.25	251	34.68	227	23.23	20.79	1.00
Previous School (No)	1.77	4.32	373	10.72	333	2.10	24.48	526	26.44	483	20.08	17.56	0.84
Oral reading fluency per minute	3.89	7.34	724	15.31	662	3.21	28.74	777	34.30	710	22.04	21.60	0.76
Male	3.17	6.52	394	14.35	381	3.67	29.19	425	33.57	397	22.10	20.66	0.73
Female	4.87	8.32	330	16.60	281	2.63	28.19	352	35.25	313	21.94	23.03	0.81
Older (10 and above)	5.91	8.99	412	18.28	378	4.62	31.27	480	37.51	432	22.27	21.71	0.76
Younger (Below 10)	1.20	5.17	312	11.34	284	1.05	24.69	297	29.35	278	20.62	19.92	0.70
LIPACE	3.51	9.54	417	15.25	385	3.45	32.33	428	40.18	381	24.11	27.41	0.96
ROCH	4.36	4.61	307	15.37	277	2.99	24.88	349	28.84	329	19.29	14.25	0.50
Previous School (Yes)	6.06	9.46	351	17.42	329	5.65	33.89	251	41.23	227	24.23	23.66	0.83
Previous School (No)	1.82	5.39	373	13.29	333	2.08	26.35	526	31.09	483	20.69	20.13	0.71
Reading comprehension (5)	0.20	0.39	719	0.81	662	0.15	1.63	762	1.94	710	1.31	1.29	0.90

Male	0.17	0.34	392	0.75	381	0.17	1.70	415	1.97	397	1.36	1.33	0.92
Female	0.24	0.44	327	0.89	281	0.12	1.55	347	1.92	313	1.27	1.27	0.88
Older (10 and above)	0.30	0.50	409	1.02	378	0.22	1.77	471	2.13	432	1.33	1.31	0.91
Younger (Below 10)	0.06	0.24	310	0.53	284	0.04	1.42	291	1.66	278	1.22	1.20	0.83
LIPACE	0.19	0.44	413	0.77	385	0.15	1.83	413	2.19	381	1.49	1.63	1.13
ROCH	0.21	0.31	306	0.86	277	0.14	1.43	349	1.71	329	1.08	0.88	0.61
Previous School (Yes)	0.34	0.51	346	0.95	329	0.32	1.99	248	2.42	227	1.47	1.54	1.07
Previous School (No)	0.06	0.27	373	0.68	333	0.07	1.47	514	1.72	483	1.21	1.14	0.79
Listening comprehension (3)	1.48	1.38	725	1.79	662	1.55	1.97	777	2.05	710	0.57	0.23	0.21
Male	1.43	1.38	395	1.73	381	1.53	1.98	425	2.06	397	0.50	0.32	0.28
Female	1.54	1.37	330	1.88	281	1.57	1.97	352	2.03	313	0.64	0.14	0.13
Older (10 and above)	1.56	1.51	413	1.87	378	1.68	2.06	480	2.07	432	0.52	0.19	0.17
Younger (Below 10)	1.36	1.20	312	1.69	284	1.36	1.84	297	2.01	278	0.58	0.27	0.24
LIPACE	1.62	1.46	418	1.90	385	1.81	2.06	428	2.14	381	0.61	0.31	0.28
ROCH	1.30	1.27	307	1.66	277	1.31	1.88	349	1.96	329	0.55	0.17	0.15

Previous School (Yes)	1.69	1.59	352	1.97	329	1.80	2.03	251	2.25	227	0.42	0.18	0.16
Previous School (No)	1.28	1.18	373	1.63	333	1.43	1.95	526	1.95	483	0.66	0.26	0.23

Note: Baseline averages differ slightly from the baseline averages at the first endline due to different sample compositions between the first and second endline.

EGMA	Control			Treatment			Treatment Effects			GSC		
Subtask	BL Avg	EL1 Avg (N = 725)	EL2 Avg (N = 662)	BL Avg	EL1 Avg (N = 777)	EL2 Avg (N = 710)	ITT EL1	ITT EL2	ITT EL2 (SDs)	BL Avg	EL1 Avg (N = 348)	EL2 Avg (N = 340)
Number identification (50)	6.29	8.83	12.70	6.25	17.69	19.79	8.25	6.90	0.68	14.63	18.65	21.49
Number identification per minute	7.07	9.53	14.97	6.76	18.97	23.15	8.90	8.13	0.60	15.62	20.40	26.24
Quantity discrimination (10)	2.46	3.56	4.52	2.75	6.38	6.69	2.47	2.06	0.61	5.19	6.20	7.12
Missing number (10)	0.28	0.68	1.36	0.28	2.49	2.52	1.84	1.38	0.62	1.00	1.79	2.95
Addition (15)	2.15	3.16	4.89	2.46	6.18	7.21	2.63	2.22	0.44	5.65	7.01	9.01
Addition per minute	2.19	3.26	5.16	2.50	6.35	7.54	2.67	2.23	0.40	5.64	7.26	9.84
Subtraction (15)	1.07	1.86	3.07	1.22	3.90	4.70	1.88	1.66	0.41	3.02	4.48	6.10

Subtraction per minute	1.06	1.92	3.16	1.24	3.97	4.79	1.84	1.62	0.38	2.94	4.55	6.28
Word problems (5)	1.23	1.54	2.13	1.26	2.35	2.60	0.75	0.48	0.32	2.00	2.58	3.13

Note: Baseline averages differ slightly from the baseline averages at the first endline due to different sample compositions between the first and second endline.

EGMA Subgroup Analysis	Control					Treatment					Treatment Effects		
Subtask	BL Avg	EL1 Avg	N EL1	EL2 Avg	N EL2	BL Avg	EL1 Avg	N EL1	EL2 Avg	N EL2	ITT EL1	ITT EL2	ITT EL2 (SDs)
Numbers identification (50)	6.29	8.83	725	12.70	662	6.25	17.69	777	19.79	710	8.25	6.90	0.68
Male	6.72	9.45	395	13.21	381	6.67	18.77	425	20.63	397	8.32	7.10	0.70
Female	5.72	8.09	330	12.00	281	5.71	16.38	352	18.73	313	8.22	6.75	0.66
Older (10 and above)	8.80	11.05	413	14.91	378	7.99	18.83	480	20.98	432	7.27	6.07	0.60
Younger (Below 10)	2.94	5.90	312	9.74	284	3.56	15.86	297	17.96	278	8.91	7.52	0.74
LIPACE	6.64	9.66	418	13.22	385	7.12	18.82	428	21.64	381	8.62	8.24	0.81
ROCH	5.88	7.79	307	12.07	277	5.45	16.48	349	18.07	329	7.73	5.27	0.52
Previous School (Yes)	9.18	11.48	352	15.08	329	9.75	19.62	251	21.80	227	7.14	6.26	0.62
Previous School (No)	3.53	6.37	373	10.42	333	4.63	16.80	526	18.86	483	9.09	7.33	0.72

Numbers identification per minute	7.07	9.53	671	14.97	662	6.76	18.97	670	23.15	710	8.90	8.13	0.60
Male	7.53	10.26	372	15.76	381	7.09	20.34	359	24.12	397	9.18	8.17	0.60
Female	6.46	8.64	299	13.90	281	6.34	17.37	311	21.90	313	8.64	8.18	0.60
Older (10 and above)	9.65	12.15	383	18.02	378	8.55	20.49	415	25.07	432	7.99	7.36	0.54
Younger (Below 10)	3.62	6.08	288	10.88	284	4.00	16.52	255	20.18	278	9.53	8.58	0.63
LIPACE	7.58	10.63	386	15.42	385	7.63	20.25	373	25.25	381	9.07	9.61	0.71
ROCH	6.45	8.17	285	14.42	277	5.95	17.59	297	21.19	329	8.59	6.26	0.46
Previous School (Yes)	10.14	12.66	315	18.02	329	10.42	21.70	212	26.06	227	8.06	7.68	0.56
Previous School (No)	4.14	6.82	356	12.05	333	5.06	17.75	458	21.80	483	9.59	8.48	0.62
Quantity discrimination (10)	2.46	3.56	725	4.52	662	2.75	6.38	777	6.69	710	2.47	2.06	0.61
Male	2.73	3.59	395	4.58	381	3.01	6.86	425	6.93	397	2.74	2.17	0.64
Female	2.10	3.52	330	4.43	281	2.42	5.79	352	6.37	313	2.14	1.93	0.57
Older (10 and above)	3.18	4.08	413	5.07	378	3.29	6.75	480	7.00	432	2.29	1.79	0.53
Younger (Below 10)	1.49	2.87	312	3.77	284	1.91	5.78	297	6.20	278	2.65	2.42	0.71
LIPACE	2.72	3.88	418	4.70	385	2.73	6.55	428	7.06	381	2.58	2.39	0.70

ROCH	2.15	3.16	307	4.30	277	2.77	6.19	349	6.34	329	2.34	1.68	0.50
Previous School (Yes)	3.13	4.36	352	5.31	329	3.53	6.89	251	7.27	227	2.11	1.56	0.46
Previous School (No)	1.81	2.82	373	3.76	333	2.39	6.14	526	6.42	483	2.77	2.46	0.72
Missing number (10)	0.28	0.68	725	1.36	662	0.28	2.49	777	2.52	710	1.84	1.38	0.62
Male	0.34	0.72	395	1.34	381	0.28	2.94	425	2.77	397	2.18	1.70	0.76
Female	0.20	0.63	330	1.38	281	0.28	1.95	352	2.19	313	1.41	0.96	0.43
Older (10 and above)	0.42	0.83	413	1.58	378	0.38	2.88	480	2.83	432	2.07	1.44	0.65
Younger (Below 10)	0.10	0.48	312	1.06	284	0.12	1.86	297	2.03	278	1.46	1.18	0.53
LIPACE	0.31	0.77	418	1.56	385	0.30	2.52	428	2.89	381	1.89	1.58	0.71
ROCH	0.25	0.57	307	1.12	277	0.26	2.46	349	2.17	329	1.79	1.16	0.52
Previous School (Yes)	0.40	0.77	352	1.65	329	0.43	2.93	251	3.13	227	2.15	1.52	0.68
Previous School (No)	0.17	0.59	373	1.08	333	0.21	2.29	526	2.23	483	1.69	1.29	0.58
Addition (15)	2.15	3.16	725	4.89	662	2.46	6.18	777	7.21	710	2.63	2.22	0.44
Male	2.46	3.48	395	5.14	381	2.75	6.85	425	7.72	397	2.78	2.33	0.47
Female	1.72	2.78	330	4.55	281	2.07	5.38	352	6.55	313	2.49	2.09	0.42

Older (10 and above)	3.05	4.13	413	5.82	378	3.21	6.93	480	8.01	432	2.47	2.08	0.42
Younger (Below 10)	0.94	1.87	312	3.65	284	1.29	4.99	297	5.97	278	2.65	2.34	0.47
LIPACE	2.33	3.51	418	5.23	385	2.86	6.46	428	7.86	381	2.63	2.42	0.48
ROCH	1.93	2.71	307	4.48	277	2.08	5.89	349	6.61	329	2.62	1.99	0.40
Previous School (Yes)	3.12	4.22	352	6.01	329	4.00	7.28	251	8.71	227	2.23	2.01	0.40
Previous School (No)	1.22	2.17	373	3.82	333	1.74	5.68	526	6.51	483	2.85	2.41	0.48
Addition per minute	2.19	3.26	725	5.16	662	2.50	6.35	777	7.54	710	2.67	2.23	0.40
Male	2.52	3.63	395	5.44	381	2.81	7.07	425	8.13	397	2.82	2.39	0.43
Female	1.76	2.82	330	4.79	281	2.11	5.49	352	6.78	313	2.53	2.05	0.37
Older (10 and above)	3.12	4.22	413	6.17	378	3.26	7.16	480	8.45	432	2.58	2.15	0.39
Younger (Below 10)	0.95	1.99	312	3.82	284	1.34	5.06	297	6.13	278	2.60	2.23	0.40
LIPACE	2.39	3.69	418	5.64	385	2.91	6.57	428	8.18	381	2.57	2.25	0.41
ROCH	1.96	2.72	307	4.60	277	2.13	6.12	349	6.94	329	2.78	2.22	0.40
Previous School (Yes)	3.18	4.38	352	6.39	329	4.04	7.52	251	9.15	227	2.24	1.97	0.36
Previous School (No)	1.26	2.22	373	3.99	333	1.79	5.81	526	6.80	483	2.92	2.49	0.45

Subtraction (15)	1.07	1.86	725	3.07	662	1.22	3.90	777	4.70	710	1.88	1.66	0.41
Male	1.30	2.13	395	3.18	381	1.28	4.33	425	5.09	397	1.95	1.86	0.46
Female	0.75	1.53	330	2.91	281	1.14	3.39	352	4.20	313	1.75	1.31	0.33
Older (10 and above)	1.48	2.42	413	3.84	378	1.61	4.49	480	5.40	432	1.90	1.58	0.39
Younger (Below 10)	0.51	1.11	312	2.04	284	0.61	2.96	297	3.62	278	1.81	1.80	0.45
LIPACE	1.31	2.19	418	3.51	385	1.46	4.09	428	5.10	381	1.82	1.49	0.37
ROCH	0.77	1.44	307	2.53	277	0.99	3.71	349	4.33	329	1.97	1.92	0.48
Previous School (Yes)	1.62	2.54	352	3.87	329	2.14	5.02	251	5.75	227	1.78	1.40	0.35
Previous School (No)	0.53	1.22	373	2.30	333	0.79	3.39	526	4.21	483	1.98	1.81	0.45
Subtraction per minute	1.06	1.92	725	3.16	662	1.24	3.97	777	4.79	710	1.84	1.62	0.38
Male	1.27	2.23	395	3.25	381	1.30	4.42	425	5.20	397	1.89	1.86	0.44
Female	0.77	1.56	330	3.04	281	1.15	3.43	352	4.27	313	1.75	1.22	0.29
Older (10 and above)	1.46	2.46	413	3.97	378	1.63	4.58	480	5.50	432	1.86	1.49	0.36
Younger (Below 10)	0.52	1.21	312	2.08	284	0.63	3.00	297	3.70	278	1.75	1.81	0.43
LIPACE	1.28	2.29	418	3.65	385	1.47	4.18	428	5.18	381	1.74	1.37	0.32

ROCH	0.79	1.46	307	2.57	277	1.02	3.75	349	4.43	329	1.97	1.98	0.47
Previous School (Yes)	1.59	2.62	352	3.95	329	2.15	5.05	251	5.85	227	1.67	1.38	0.33
Previous School (No)	0.55	1.28	373	2.41	333	0.81	3.47	526	4.30	483	2.01	1.77	0.42
Word problems (5)	1.23	1.54	713	2.13	662	1.26	2.35	774	2.60	710	0.75	0.48	0.32
Male	1.30	1.58	390	2.21	381	1.27	2.45	424	2.77	397	0.77	0.55	0.37
Female	1.13	1.48	323	2.03	281	1.25	2.22	350	2.39	313	0.73	0.38	0.25
Older (10 and above)	1.48	1.86	407	2.45	378	1.50	2.56	479	2.86	432	0.69	0.43	0.28
Younger (Below 10)	0.89	1.11	306	1.71	284	0.88	2.01	295	2.21	278	0.82	0.51	0.34
LIPACE	1.29	1.48	414	2.20	385	1.34	2.32	426	2.75	381	0.75	0.56	0.37
ROCH	1.15	1.61	299	2.06	277	1.18	2.38	348	2.47	329	0.74	0.38	0.25
Previous School (Yes)	1.49	1.81	349	2.45	329	1.69	2.53	250	2.96	227	0.60	0.34	0.23
Previous School (No)	0.98	1.28	364	1.83	333	1.06	2.27	524	2.44	483	0.89	0.61	0.41

Note: Baseline averages differ slightly from the baseline averages at the first endline due to different sample compositions between the first and second endline.

Appendix A4 : Gender-wise results

Subtask	C BL (Boys)	C EL2 (Boys)	T BL (Boys)	T EL2 (Boys)	ITT EL2 (Boys)	C BL (Girls)	C EL2 (Girls)	T BL (Girls)	T EL2 (Girls)	ITT EL2 (Girls)	Gender Interaction Coefficient EL2	Gender Interaction Coefficient EL2 (p-value)
EGRA												
Letter name identification (100)	28.51	50.76	31.16	79.34	23.72	29.51	54.57	33.16	81.94	23.38	-0.79	0.75
Letter name identification per minute	28.60	53.35	31.38	85.32	26.59	29.78	57.19	33.32	88.44	26.43	-0.32	0.91
Phonemics (10)	2.92	4.20	3.07	5.55	1.49	3.11	4.59	2.92	5.24	0.78	-0.67	0.01
Familiar word reading (50)	2.02	8.11	1.82	23.46	15.33	2.47	9.31	1.33	24.11	17.09	0.38	0.77
Familiar word reading per minute	1.95	9.88	1.77	26.24	16.20	2.18	11.13	1.33	27.24	18.65	0.97	0.55
Non-word reading (50)	0.22	1.55	0.06	8.76	8.01	0.47	1.58	0.03	6.86	5.99	-1.68	0.06
Nonword reading per minute	0.10	1.58	0.06	8.66	7.84	0.16	1.59	0.04	6.75	5.82	-1.68	0.04
Passage reading	3.16	11.60	3.72	28.80	18.36	4.84	13.03	2.64	29.38	19.71	0.52	0.80

Passage reading per minute	3.17	14.35	3.67	33.57	20.66	4.87	16.60	2.63	35.25	23.03	1.43	0.61
Reading comprehension (5)	0.17	0.75	0.17	1.97	1.33	0.24	0.89	0.12	1.92	1.27	-0.12	0.42
Listening comprehension (3)	1.43	1.73	1.53	2.06	0.32	1.54	1.88	1.57	2.03	0.14	-0.12	0.30
EGMA												
Number identification (50)	6.72	13.21	6.67	20.63	7.10	5.72	12.00	5.71	18.73	6.75	-0.55	0.48
Number identification per minute	7.53	15.76	7.09	24.12	8.17	6.46	13.90	6.34	21.90	8.18	-0.49	0.64
Quantity discrimination (10)	2.73	4.58	3.01	6.93	2.17	2.10	4.43	2.42	6.37	1.93	-0.29	0.38
Missing number (10)	0.34	1.34	0.28	2.77	1.70	0.20	1.38	0.28	2.19	0.96	-0.75	0.00
Addition (15)	2.46	5.14	2.75	7.72	2.33	1.72	4.55	2.07	6.55	2.09	-0.51	0.26
Addition per minute	2.52	5.44	2.81	8.13	2.39	1.76	4.79	2.11	6.78	2.05	-0.63	0.19
Subtraction (15)	1.30	3.18	1.28	5.09	1.86	0.75	2.91	1.14	4.20	1.31	-0.84	0.02
Subtraction per minute	1.27	3.25	1.30	5.20	1.86	0.77	3.04	1.15	4.27	1.22	-0.92	0.01

Word problems (5)	1.30	2.21	1.27	2.77	0.55	1.13	2.03	1.25	2.39	0.38	-0.24	0.06

Note: Baseline averages differ slightly from the baseline averages at the first endline due to different sample compositions between the first and second endline.

Appendix A5: IPW Analysis

Subtask	ATE EL1	ATE EL2
Letter name identification	39.22	26.84
Letter names per minute	41.59	30.15
Phonemics	1.75	1.20
Familiar word reading	18.75	16.80
Familiar word per minute	19.12	18.21
Non-word reading	7.94	7.20
Non-word reading per minute	7.64	7.04
Passage reading	21.81	18.98
Passage score per minute	22.74	21.69
Reading comprehension	1.35	1.30
Listening comprehension	0.62	0.29

Number identification	8.68	7.53
Numbers per minute	9.47	8.88
Quantity discrimination	2.74	2.25
Missing number	1.88	1.36
Addition	2.76	2.39
Addition per minute	2.79	2.41
Subtraction	1.98	1.75
Subtraction per minute	1.94	1.72
Word problems	0.79	0.53

Appendix B: Analytical Modelling

Table B1: IRT Results

Assessment	ITT (SDs) EL1	ITT (SDs) EL2
EGRA	1.26	0.82
EGMA	0.81	0.58

Methodology

Item response theory (IRT) is a statistical method used to estimate underlying or 'latent' ability and traits that cannot be directly observed. In education research IRT has been used to estimate attributes such as literacy skills and numeracy skills. The main underlying assumption in IRT is that each question in a well-designed test gives us information about a participant's latent trait or ability (Columbia University, 2023). In our context, when we use IRT we assume that a child's performance on the EGRA and EGMA subtasks can be explained by a measure of their latent literacy and numeracy ability, respectively. Before presenting the various steps we took to estimate the Luminos program's effect on numeracy and literacy, we highlight that IRT models typically require test questions to be dichotomous (i.e., 0/1) or ordinal (i.e., Likert scales). In order to use IRT with our data we have transformed all the scores on EGRA and EGMA subtasks to dichotomous variables by defining success (i.e., 1) as when a child scored more than 50% of the possible points on a subtask, and a child failed (i.e., 0) otherwise. For example, for the letter identification task, which has a maximum possible score of 100, if a child was able to identify between 51 – 100 letters they succeeded (i.e., assigned 1) at the task and if they identified 0 – 50 they failed the task (i.e., assigned 0). This 50% cutoff was set to ensure that we had enough variation at baseline and endline. A higher cut-off, such as greater than 80%, leads to some subtasks having no child marked as succeeding at the subtask at baseline or endline.

There are three common dichotomous IRT models. First, a 1-parameter which estimates each test item's difficulty. For example, word identification is a task that was answered correctly by a smaller proportion of our returning OOSC

sample (24%) than the letter name identification (63%). The 1-parameter model will use this information to assign a higher difficulty to the word identification subtask than the letter name identification task. A 2-parameter model also assumes that questions can vary in their discrimination. Discrimination is defined as a question's ability to differentiate between participants with larger and smaller values of the latent trait being estimated. For example, a numeracy task with perfect discrimination would be one that is answered incorrectly by all students below a fixed value of numeracy ability but is then answered correctly by all students above the same fixed value. A 1-parameter model still calculates a discrimination parameter but it assumes all items in the test discriminate high-performing and low-performing participants in the same way. In a 2-parameter model the model assumes that discrimination can vary across questions (StataCorp, 2023). A 3-parameter model adds a guessing parameter which assumes that some of the variation in success rates on items can be explained by the fact that a participant might have guessed the right answer. So, if a question is difficult and discriminates well but is answered correctly by a weaker performing student, a 3-parameter model will try to capture how often participants guessed the right answer to a question.

We conducted all our IRT analysis in STATA 18. We use a 1-parameter IRT model for both EGRA and EGMA scores. We do not use a 2-parameter model because for EGRA subtasks a 2-parameter model failed to converge. This means that when we allow EGRA items to vary in how well they discriminate between stronger and weaker students, the model could not calculate stable estimates of the model's parameters. This might be because high-performing children in our sample failed easier questions in patterns that were hard to explain by the model, or vice-versa. Another explanation could be that the IRT model we use, which is a unidimensional model of latent ability, may be too simple to explain literacy ability in our sample. Our model assumes that literacy can be measured by one latent trait. It may be that more than one ability is responsible for higher performance on literacy tasks and literacy ability should be split into more than one trait. For example, perhaps literacy should be measured using a multi-dimensional IRT, as described by Hartig and Höhler, 2009, which calculates two latent traits such as (1) "listening" and (2) "reading comprehension" for literacy. The fact that 18% (133/725) of control students got the maximum score on the listening comprehension task at endline while only 2% of control students (15/725) got the maximum score on passage reading at endline might indicate that language ability should be modeled as having of several underlying competencies and can't be summarized into one value for literacy ability in our sample. We note that we are able to successfully estimate a 2-parameter model for EGMA subtasks but we model numeracy using a 1-parameter model to be consistent with our calculation of literacy ability.

For both EGRA and EGMA, we run an IRT model on baseline, endline, and second endline data that produces estimates of item difficulty and each student's ability in each phase. The raw latent ability measure that is calculated by the model typically ranges from -4 to 4 with smaller values indicating lower values of the latent trait and higher values indicating larger values of the latent trait. However, these values are often hard to interpret so we report all results for our regressions using standardized values of the latent ability trait. We include in the EGRA IRT model 7 dichotomous test items for success or failure on each, and for EGMA IRT model we use 6 dichotomous items for success or failure in each EGMA subtask. We run the IRT model and store estimates of baseline and endline numeracy and literacy ability from the IRT model. We then run our ITT and TOT regressions using our measure of literacy and numeracy ability instead of the raw endline and baseline scores. This leads to the following analytical model:

$$\theta_{ij}^{el} = \beta^*_i T_j + \beta^* \theta_{ij}^{bl} + X_{-\theta_{bl}}{}'_{ij}\gamma + \propto'_s + \varepsilon^*_{ij}$$

Where:

- θ_{ij}^{el} denotes the numeracy or literacy for child i in community j at first endline or second endline calculated using IRT
- θ_{ij}^{bl} denotes the numeracy or literacy for child i in community j at baseline calculated using IRT

www.IDinsight.org

@IDinsight

IDinsight