How Sustainable Are Learning Gains After School Closures?

Findings From a 5-Year Randomized Controlled Trial in Zambia

Thomas De Hoop, Hannah Ring, Garima Siwach, Paula Dias, Gelson Tembo

SEPTEMBER 2025

Acknowledgements

The authors would like to thank the AIR Opportunity Fund for funding this study. We would especially like to thank Laura Evans, Chesci Horn, Amosi Chipasi, and Katie Kerr at Promoting Equality in African Schools, Michael J. Weiss at MDRC, and Jordan Rickles, Pooja Nakamura, David Seidenfeld, and Reshma Patel at the American Institutes for Research for their valuable comments and feedback on this study. All errors are our own. Our acknowledgements would be incomplete without mentioning the team of very able data collectors and team leaders at Palm Associates in Zambia. We also acknowledge the input of the full team at Impact Network.

Co-pilot tools were used to copy-edit the paper. The authors reviewed all these copy-edits before submission.

Abstract

Evidence on post-COVID-19 learning losses in Africa remains limited, and even less is known about whether interventions that improved learning outcomes before pandemic-related school closures remained effective after schools reopened. This study addresses that gap by evaluating the persistence of learning gains from a technology-aided structured pedagogy program in rural Zambia, more than a year after schools resumed. Drawing on a five-year cluster-randomized controlled trial, we find that literacy improvements were sustained at 0.26 standard deviations, though smaller than pre-closure effects. In contrast, significant gains in mathematics were observed only among boys and students with above-median baseline scores, and these effects lost statistical significance after correcting for multiple hypothesis testing. While pandemic-related disruptions complicate interpretation, the findings suggest that structured pedagogy programs, previously shown to be effective, can partially mitigate the negative impact of school closures on learning outcomes.

1. Introduction

COVID-19-related school closures caused widespread learning losses globally (e.g., Betthäuser et al., 2023; Di Pietro, 2023; Goldhaber et al., 2023; Singh et al., 2024). However, rigorous evidence on the pandemic's effects on learning outcomes in sub-Saharan Africa, particularly in low-income countries, remains limited. Two recent meta-analyses included only one study from the region, focused on South Africa, a middle-income country (Ardington et al., 2023; Betthäuser et al., 2023; Di Pietro, 2023). This lack of evidence leaves open critical questions about the magnitude and persistence of post-COVID-19 learning losses in low-income African contexts.

While evidence on post-COVID-19 learning losses in Africa is scarce, the gap in research on effective responses is even more pronounced. Most existing studies focus on programs specifically designed to mitigate pandemic-related learning losses (e.g., Angrist et al., 2022; Crawfurd et al., 2023; Hassan et al., 2021; Radhakrishnan et al., 2021; Schueler & Rodriguez-Segura, 2021), with far less attention paid to interventions that were implemented before the pandemic and continued through school closures and the return to in-person learning.

To address this gap, we examine whether learning gains from a prepandemic technology-aided education intervention were sustained after extended school closures in rural Zambia. We estimate the four-and-a-half-year impacts of the eSchool 360 model, a multifaceted, technology-aided instruction program that was implemented by Impact Network[1], using data from a 5-year cluster-randomized controlled trial that included substantial pauses due to school closures. At the time of this study, the eSchool 360 model combined a standardized e-learning curriculum and pedagogy, ongoing teacher training, community engagement through local teacher recruitment and parent-teacher associations (PTAs), and free primary education, including uniforms and books.[2]

To estimate intention-to-treat (ITT) effects, we randomly sampled 30 households from a census-generated frame across 30 treatment and 33 control schools. We applied an ANCOVA model to estimate ITT effects on school enrollment and student achievement in literacy and mathematics, 14 months after program launch (midline, reported in de Hoop et al., 2024) and

[1] The eSchool 360 program is now implemented by Promoting Equality in African Schools (PEAS), which merged with Impact Network in Zambia in January 2025.
[2] The Ministry of Education also provides free primary education, but in some cases the schools charge for uniforms, books, or other school materials. The midline analysis indeed showed lower education expenditures for students in Impact Network Schools (de Hoop et al., 2024).

again 4.5 years after launch (endline). We complement these estimates with a qualitative process evaluation and data on how the program adapted to COVID-19-related disruptions.

After 4.5 years, we find statistically significant ITT effects of 0.26 standard deviations on literacy scores, though smaller than those observed before the pandemic. At midline, 14 months after launch, ITT effects included an 8-percentage-point increase in primary school enrollment, a 0.40 SD increase in literacy scores (measured through aggregate early grade reading assessment [EGRA] scores), a 0.21 SD increase in mathematics scores (measured through aggregate early grade mathematics assessment [EGMA] scores), a 0.15 SD increase in Zambian Achievement Test (ZAT) scores, and a 0.25 SD increase in oral vocabulary scores (de Hoop et al., 2024). Postpandemic, no statistically significant ITT effects were found for EGMA, ZAT, or oral vocabulary scores. While point estimates were positive, they were not statistically significant. Some positive impacts on EGMA scores were observed for boys and children with higher baseline scores, but these were not robust to multiple hypothesis testing.

Our findings suggest that structured pedagogy programs, previously shown to improve learning outcomes (e.g., Eble et al., 2021; Glewwe & Muralidharan, 2016; Muralidharan et al., 2019), can also help mitigate learning losses (particularly related to literacy) caused by COVID-19-related school closures. Structured pedagogy is “a coordinated, combined approach including lesson plans + student materials + training + ongoing support (e.g., coaching)” (Piper et al., n.d., p. 1). The eSchool 360 program is aligned with structured pedagogy because its technology component is designed solely to strengthen the effectiveness of lesson plans and teacher training (de Hoop et al., 2024).

The moderate sustained program effects are consistent with a recent large-scale quasi-experimental study in India, which found that afterschool remediation programs with structured pedagogy elements contributed to post-COVID-19 learning recovery. A panel survey of approximately 19,000 students found that two-thirds of learning loss was recovered within 6 months of reopening, with the program explaining roughly 24% of the catch-up (Singh et al., 2024). This program shared features with other structured pedagogy interventions, including “teaching at the right level” (Banerjee et al., 2017; Duflo et al., 2024). We are not aware of similar studies examining the sustainability of post-COVID-19 learning gains caused by structured pedagogy programs in sub-Saharan Africa.

Several factors complicate the interpretation of our results. The original experimental design anticipated an endline study after 3 years (de Hoop et al., 2017), which informed the decision to roll out the program across only two cohorts. Although the limited rollout was planned, the COVID-19 pandemic and associated school closures disrupted implementation and research timelines. These disruptions may have contributed to the program’s relatively low dosage: only

49% of sampled children living near treatment schools ever enrolled in an Impact Network school, and just 20% received more than 4 years of exposure. Moreover, some students, particularly lower-performing ones, did not return to school immediately or at all after the pandemic, leading to gaps in learning continuity.

Implementation activities during the pandemic further complicate interpretation. Although schools were officially closed, Impact Network teachers visited students and distributed “homework packs” to support home-based learning. In some cases, teachers met with small groups, indicating that the intervention continued in modified form rather than pausing entirely.

Overall, the findings suggest that structured pedagogy programs that showed promise in addressing the global learning crisis before the pandemic can also mitigate learning losses in rural Africa following the COVID-19 pandemic. However, questions remain about the extent to which such programs can fully offset these losses, and further research is needed to assess their scalability and external validity across diverse sub-Saharan African contexts.

2. Background

This section provides an overview of Zambia’s education system and reviews evidence on technology-in-education and structured pedagogy programs in sub-Saharan Africa, updating the literature presented in de Hoop et al. (2024) following 14 months of program implementation.

Zambia has a large community schooling system that has expanded significantly over the past 2 decades to improve access to education in remote rural areas (de Hoop et al., 2024). While community schools have some autonomy, they must register with the Ministry of Education and re-recruit government teachers, in addition to following Ministry of Education guidelines. At the time of the study, the eSchool 360 model aimed to enhance education quality in these community schools through a multifaceted, technology-aided instruction program that included: (a) a standardized e-learning curriculum and pedagogy; (b) ongoing teacher training and professional development; (c) community ownership via local teacher recruitment and parent-teacher associations (PTAs); and (d) free primary education, including books and uniforms.

Impact Network supplies solar-powered electricity and uses e-learning technology—tablets and projectors provided by Mwabu—loaded with government-approved lesson plans and interactive content in local languages to deliver the eSchool 360 model. The model includes an emphasis on phonics instruction by learning letter sounds and blending them into words in the local language.[3] The lessons use a "learning-by-doing" activity-based approach instead of rote memorization, with a focus on play-based interactive learning approaches. Impact Network delivers weekly on-site coaching, including classroom observations and encouragement of teachers to self-identify strengths and challenges from the lessons. The model also includes monthly training workshops to cover pedagogical techniques and curriculum updates, among other topics. Initially, teachers were locally recruited and lacked formal training, but by 2022, the organization began hiring government-certified teachers in response to new national regulations.

The eSchool 360 model is aligned with implementation of the Catch-Up program, which implements the "Teaching at the Right Level" methodology that groups children based on their learning level and was scaled up in all community schools in Zambia between the midline and the endline survey. Catch-Up classes in literacy and numeracy took place twice a week during 40-minute sessions.

While integrating technology into education can improve quality, evidence shows that simply providing devices is insufficient. Effective programs either support self-led learning or enhance instruction through structured pedagogy, offering teachers scripted lesson plans, training, and continuous feedback (Rodriguez-Segura, 2022). For example, the Mindspark program in India individualized content to students' learning levels and significantly improved math and reading outcomes at low cost (Muralidharan et al., 2019). However, technology tends to be less effective when not embedded in the curriculum or when teachers are not supportive (Ale et al., 2017; Barrera-Osorio & Linden, 2009; Rodriguez-Segura, 2022; Tauson & Stannard, 2018; Valiente, 2010).

Systematic reviews and impact evaluations show that structured pedagogy programs, especially those combining multiple components such as para-teachers, scripted lessons, and frequent monitoring, can significantly improve literacy and numeracy in rural sub-Saharan Africa (Eble et al., 2021; Eble & Escueta, 2023; Glewwe & Muralidharan, 2016; Snilstveit et al., 2016). For instance, a program in The Gambia demonstrated large learning gains using this approach. Similarly, a Kenyan initiative that combined tablets with detailed lesson guides, daily teacher monitoring, school construction, and financial management led to substantial test score improvements over 2 years (Gray-Lobe et al., 2022). Evidence from South Africa also suggests that structured pedagogy programs targeting instruction in children's home languages can

[3] This approach was recently augmented through the dedicated "*Read Smart"* program by introducing visual phonemic exercises for Grades 1–2 to strengthen early decoding skills. "Read Smart" was, however, not yet introduced as part of the eSchool 360 model at the time of this study. A separate cluster-RCT will examine the impact of the Read Smart program specifically.

improve reading proficiency, while those focused solely on English may negatively affect home language literacy (Moholwane et al., 2023).

These findings are consistent with the midline results of the eSchool 360 impact evaluation, which showed substantial gains in enrollment and learning outcomes 14 months after program launch (de Hoop et al., 2023).

3. Theory of Change

The theory of change aligns with a structured pedagogy approach, which emphasizes the use of lesson plans, teacher training, classroom observations, and feedback to improve instructional quality. Structured pedagogy programs have been shown to improve learning outcomes by enhancing teacher knowledge and classroom practices (Evans & Acosta, 2021; Gray-Lobe et al., 2022; Piper et al., 2015). In the eSchool 360 model, technology reinforces this approach by supporting teachers in delivering activity-based lessons aligned with the Zambian national curriculum. This design is consistent with evidence from a systematic review showing that additional resources improve student achievement only when they lead to meaningful changes in students' daily classroom experiences (Ganimian & Murnane, 2016).

The program is expected to improve learning outcomes through several pathways that enhance both the supply of and demand for education:

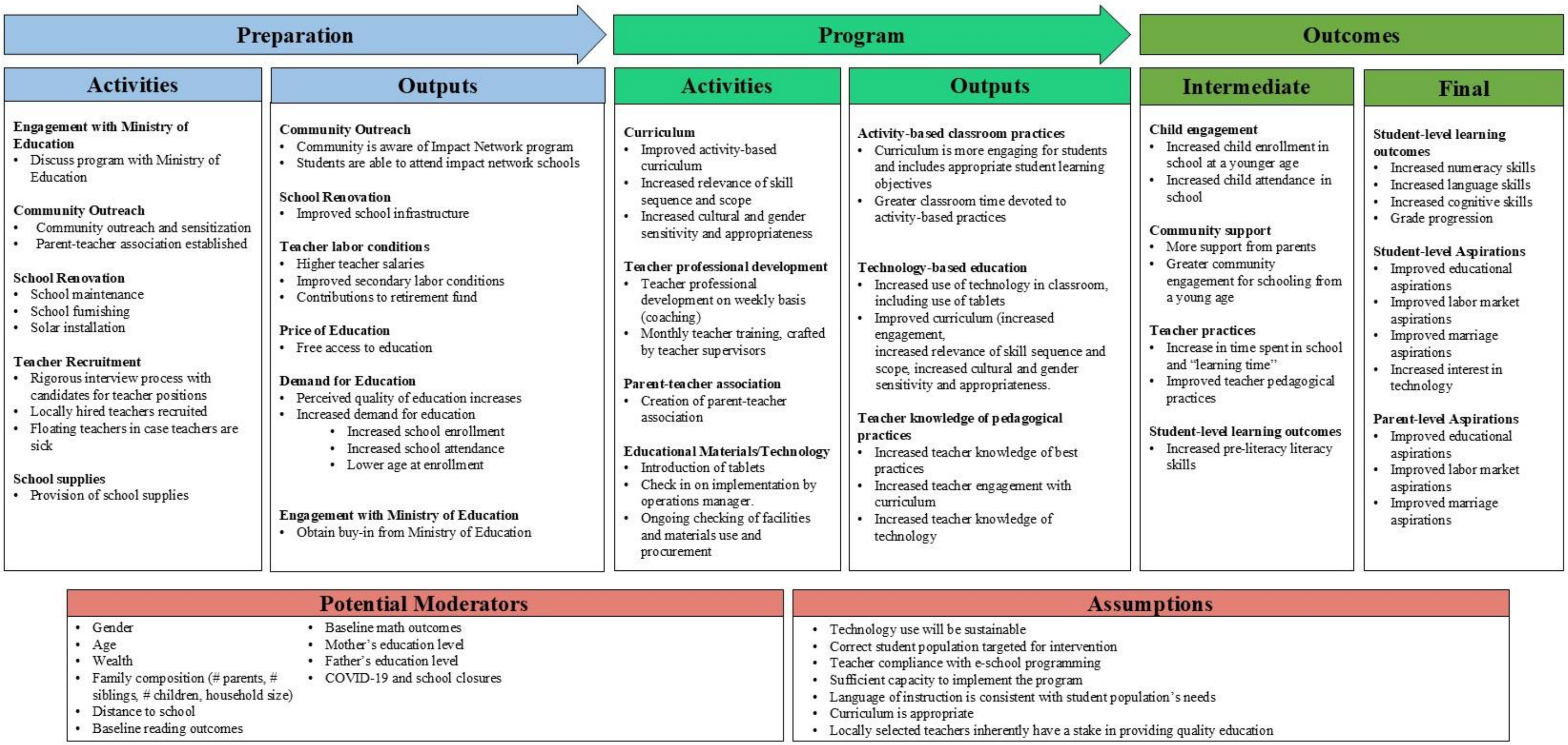
1. **Improved instructional quality through teacher development**: Regular professional development may enhance the knowledge and practices of teachers without formal training, enabling them to implement activity-based learning and deliver a more effective curriculum. These improvements are expected to translate into gains in pre-literacy, early grade reading, and mathematics.
2. **Technology-aided instruction**: Tablets preloaded with sequenced, government-approved lessons help standardize instruction and support teachers without formal training. These devices are monitored by operations staff to ensure consistent use and fidelity to the curriculum.
3. **Increased demand for education**: Infrastructure improvements and the provision of free primary education (including uniforms and books) may increase school enrollment and attendance, particularly among younger children. Increased time spent in school is expected to contribute to improved learning outcomes.
4. **Influence on aspirations of students and parents**: Improvements in the perceived quality of education could raise expectations for children's futures, potentially increasing aspirations related to education, labor market participation, and family outcomes. However, midline

results did not show statistically significant impacts on aspirations of either students or parents (de Hoop et al., 2024).

These mechanisms are illustrated in **Figure 1** (adapted from de Hoop et al., 2024), which presents the theory of change used to guide our hypotheses and evaluation strategy.

While the evaluation is designed to estimate the overall impact of the eSchool 360 program as a bundled intervention, it does not allow for disaggregating the effects of individual program components. Identifying the causal impact of specific elements, such as teacher training, technology use, or community engagement, would require a multi-arm experimental or quasi-experimental design, which was not feasible in the current study due to the limited number of treatment schools.

Figure 1. Theory of Change



4. Research Design

4.1 Quantitative Analysis

As described in de Hoop et al. (2024), we randomly assigned the eSchool 360 program across 63 community schools that met Impact Network's eligibility criteria related to geography, infrastructure, and organizational structure. Of the 149 initially identified schools, 63 were eligible. In May 2017, with oversight from the research team, local representatives from the Zambian Ministry of Education conducted the randomization: 30 schools (5 in Katete, 9 in Sinda, and 16 in Petauke) were assigned to the treatment group[4], and 33 to the control group.

To estimate ITT effects, we compared all eligible children living near the 30 treatment schools with those near the 33 control schools—regardless of enrollment. Eligibility was determined through a household census, identifying children aged 6 or older in January 2018 who had not yet attended first grade.

We randomly sampled 30 households per school from this census frame. In households with multiple eligible children, we selected the oldest. To increase statistical power, we oversampled 8- and 9-year-olds, who were more likely to be enrolled in first grade (Ministry of General Education, 2016). Where there were insufficient 7–9-year-olds within 1.5 km of a school, we included 6-year-olds and expanded the radius in 0.5 km increments until the sample was complete. This procedure was applied consistently across treatment and control areas. The final baseline sample included 888 children from treatment areas and 977 from control areas.

We collected data at three points: baseline (early 2017), midline (early 2019), and endline (mid-2022). Enumerators, fluent in local languages and trained in the learning assessment tools, used SurveyCTO on tablets to collect data, including GPS coordinates for household tracking. We measured primary learning outcomes using EGRA, EGMA, and ZAT instruments previously validated in Zambia. We also administered an oral vocabulary assessment and collected data on enrollment, attendance, parental perceptions, education expenditures, aspirations, and food security.

[4] While originally three districts, a new district, Lusangazi, was formed, made independent from Petauke district between midline and endline.

We estimated ITT effects using an ANCOVA model:

$$Y_{i2022} = \alpha + \beta\, IN_i + \delta S_i + \sigma Y_{i2018} + \mu\, C_i + \epsilon_i$$

where Y_{i2022} is the outcome at endline (measured in 2022) for child i; IN_i is an indicator for whether child i resides in a treatment area of an Impact Network school; S_i is a vector of district strata dummies; Y_{i2018} is the baseline value of the outcome of interest for child i; C_i is a vector of other control variables; and ϵ_i is the conditionally mean-zero error term. We used ordinary least squares (OLS) regression, with standard errors clustered at the school catchment area level to account for intra-cluster correlation.[5]

The study protocol was registered with RIDIE (ridie.3ieimpact.org) after the baseline survey. The pre-analysis plan (de Hoop et al., 2017) specified primary outcomes (EGRA, EGMA, ZAT, oral vocabulary), secondary outcomes (e.g., enrollment, perceptions, aspirations), and the use of OLS and multiple comparison corrections. As in the midline analysis, we did not implement the planned factor analysis and instead focused on EGRA and EGMA subscores (de Hoop et al., 2024). We made this decision because both EGRA and EGMA measure distinct subskills (e.g., letter naming vs. reading comprehension).

We explored treatment heterogeneity by age, gender, region, socioeconomic status, caregiver education, and baseline test scores. For continuous moderators, we created binary indicators for above/below median values and interacted them with treatment status. Socioeconomic status was measured using an asset index derived from principal component analysis (Filmer & Pritchett, 2001). We applied the Benjamini-Hochberg procedure (Banerjee et al., 2015) to correct for multiple comparisons within EGRA and EGMA domains, as prespecified. We also used randomization inference (Heß, 2017; Young, 2019) to assess the robustness of ITT estimates.

We also estimated two types of treatment-on-the-treated (TOT) effects: (a) enrollment-based TOT, which uses program assignment as an instrument for enrollment in the past year; and (b) attendance-based TOT, which uses program assignment as an instrument for attending an Impact Network school at least three times in the week before the survey.

To address potential composition effects, we sampled households independently of school enrollment. This approach avoids bias from comparing students who may differ systematically across treatment and control schools due to observable or unobservable characteristics.

[5] We conducted analyses with and without applying post-stratification weights. This paper discusses results without post-stratification weights related to student age, but all findings are robust to applying weights.

4.2 Qualitative Analysis

We conducted two rounds of qualitative process evaluation: midline (early 2019) and endline (early 2022). Our process evaluation approach included three data collection methods: (a) key informant interviews (KIIs) with teachers, supervisors, and Impact Network staff; (b) focus group discussions (FGDs) with students and parents; and (c) classroom observations in treatment schools.

KIIs explored perceptions of program implementation and its effects on teaching and learning. FGDs provided insights into parent and student experiences, including engagement with the program and use of school facilities. Classroom observations assessed the use of e-learning tools and pedagogical practices, as well as the child-friendliness of the school environment.

Data collection took place in Katete, Petauke, Sinda, and Lusangazi. Schools were selected based on observable characteristics (e.g., size, distance from district center, prior performance). Grade 5 students were chosen for FGDs to ensure variation in gender and academic performance. Parents of Grade 5 students, including PTA members, were also included. All interviews were conducted in Nyanja or Nsenga dialects, except for those with Impact Network leadership, which were conducted in English.

Table 1. Summary of Qualitative Methods for eSchool 360 Process Evaluation

Method	Respondents	Number
FGD	Students	6
FGD	Parents/PTA members	6
KII	Teachers	12
KII	Teacher supervisors	3
KII	IN staff	10
Classroom observation	N/A	13
Total		50

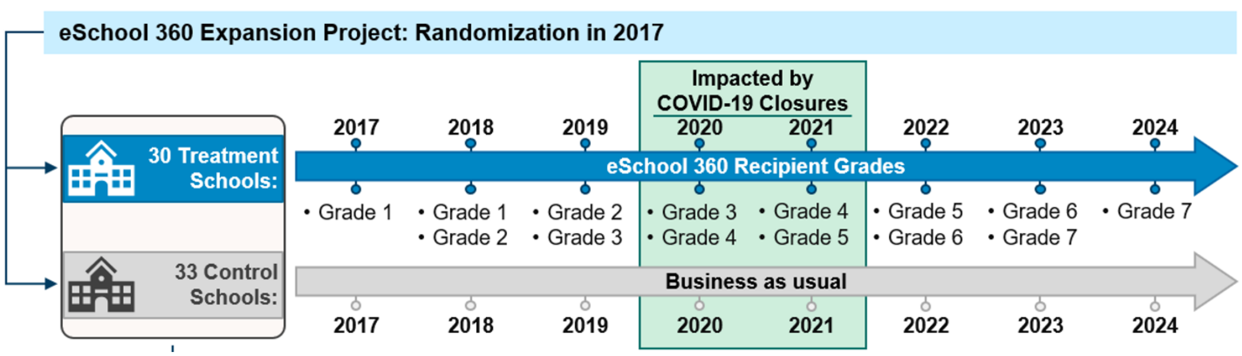
All qualitative data were coded and analyzed using NVivo. A preliminary coding framework was developed based on research questions and interview protocols. Coders double-coded a sample of transcripts to ensure interrater reliability. Using grounded theory, we identified themes and patterns that confirmed or challenged initial hypotheses, characterized response prevalence, and highlighted key findings.

5. Adaptations to Research and Programming Because of School Closures

Due to COVID-19, schools in Zambia were closed for extended periods, disrupting both student enrollment and attendance in community schools. The Zambian government first closed learning institutions in March 2020. Although schools briefly reopened in September 2020, intermittent closures continued until mid-2021, resulting in prolonged interruptions to students' education.

These disruptions significantly extended the study timeline. Originally, endline data collection was planned for late 2020, after children were expected to have completed third grade (de Hoop et al., 2017). In response to the pandemic, we postponed data collection to mid-2022, more than a year after schools had reopened, to ensure that children had sufficient time to return to school before the final round of data collection. **Figure 2** illustrates the project rollout and timeline, highlighting the impact of COVID-19-related school closures.

Figure 2. Summary of Project Rollout and Timeline



This revised timeline allowed us to estimate the four-and-a-half-year impacts of the eSchool 360 program for children eligible to enroll in first grade in 2017 and 2018 and who lived near one of the 63 study schools. These cohorts were targeted because the program was designed to expand to an additional grade each year. The study followed each sampled child regardless of their enrollment trajectory, enabling us to estimate both the ITT effects (the impact of the opportunity to enroll in an Impact Network school) and the TOT effects (the impact on children who enrolled and actively participated in the program.)

However, school closures reduced students' exposure to the intervention. The eSchool 360 model was designed to serve two cohorts, those entering Grade 1 in 2017 and in 2018, and to follow them as they progressed through primary school. Students who repeated grades or enrolled in Impact Network schools after a delay received less exposure to the structured pedagogy curriculum, though they still benefited from school infrastructure.

During school closures, Impact Network adapted its programming to support continued learning. Teachers were tasked with visiting students regularly and distributing “homework packs.” In some cases, they met with small groups of students and provided health guidance related to COVID-19. Teachers reported receiving daily targets for home visits. For instance, a teacher from Petauke stated, “We were given the targets, like each and every day we are supposed to visit about a minimum of five children and a maximum of eight children.”

Parents and students confirmed the usefulness of these visits. As a parent from Katete explained,

> “[Teachers] would go through homes and give the children homework to do and then they would come again the following week to check up on the work that was given to them the previous week.”
>
> – *Katete Parent*

In addition, Impact Network partnered with local radio stations, Mphangwe Radio, Breeze FM, and Radio Explorer, to broadcast educational content. These radio lessons were accessible to both treatment and control group students. However, respondents noted that poor signal strength often interfered with reception. The radio programming was similar to that used by Rising Academies in Sierra Leone, which did not show significant impacts on learning outcomes (Crawfurd et al., 2023).

6. Results

6.1 Descriptive Statistics

We begin by presenting descriptive statistics of the sample before turning to the program’s impact estimates. Specifically, we examine attrition levels, differential attrition between treatment and control groups, and baseline balance. These analyses provide insight into the internal validity of the cluster-randomized controlled trial.

6.1.1 Attrition

Figure 3 presents the CONSORT flow diagram, which outlines the sampling strategy—from eligibility screening and random assignment to the census, age-based stratification, and practical adjustments made during data collection. The diagram also reports attrition rates at midline and endline.

Although endline data collection was originally scheduled for 2020, it was postponed to mid-2022 due to COVID-19-related school closures. This delay allowed children to return to school for over a year before the final survey but also increased the risk of attrition due to the extended gap between baseline and endline.

Despite these challenges, we successfully tracked and surveyed 93% of the original baseline sample (*N* = 1,729 out of 1,865 children), a higher follow-up rate than at midline. Attrition was slightly higher in control areas (9.0%, *N* = 88) than in treatment areas (5.5%, *N* = 48). The main reasons for attrition included: households no longer identifiable in the community (52 cases), household relocation to distant areas (79 cases), and child mortality (5 cases).

These attrition rates are relatively low given the study's duration and the disruptions caused by the pandemic, supporting the credibility of the impact estimates that follow. While the difference in attrition between treatment and control areas is statistically significant, it is small and unlikely to bias the results meaningfully. We did not find statistically significant differences in baseline EGRA, EGMA, ZAT, or oral vocabulary scores between children who remained in the sample and those who were lost to follow-up. **Table 2** presents the results of our differential attrition tests.

To locate households that were initially unidentifiable, we used a combination of three tracking strategies:

1. Snowballing: Asking other community members to help locate missing households
2. GIS-based tracking: Using GPS coordinates collected at baseline and midline
3. Phone follow-up: Contacting households using phone numbers collected during baseline and midline surveys

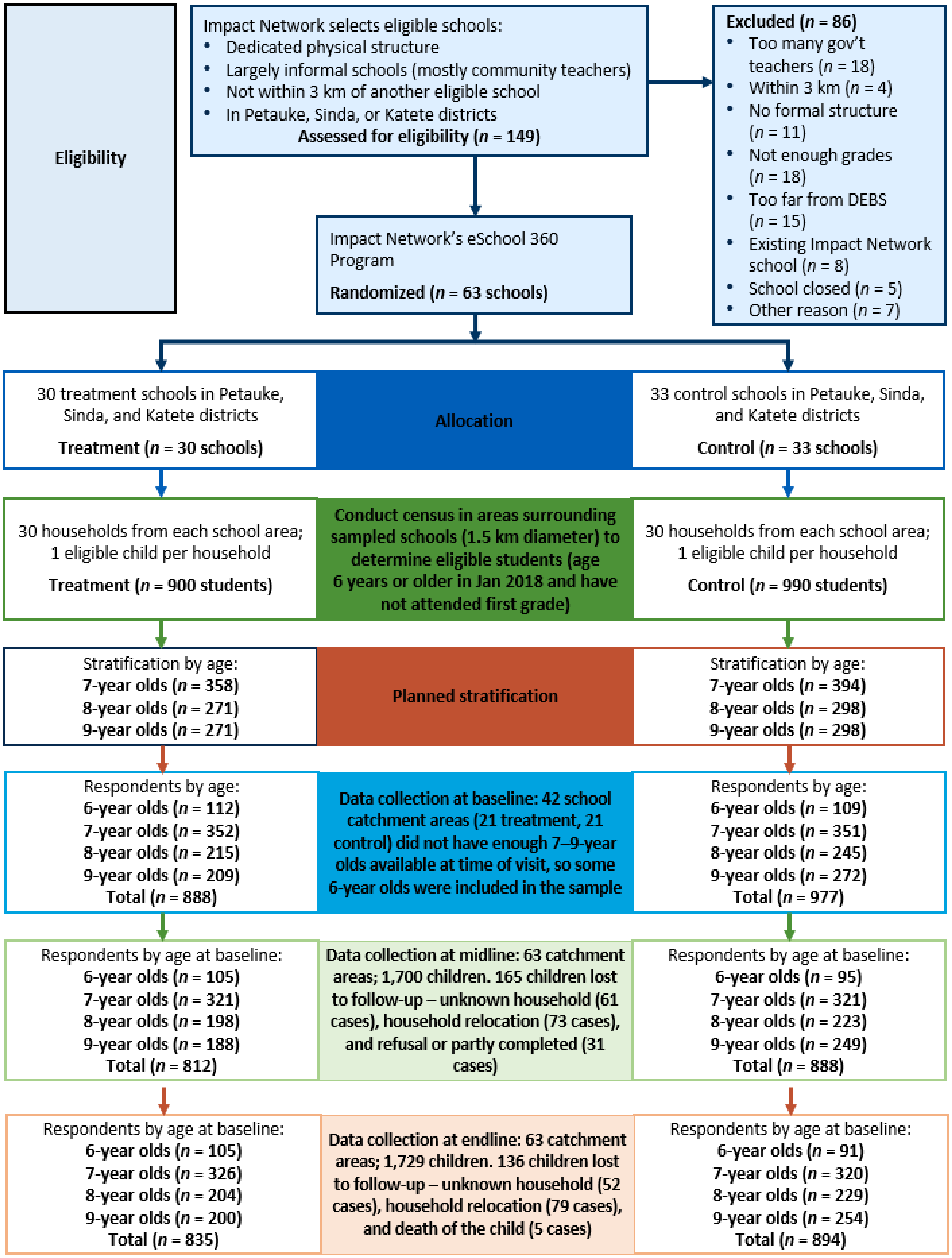
These methods were applied interchangeably and consistently across treatment and control areas to minimize attrition and maintain the integrity of the sample.

Table 2. Differential Attrition

	Attrited		Non-Attrited		T-test	
Variable	***N*/Clusters**	**Mean/(SE)**	***N*/Clusters**	**Mean/(SE)**	**Mean difference**	**Std. difference**
ZAT (proportion correct)	136	0.446	1729	0.446	0.001	0.002
	46	(0.021)	63	(0.011)		
EGRA (proportion correct)	136	0.051	1729	0.051	−0.000	−0.003
	46	(0.003)	63	(0.002)		
EGMA (proportion correct)	136	0.083	1729	0.069	0.014	0.116
	46	(0.014)	63	(0.004)		
Oral vocabulary (proportion correct)	136	0.596	1729	0.608	−0.012	−0.042
	46	(0.026)	63	(0.012)		

Note. Standard errors (SE) clustered at the school level.

Figure 3. Consort Flow Chart



6.1.2 *Balance in Observable Characteristics*

We did not find statistically significant differences in most background characteristics between households in Impact Network (IN) catchment areas and those in control areas for whom we collected both baseline and endline data. Child- and household-level characteristics, including district of residence, household perceptions of poverty, and distance to the nearest school, were similar across groups.

Table 3 presents descriptive statistics on baseline characteristics by treatment status. While a few differences are statistically significant (e.g., age of the index child, caregiver education, and baseline EGMA scores), the standardized differences are small, and the overall pattern suggests that randomization was successful in creating comparable groups.

Table 3. Baseline Equivalence

	Control		Treatment		T-test	
Variable	***N*/ Clusters**	**Mean/ (SE)**	***N*/ Clusters**	**Mean/ (SE)**	**Mean difference**	**Std. difference**
HH very poor at baseline	894	0.496	835	0.469	0.026	0.052
	33	(0.025)	30	(0.024)		
HH food secure at baseline (never had no food in HH in past 4 weeks)	894	0.416	835	0.416	0.001	0.001
	33	(0.037)	30	(0.037)		
HH distance to catchment school at baseline	894	0.834	835	1.063	−0.229	−0.082
	33	(0.076)	30	(0.157)		
Index child was 8 years old or older at baseline	894	0.540	835	0.484	0.056**	0.113
	33	(0.017)	30	(0.022)		
Child is female	894	0.475	835	0.442	0.033	0.067
	33	(0.018)	30	(0.017)		
Caregiver attended school	849	0.588	779	0.653	−0.066**	−0.136
	33	(0.020)	30	(0.026)		
ZAT (proportion correct)	894	0.454	835	0.436	0.018	0.081
	33	(0.014)	30	(0.018)		
EGRA (proportion correct)	894	0.047	835	0.048	−0.000	−0.015
	33	(0.002)	30	(0.002)		
	894	0.078	835	0.060	0.018**	0.170

	Control		Treatment		T-test	
EGMA (proportion correct)	33	(0.007)	30	(0.005)		
Oral vocabulary (proportion correct)	894	0.621	835	0.594	0.027	0.097
	33	(0.018)	30	(0.015)		

Note. Standard errors (SEs) for t-tests clustered at the school level.
** $p < .05$.

6.2 Impact Estimates

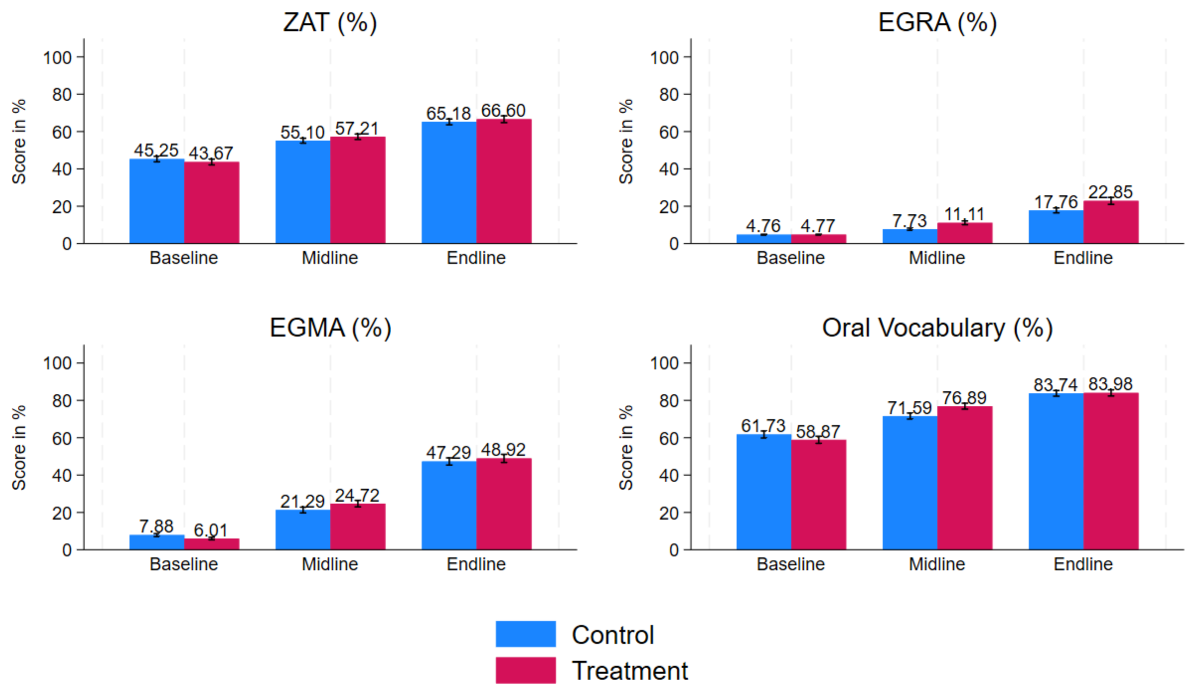
6.2.1 Intent-to-Treat Impacts on Primary Outcomes

Initial comparisons in test scores between the treatment and control groups suggest positive program effects on literacy outcomes. However, differences in mathematics scores were modest, and we observed little to no difference in ZAT and oral vocabulary scores 4.5 years after the program's launch.

Results from the formal ANCOVA framework show positive point estimates across all primary test scores. However, statistically significant effects were consistently observed only for literacy outcomes. Specifically, residing in an Impact Network (IN) catchment area—our treatment assignment—was associated with a 0.26 standard deviation increase in EGRA scores, equivalent to a 5.1-percentage-point gain ($p < .01$), as shown in **Table 4**.

For the other assessments, EGMA, ZAT, and oral vocabulary, the estimated coefficients were positive but not statistically significant. These findings suggest that, by 2022, the differences between treatment and control groups observed at midline had narrowed, including in literacy, where we had previously found larger effects after 14 months of implementation (0.40 standard deviations).

Figure 4. Test Scores by Treatment Status, Baseline, Midline, and Endline



Note. This graph is based on observations we had for all three waves and only supports a visual of trends. *N* = 1,626. ZAT refers to the Zambian achievement test, EGRA to the early grade reading assessment, EGMA to the early grade math assessment, and oral vocabulary to an oral vocabulary test developed by AIR.

Table 4. ITT Impacts on Primary Scores

	ZAT (%)	ZAT (SD)	EGRA (%)	EGRA (SD)	EGMA (%)	EGMA (SD)	Oral vocabulary (%)	Oral vocabulary (SD)
Treatment	0.021	0.089	0.051***	0.263***	0.024	0.084	0.011	0.048
	(0.015)	(0.064)	(0.015)	(0.075)	(0.018)	(0.063)	(0.014)	(0.059)
Comparison mean	0.646	0.646	0.175	0.175	0.468	0.468	0.832	0.832
N	1729	1729	1729	1729	1729	1729	1729	1729
Unadjusted *p* value	0.17	0.17	0.00	0.00	0.19	0.19	0.42	0.42
RIT adjusted *p* value	0.18	0.18	0.00	0.00	0.21	0.21	0.44	0.44

Note. Clustered standard errors in parentheses. Estimations control for district fixed effects, baseline outcomes, age at baseline, distance from school at baseline, and baseline EGMA score.
$^*p < .1$. $^{**}p < .05$. $^{***}p < .01$.

The ITT estimates show statistically significant impacts, exceeding 0.20 standard deviations, on 6 out of 8 EGRA subscales (**Table 5**). While positive point estimates were observed for 8 of the 10 EGMA subscales (**Table 6**), only three (Pattern Completion, Addition Questions 2, and Subtraction Questions 2) were statistically significant. Addition Questions 2 and Subtraction Questions 2 are relatively difficult, suggesting that the eSchool 360 model may enhance performance on more challenging EGMA tasks. However, as shown in **Tables A1 and A2** in the Appendix, these results did not remain robust after correcting for multiple comparisons using the Benjamini-Hochberg method, as indicated by the adjusted *q*-values.

Several factors may explain the absence of significant impacts on non-EGRA outcomes. First, COVID-19-related school closures and the resulting disruptions likely reduced students' exposure to the intervention and affected treatment compliance. Second, ceiling effects may have limited the ability to detect gains, particularly for the Zambian Achievement Test and oral vocabulary assessments. Third, the control group may have caught up after the midline data collection.

Table 5. ITT Impacts on EGRA Subscores

	Orientation to Print	Letter Sound Knowledge	Nonword Decoding	Oral Passage Reading 1	Reading Comprehension 1	Oral Passage Reading 2	Reading Comprehension 2	Listening Comprehension
Treatment	−0.007	0.177**	0.273***	0.289***	0.321***	0.292***	0.310***	0.010
	(0.056)	(0.080)	(0.078)	(0.072)	(0.064)	(0.061)	(0.064)	(0.072)
N	1729	1710	1729	1729	1729	1729	1729	1729

Note. Clustered standard errors in parentheses. Statistical significance was robust to Benjamini-Hoschberg correction to *p* values. Estimations control for district fixed effects, baseline outcomes, age at baseline, distance from school at baseline, and baseline EGMA score.
$^*p < .1$. $^{**}p < .05$. $^{***}p < .01$.

Table 6. ITT Impacts on EGMA Subscores

	Oral Counting	Rational Counting	Number Recognition	Quantity Discrimination	Pattern Completion	Word Problems	Addition Questions 1	Addition Questions 2	Subtraction Questions 1	Subtraction Questions 2
Treatment	−0.004	0.083	0.104	0.018	0.139**	0.032	−0.035	0.144**	0.037	0.132**
	(0.044)	(0.073)	(0.068)	(0.056)	(0.065)	(0.059)	(0.064)	(0.062)	(0.061)	(0.064)
N	1729	1729	1729	1729	1729	1729	1729	1729	1729	1729

Note. Clustered standard errors in parentheses. While there are three estimates that are statistically significant at 5% alpha, none of these were statistically significant after Benjamini-Hoschberg correction to *p* values. Estimations control for district fixed effects, baseline outcomes, age at baseline, distance from school at baseline, and baseline EGMA score.
$^*p < .1$. $^{**}p < .05$. $^{***}p < .01$.

6.2.2 *Heterogeneous Effects by Demographic Characteristics*

We estimated treatment effects by subgroups and found evidence of variation by gender and baseline assessment scores. Children who scored above the median on the ZAT, oral vocabulary, or EGMA at baseline experienced positive and statistically significant impacts on ZAT and EGMA outcomes, in addition to EGRA. One possible explanation for these larger effects is that higher-scoring students were more likely to return to school following COVID-19-related closures. Analyses of enrollment rates support this, showing that children with higher baseline scores were significantly more likely to reenroll postpandemic. However, the effects do not remain statistically significant after correcting for multiple hypothesis testing, as shown in Appendix B.

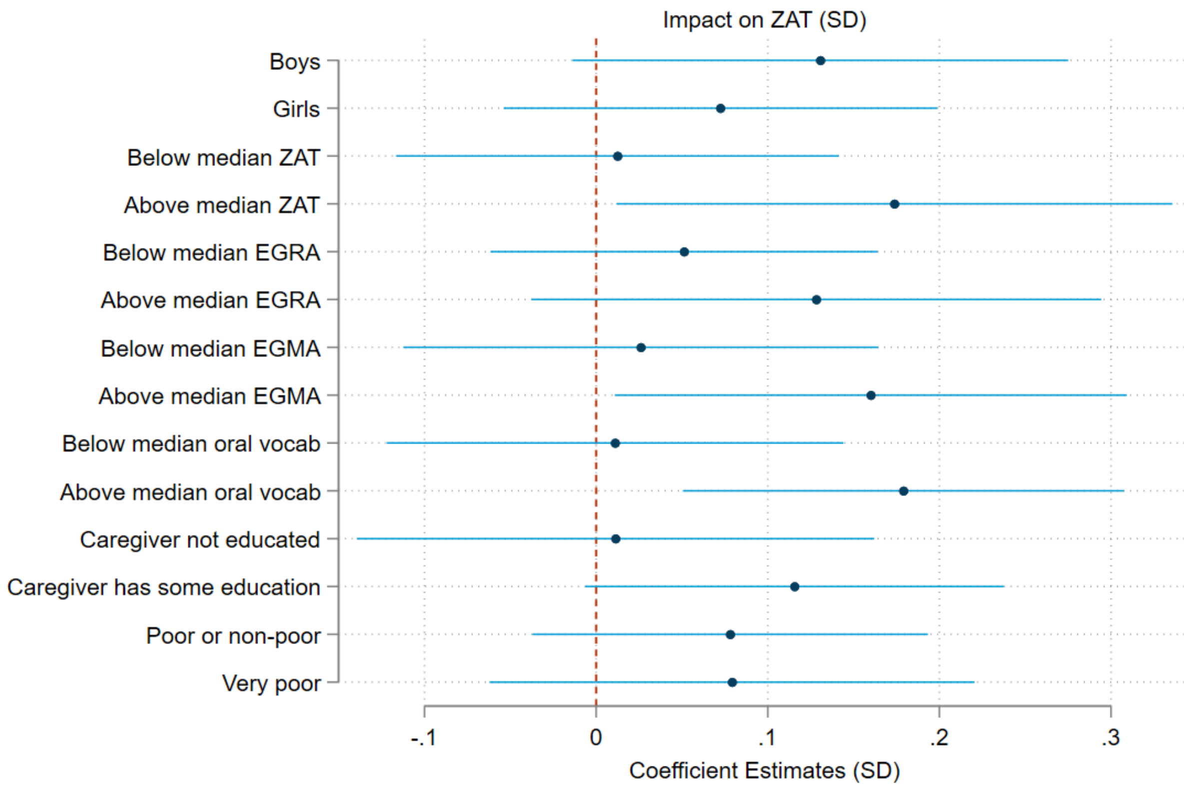
We also observed meaningful and statistically significant treatment effects for boys: 0.28 SD on EGRA, 0.20 SD on EGMA, and 0.15 SD on oral vocabulary. Importantly, boys were not overrepresented in the subgroup with above-median baseline scores. While 55% of the full sample is male, only 49% of those scoring above the median on EGMA at baseline are male. This suggests that the gender-based heterogeneity in impacts is distinct from the heterogeneity based on baseline performance.

The gender differences are notable and somewhat unexpected, given the typically lower enrollment and achievement rates among boys in the Zambian context. Higher dropout rates among boys would generally be expected to reduce effect sizes. However, while the treatment effects for boys are substantial, they do not remain statistically significant after corrections for multiple hypothesis testing, as shown in Appendix B.

We did not find strong evidence of other subgroup heterogeneities. Nonetheless, these findings should be interpreted with caution, as the study may lack sufficient statistical power to detect smaller differences in effect sizes across subgroups.

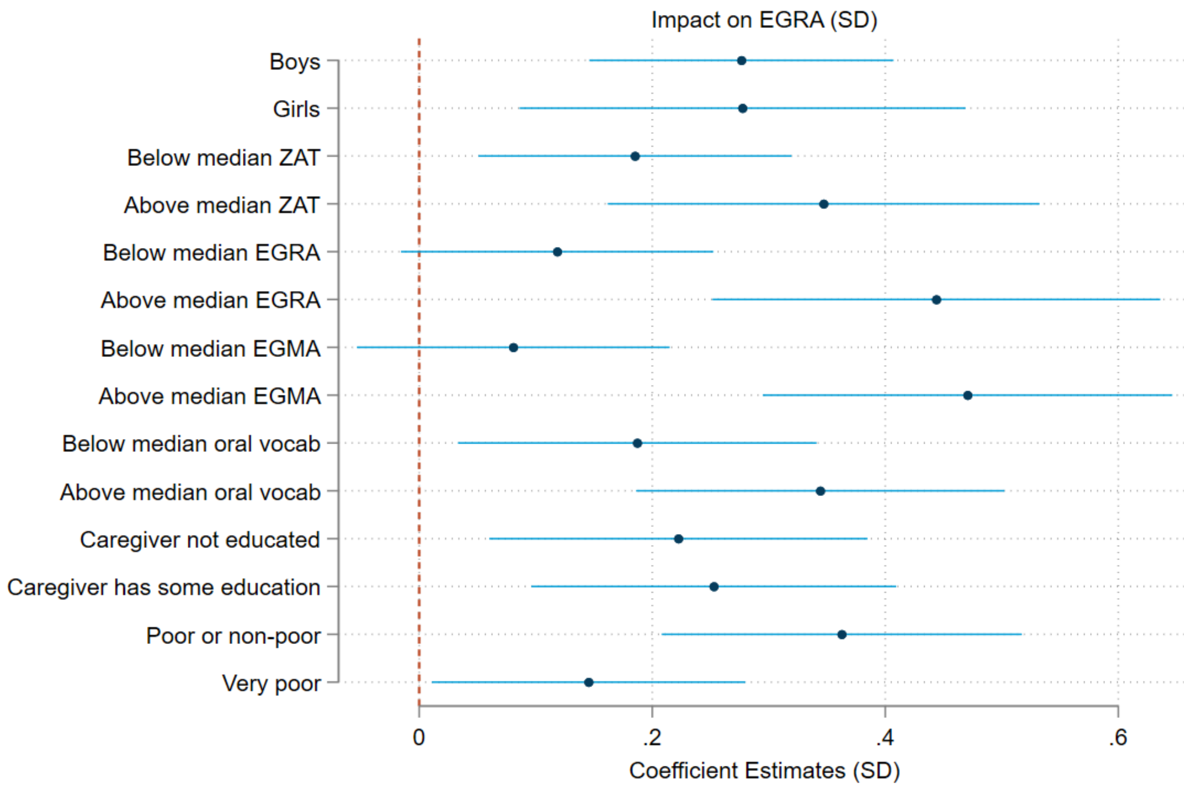
These subgroup results are illustrated in **Figures 5–8**, which show treatment effects by baseline performance and gender for each of the four primary outcomes: ZAT, EGRA, EGMA, and oral vocabulary.

Figure 5. Heterogeneous Effects—ZAT



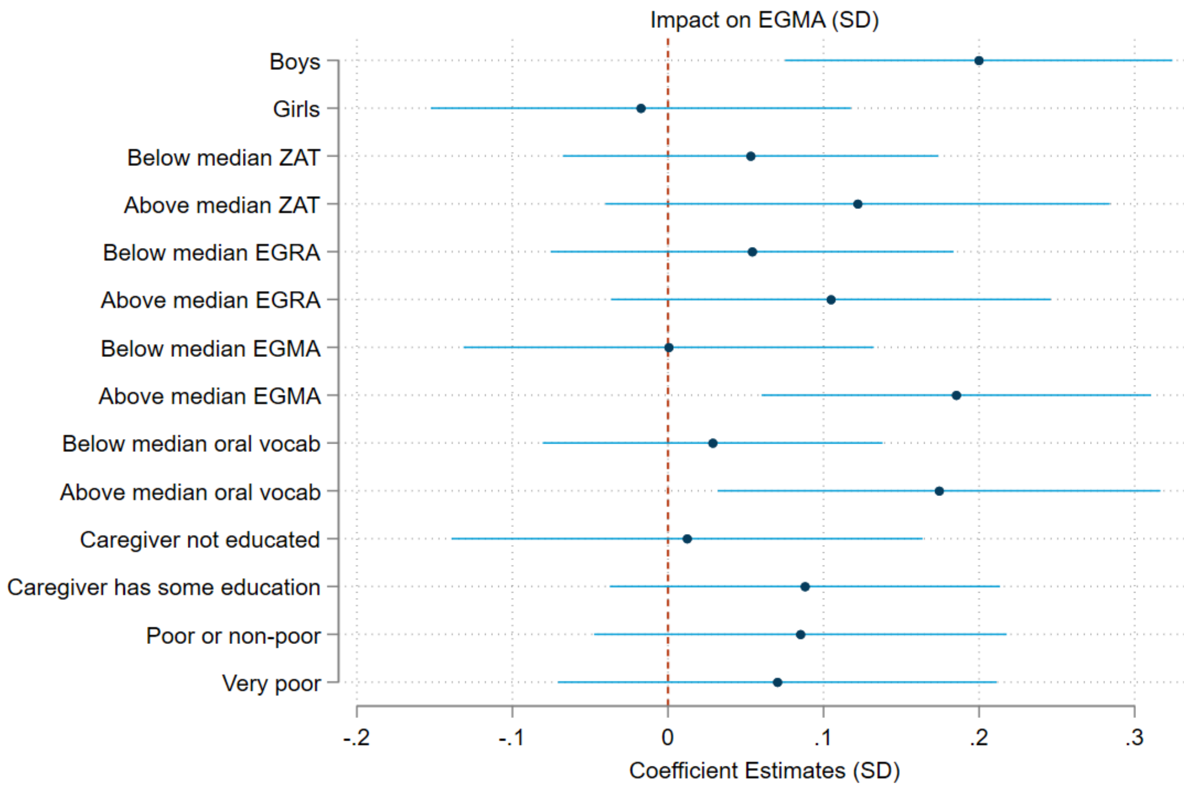
Note. Coefficient estimates show ITT effects on different subsamples indicated on the y-axis. Red bars indicate 90% confidence intervals.

Figure 6. Heterogeneous Effects—EGRA



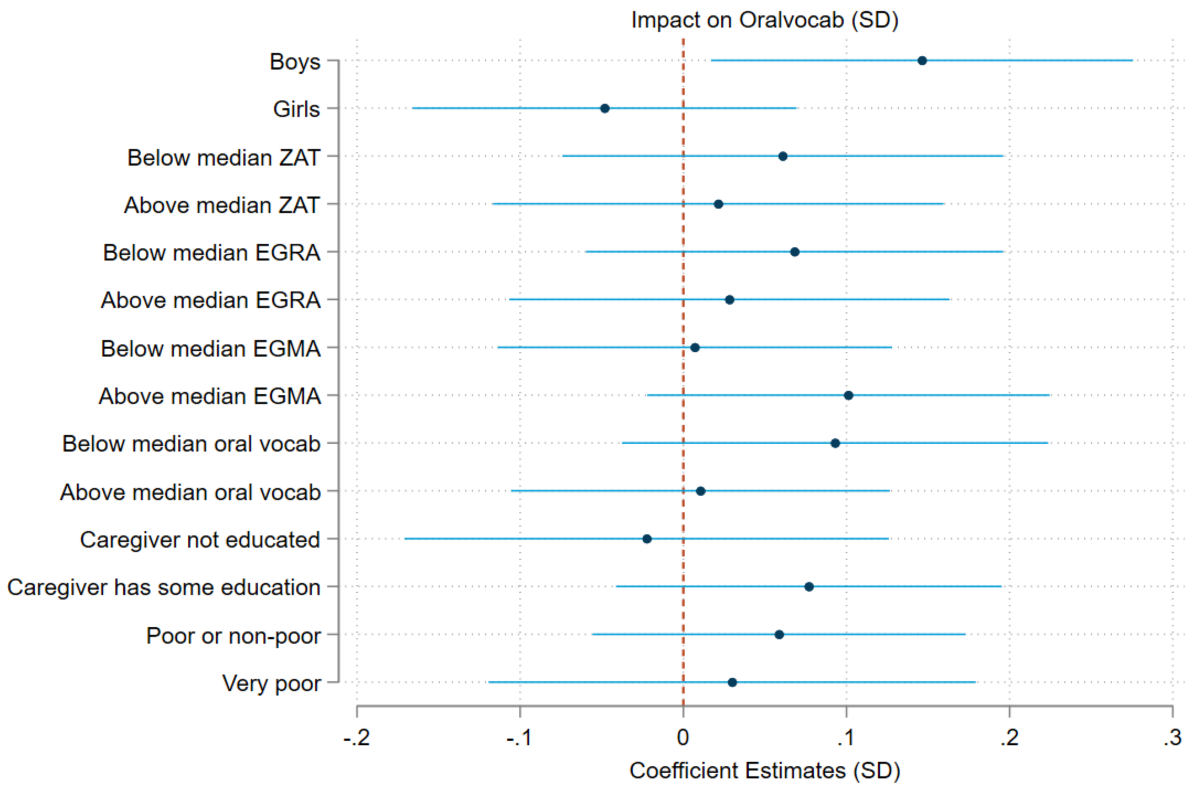
Note. Coefficient estimates show ITT effects on different subsamples indicated on the y-axis. Red bars indicate 90% confidence intervals.

Figure 7. Heterogeneous Effects—EGMA



Note. Coefficient estimates show ITT effects on different subsamples indicated on the y-axis. Red bars indicate 90% confidence intervals.

Figure 8. Heterogeneous Effects—Oral Vocabulary



Note. Coefficient estimates show ITT effects on different subsamples indicated on the y-axis. Red bars indicate 90% confidence intervals.

6.2.3 Mechanisms of Change

Various demand-side and supply-side mechanisms may have contributed to the sustained impacts on EGRA outcomes and the reduced impacts on EGMA, ZAT, and oral vocabulary scores. Demand-side mechanisms relate to effects on school enrollment and attendance, while supply-side mechanisms are associated with improvements in the quality of education.

Before the Pandemic: The study showed strong evidence for increased demand for education caused by the eSchool 360 program after 14 months of implementation. Treatment students were 8 percentage points more likely to enroll in school, and the program led to an increase of 0.36 days of weekly attendance, on average. Age at enrollment also decreased significantly. Qualitative research revealed that children were motivated to attend class due to interest in using tablets and technology.

We also found strong evidence for improvements in education quality. Caregivers in treatment areas reported a 0.29-point higher level of satisfaction with education than caregivers in control areas ($p < .05$). Qualitative data also indicated higher teacher attendance and positive perceptions of teacher quality in Impact Network schools.

After School Closures: We no longer found impacts on school enrollment and attendance after 4.5 years. ITT effects on these outcomes were statistically insignificant, and the point estimate for attendance was negative, as shown in **Table 7**.

Table 7. ITT Impacts on School Enrollment and Attendance

	Index child is enrolled at any school (last year or this year)	Weekly attendance (number of days attended)
Treatment	−0.019	−0.019
	(0.027)	(0.144)
Comparison mean	0.679	2.621
N	1729	1729

Note. Clustered standard errors in parentheses. Estimations control for district fixed effects, age at baseline, distance from school at baseline, and baseline EGMA score.
$^{*}p < .1$. $^{**}p < .05$. $^{***}p < .01$.

COVID-19 and the associated school closures changed the mechanisms through which the program affected demand. While we did not find statistically significant impacts on parents' reported enrollment decisions, as shown in **Table 8**, students with lower baseline scores were significantly less likely to enroll in school at endline, strongly suggesting that they were also more likely to return to school. Among those who were enrolled at midline, children scoring below median on all scores, and particularly children with lower ZAT and oral vocabulary scores, were more likely to drop out by endline. In addition, over 12% of boys enrolled at midline dropped out by endline, compared to 9.7% of girls. This changed the composition of children who benefited from the program.

However, treatment students were more likely to attend Grade 5 or higher after 4.5 years—12 percentage points more likely than control students. This pattern held for both boys and girls, though boys were less likely overall to reach higher grades. These results are presented in **Table 8**.

Table 8. COVID-19 Impacts

	Decision to enroll child at school affected by COVID-19	Delayed or stopped school because of COVID-19	Grade 5 or 6 by endline
Treatment	−0.038	−0.016	0.118***
	(0.026)	(0.016)	(0.024)
Comparison mean	0.318	0.100	0.162
N	1724	1724	1616

Note. Clustered standard errors in parentheses. Estimations control for district fixed effects, age at baseline, distance from school at baseline, and baseline EGMA score.
*$p < .1$. **$p < .05$. ***$p < .01$.

Despite the lack of enrollment effects, the program continued to deliver high-quality education. Parents in treatment areas reported significantly higher satisfaction with schools and greater perceptions of their children's engagement. These findings are shown in **Table 9**. Qualitative interviews echoed these results, with students, teachers, and parents praising the quality of instruction and the learning environment. For instance, a parent from Chatata school noted, "I think the children here at Impact Network schools are well ahead of the other schools in terms of how the children learn."

Table 9. ITT Impacts on Caregiver Satisfaction and Parent Engagement

	Scale for caregiver's satisfaction on teachers, staff, and school structure	Caregiver perception of child engagement scale
Treatment	0.197**	0.595***
	(0.081)	(0.220)
Comparison mean	2.364	8.341
N	1055	1055

Note. Clustered standard errors in parentheses. Estimations control for district fixed effects, age at baseline, distance from school at baseline, and baseline EGMA score.
*$p < .1$. **$p < .05$. ***$p < .01$.

These findings, combined with the statistically insignificant effects on school attendance and enrollment, suggest that the higher quality of education in Impact Network schools likely contributed to improvements in literacy. Respondents in qualitative interviews, including students, teachers, and parents, voiced consistently positive perceptions of education quality in Impact Network schools.

Students frequently cited the quality of their teachers, noting that teachers paused to explain materials and checked for understanding before moving on. Teachers emphasized the design of the Impact Network program and the quality of the school environment. One teacher explained, “Impact Network lessons are well designed. … You’ll follow the laid down plans, you have the schemes of work, they are well aligned, you follow them.”

Another teacher highlighted how the program’s training and behavioral guidelines created a supportive learning environment: “We are looking at the rules which the Impact Network is giving us as teachers on how to handle Impact Network learners in short … We are even guided in which punishment to give them, there is no beating, so they have … It is a conducive environment for them.”

As a result, many teachers interviewed in this study believed that Impact Network students learned quickly and performed well, especially in comparison to students in other schools.

6.2.4 Treatment Effects on the Treated

As expected, treatment-on-the-treated (TOT) effects, defined as attending school at least three times in the week prior to the survey, are substantially larger than the ITT effects. Specifically, attending an Impact Network school for at least 3 days in the past week improved EGRA outcomes by 10 percentage points, or 0.51 standard deviations. However, impact estimates for non-EGRA assessments remain statistically insignificant, as shown in **Table 10**. The point estimate for EGMA is a meaningful 0.16 standard deviations, but the standard error is large, likely due to reduced statistical power.

Similarly, estimates of the impact of having ever enrolled in an Impact Network school during the year prior to the survey show positive effects on EGRA but statistically insignificant effects on other outcomes. Enrollment at endline increased EGRA scores by 8.1 percentage points, or 0.42 standard deviations. These results are presented in **Table 11**. The first-stage estimate shows that assignment to an Impact Network school increased the likelihood of enrollment by 63 percentage points, providing a strong instrument for the second-stage analysis.

One reason for the smaller ITT impacts is the relatively low program dosage among children in the treatment group. Because the program was implemented in only two cohorts—consistent with the original analysis plan that anticipated an endline study after 3 years—students who repeated grades or enrolled late did not fully benefit from the eSchool 360 model. In fact, only 49% of sampled children in treatment areas ever enrolled in an Impact Network school, and just 20% received more than 4 years of program exposure.

The effects for students who enrolled in school and those who attended regularly were often larger, but the point estimates were not always statistically significant. We find sizeable and statistically significant effects on EGRA outcomes for those who enrolled and regularly attended school, but the effects on EGMA, ZAT, and oral vocabulary scores were generally not statistically significant for these subgroups, despite the larger point estimates.

While the results are vulnerable to weak instrument bias, we did find sizeable and statistically significant effects on both EGRA and EGMA outcomes for students who enrolled in Impact Network classes for the full duration of the program (approximately 20% of the treatment group). These students also had lower rates of grade repetition compared to their peers in control schools, which may have contributed to the larger observed impacts.

Table 10. TOT Impacts on Primary Scores—IN Attendance

	Weekly attendance IN school 3+ days	ZAT (%)	ZAT (SD)	EGRA (%)	EGRA (SD)	EGMA (%)	EGMA (SD)	Oral vocabulary (%)	Oral vocabulary (SD)
Treatment	0.513***								
	(0.022)								
3+ days weekly attendance		0.040	0.173	0.100***	0.515***	0.047	0.163	0.022	0.093
		(0.028)	(0.120)	(0.026)	(0.135)	(0.033)	(0.117)	(0.026)	(0.111)
Comparison mean		0.614	0.614	0.156	0.156	0.420	0.420	0.815	0.815
N		1729	1729	1729	1729	1729	1729	1729	1729

Note. Clustered standard errors in parentheses. Estimations control for district fixed effects, baseline outcomes, age at baseline, distance from school at baseline, and baseline EGMA score.
**p* < .1. ***p* < .05. ****p* < .01.

Table 11. TOT Impacts on Primary Scores—IN Enrollment

	Enrolled in an IN school	ZAT (%)	ZAT (SD)	EGRA (%)	EGRA (SD)	EGMA (%)	EGMA (SD)	Oral vocabulary (%)	Oral vocabulary (SD)
Treatment	0.632***								
	(0.025)								
Enrolled in an IN school		0.032	0.140	0.081***	0.418***	0.038	0.133	0.017	0.076
		(0.023)	(0.098)	(0.022)	(0.112)	(0.027)	(0.096)	(0.021)	(0.090)
Comparison mean		0.608	0.608	0.150	0.150	0.414	0.414	0.814	0.814
N		1729	1729	1729	1729	1729	1729	1729	1729

Note. Clustered standard errors in parentheses. Estimations control for district fixed effects, baseline outcomes, age at baseline, distance from school at baseline, and baseline EGMA score.
**p* < .1. ***p* < .05. ****p* < .01.

7. Conclusion

Using a 5-year cluster-randomized controlled trial, this study evaluates the long-term effects of the eSchool 360 program in rural Zambia, with a focus on post-COVID-19 learning recovery. Literacy gains persisted at 0.26 standard deviations, though the magnitude was lower than prepandemic estimates. In contrast, improvements in mathematics were concentrated among boys and students with above-median baseline scores, but these effects did not remain statistically significant after adjusting for multiple comparisons.

These findings suggest that structured pedagogy programs like the eSchool 360 model can partially mitigate pandemic-related learning losses, particularly in foundational literacy. However, the benefits were limited to students who returned to school after closures—those who remained out of school did not experience gains. This underscores the importance of pairing instructional interventions with strategies to reengage out-of-school children. The absence of robust improvements in mathematics and the heterogeneity of impacts raise important questions about how such programs might perform in other sub-Saharan African contexts, where barriers to school reentry and baseline learning levels may differ.

Nonetheless, the qualitative pattern of results aligns with findings from a large-scale panel study in India, where remedial education incorporating structured pedagogy supported postpandemic learning recovery (Singh et al., 2024). Evidence from both contexts suggests that structured pedagogy may offer a scalable approach to addressing foundational learning deficits in low-resource settings. However, the extent of its impact will depend on reenrollment rates and the ability to reach students most at risk of being left behind.

Structured pedagogy programs were not developed in response to COVID-19. They have long been proposed as a solution to the global learning crisis, particularly in sub-Saharan Africa, where instructional quality and teacher support are often limited. Rigorous evidence from multiple settings (e.g., Eble et al., 2021; Eble & Escueta, 2023; Glewwe & Muralidharan, 2016; Kerwin & Thornton, 2021; Rodriguez-Segura, 2022; Snilstveit et al., 2016) supports their effectiveness. Our findings reinforce their relevance not only for pandemic recovery, but also as a broader strategy to improve foundational learning outcomes at scale.

While this study focused primarily on learning outcomes, complementary cost-effectiveness analysis suggests that structured pedagogy programs can deliver substantial gains at relatively low cost. Our estimates indicate that for every $100 invested, the program generated between 0.35 and 1.40 standard deviation improvements in literacy by endline, depending on the scenario. These figures are based on 5 years of program expenditure, expressed in 2022 USD, when the endline assessments were conducted. Given constrained education budgets in many low- and middle-income countries, future research should prioritize multi-arm cluster-randomized trials to identify the most cost-effective combinations of instructional and reengagement strategies. Although this study lacked sufficient scale to implement a multi-arm design, such evidence is essential to guide investment decisions in the wake of the pandemic and amid compounding crises such as conflict and climate shocks. To improve external validity, it is also critical to examine the impacts of the eSchool 360 model in government schools. Such analyses will help to determine the scalability of the program.

8. References

Angrist, N., Bergman, P., & Matsheng, M. (2022). Experimental evidence on learning using low-tech when school is out. *Nature Human Behaviour, 14*(7), 941–950. https://doi.org/10.1038/s41562-022-01381-z

Ale, K., Loh, Y. A. C., & Chib, A. (2017). Contextualized OLPC education project in rural India: Measuring learning impact and mediation of computer self-efficacy. *Educational Technology Research and Development*, *65*(3), 769–794. https://doi.org/10.1007/s11423-017-9517-2

Ardington, C., Wills, G., & Kotze, J. (2021). COVID-19 learning losses: Early grade reading in South Africa. *International Journal of Educational Development*, *86,* 102480.

Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parienté, W., Shapiro, J., Thuysbaert, B., & Udry, C. (2015). A multifaceted program causing lasting progress for the very poor: Evidence from six countries. *Science (New York, N.Y.)*, *348*(6236), 1260799–1260799. https://doi.org/10.1126/science.1260799

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., & Walton, M. (2017). From proof of concept to scalable policies: Challenges and solutions, with an application. *Journal of Economic Perspectives, 31*(4), 73–102. https://doi.org/10.1257/jep.31.4.73

Barrera-Osorio, F., & Linden, L. L. (2009). *The use and misuse of computers in education: Evidence from a randomized experiment in Colombia* (World Bank Policy Research Working Paper No. 4836). http://nrs.harvard.edu/urn-3:HUL.InstRepos:8140109

Betthäuser, B. A., Bach-Mortensen, A. M., & Engzell, P. (2023). A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic. *Nature Human Behaviour, 7*(4), 375–385. https://doi.org/10.1038/s41562-022-01506-4

Crawfurd, L., Hares, S., Minardi, A. L., & Sandefur, J. (2023). Radio education during COVID-19: Evidence from Sierra Leone*. Journal of Development Economics, 166*, Article 102148.

De Hoop, T., Brudevold-Newman, A., Ring, H., Zanolini, A., Nowlin, C., & Davis, D. (2017). Inception report for the mixed-methods cluster-randomized controlled trial of Impact Network's eSchool 360 model in rural Zambia. https://www.air.org/sites/default/files/2021-06/Inception-Report-eSchool-360-Model-Rural-Zambia-Sept-2017.pdf de Hoop, T., Ring, H., Siwach, G., Dias, P., Tembo, G., Rothbard, V., & Toungui, A. (2024). Impact of technology-aided activity-based learning approaches on learning outcomes: Experimental evidence from community schools in rural Zambia. *Journal of Research on Educational Effectiveness, 17*(4), 836–873. https://doi.org/10.1080/19345747.2023.2268072

Di Pietro, G. (2023). The impact of Covid-19 on student achievement: Evidence from a recent meta-analysis. *Educational Research Review, 39,* 100530.

Duflo, A., Kiessel, J., & Lucas, A. M. (2024). Experimental evidence on four policies to increase learning at scale. *Economic Journal, 134*(661), 1985–2008. https://doi.org/10.1093/ej/ueae003

Eble, A., Frost, C., Camara, A., Bouy, B., Bah, M., Sivaraman, M., Pei-Tseng, J. H., Jayanty, C., Brady, T., Gawron, P., Vansteelandt, S., Boone, P., & Elbourne, D. (2021). How much can we remedy very low learning levels in rural parts of low-income countries? Impact and generalizability of a multi-pronged para-teacher intervention from a cluster-randomized trial in the Gambia. *Journal of Development Economics*, *148*, 102539. https://doi.org/10.1016/j.jdeveco.2020.102539

Eble, A., & Escueta, M. (2023). *When your bootstraps are not enough: How demand and supply interact to generate learning in settings of extreme poverty* (NBER Working Paper No. 31388). National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w31388/w31388.pdf

Eble, A., Frost, C., Camara, A., Bouy, B., Bah, M., Sivaraman, M., Pei-Tseng, J. H., Jayanty, C., Brady, T., Gawron, P., Vansteelandt, S., Boone, P., & Elbourne, D. (2021). How much can we remedy very low learning levels in rural parts of low-income countries? Impact and generalizability of a multi-pronged para-teacher intervention from a cluster-randomized trial in the Gambia. *Journal of Development Economics, 148*, Article 102539. https://doi.org/10.1016/j.jdeveco.2020.102539

Evans, D., & Mendez Acosta, A. (2021). Education in Africa. What are we learning? *Journal of African Economies, 30*(1), 13–54. https://doi.org/10.1093/jae/ejaa009

Filmer, D., & Pritchett, L. (2001). Estimating wealth effects without expenditure data—Or tears: An application to educational enrollments in states in India. *Demography*, *38*(1), 115–132. https://doi.org/10.2307/3088292

Ganimian, A., & Murnane, R. (2016). Improving education in developing countries: Lessons from rigorous impact evaluations. *Review of Educational Research*, *86*(3), 719–755. https://doi.org/10.3102/0034654315627499

Glewwe, P., & Muralidharan, K. (2016). Chapter 10 - Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications. *Handbook of the Economics of Education*, *5*, 653–743.

Goldhaber, D., Kane, T.J., McEachin, A., Morton, E., Patterson, T., & Staiger, D.O. (2023). The educational consequences of remote and hybrid instruction during the pandemic. *American Economic Review: Insights, 5*(3), 377-392.

Gray-Lobe, G., Keats, A., Kremer, M., Mbiti, I., & Ozier, O. (2022). *Can education be standardized? Evidence from Kenya* (Working Paper No. 2022-68). University of Chicago, Becker Friedman Institute for Economics. https://bfi.uchicago.edu/wp-content/uploads/2022/06/BFI_WP_2022-68-1.pdf

Hassan, H., Islam, A., Siddique, A., & Wang, L.C., 2021. Telementoring and Homeschooling During School Closures: A Randomized Experiment in Rural Bangladesh. Technical Report.

Heß, S. (2017). Randomization inference with Stata: A guide and software. *The Stata Journal: Promoting Communications on Statistics and Stata*, *17*(3), 630–651. https://doi.org/10.1177/1536867X1701700306

Moholwane, N., Taylor, S., & Shepherd, D. (2023). Reading skills transfer best from home language to a second language: Policy lessons from two field experiments in South Africa. *Journal of Research on Educational Effectiveness, 17*(4), 687–710. https://doi-org.air.idm.oclc.org/10.1080/19345747.2023.2279123

Muralidharan, K., Singh, A., & Ganimian, A. J. (2019). Disrupting education? Experimental evidence on technology-aided instruction in India. *American Economic Review*, *109*(4), 1426–1460. https://doi.org/10.1257/aer.20171112

Piper, B., Zuilkowski, S. S., & Mugenda, A. (2015). PRO-POOR PRIMR: Improving early literacy skills for children from low-income families in Kenya. *Africa Education Review, 12*(1), 67–87. https://doi.org/10.1080/18146627.2015.1036566

Piper, B., Jepkemei, E., & Kibukho, K. (2015). Pro-poor Primr: Improving early literacy skills for children from low-income families in Kenya. *Africa Education Review*, *12*(1), 67–87. https://doi.org/10.1080/18146627.2015.1036566

Piper, B., Ralaingita, W., Mejia, J., Dubeck, M., DeStefano, J., Stern, J., Jordan, R., & Sitabkhan, Y. (n.d.). *Structured pedagogy: How-to guides and literature review*. Science of Teaching. https://docs.opendeved.net/lib/754EK7TT/download/JU4VK9TN/Structured%20Pedagogy%20How-To%20Guides%20and%20Literature%20Review..pdf

Radhakrishnan, K., Sabarwal, S., Sharma, U., Cullen, C., Crossley, C., & Angrist, N. 2021. Remote Learning: Evidence from Nepal during COVID-19. World Bank Policy Brief.

Rodriguez-Segura, D. (2022). EdTech in developing countries: A review of the evidence. *World Bank Research Observer, 37*(2), 171–203. https://doi.org/10.1093/wbro/lkab011

Schueler, B.E., & Rodriguez-Segura, D., 2021. A Cautionary Tale of Tutoring Hard-to-Reach Students in Kenya. Technical Report.

Singh, A., Romero, M., & Muralidharan, K. (2024). Learning recovery after COVID-19: Experimental evidence from a large-scale panel study in India. *Economic Journal, 134*(661), 2009–2036.

Snilstveit, B., Stevenson, J., Menon, R., Phillips, D., Gallagher, E., Geleen, M., Jobse, H., Schmidt, T., & Jimenez, E. (2016). The impact of education programmes on learning and school participation in low- and middle-income countries: A systematic review summary report. *3ie Systematic Review Summary 7*, 1–40. https://www.3ieimpact.org/evidence-hub/publications/systematic-review-summaries/impact-education-programmes-learning-school-participation-low-and-middle-income-countries

Tauson, M., & Stannard, L. (2018). *EdTech for learning in emergencies and displaced settings: A rigorous review and narrative synthesis*. Save the Children. https://resourcecentre.savethechildren.net/pdf/edtech-learning.pdf

Valiente, O. (2010). *1-to-1 in education: Current practice, international comparative research evidence and policy implications* (OECD Education Working Paper No. 44). Organisation for Economic Co-operation and Development. https://www.oecd.org/content/dam/oecd/en/publications/reports/2010/03/1-1-in-education_g17a1e15/5kmjzwfl9vr2-en.pdf

Young, A. (2019). Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *Quarterly Journal of Economics, 134*(2), 557–598. https://doi.org/10.1093/qje/qjy029

Appendix A. Adjusted *p* Values for Multiple Comparisons in EGRA and EGMA Subscores

Table A1. Adjusted Standard Errors for EGRA Subscores (standard deviation)

EGRA subscore	Coefficient	Std. error	Unadjusted *p* value	Adjusted *q* value
Orientation to Print	–0.007	0.056	0.900	0.900
Letter Sound Knowledge	0.177	0.080	0.031	0.042
Nonword Decoding	0.273	0.078	0.001	0.001
Oral Passage Reading 1	0.289	0.072	0.000	0.000
Reading Comprehension 1	0.321	0.064	0.000	0.000
Oral Passage Reading 2	0.292	0.061	0.000	0.000
Reading Comprehension 2	0.310	0.064	0.000	0.000
Listening Comprehension	0.010	0.072	0.894	0.900

Table A2. Adjusted Standard Errors for EGMA Subscores (standard deviation)

EGMA subscore	Coefficient	Std. error	Unadjusted *p* value	Adjusted *q* value
Oral Counting	–0.004	0.044	0.926	0.926
Rational Counting	0.083	0.073	0.261	0.523
Number Recognition	0.104	0.068	0.131	0.327
Quantity Discrimination	0.018	0.056	0.753	0.837
Pattern Completion	0.139	0.065	0.037	0.143
Word Problems	0.032	0.059	0.595	0.744
Addition Questions (Level 1)	–0.035	0.064	0.587	0.744
Addition Questions (Level 2)	0.144	0.062	0.022	0.143
Subtraction Questions (Level 1)	0.037	0.061	0.553	0.744
Subtraction Questions (Level 2)	0.132	0.064	0.043	0.143

Appendix B. Adjusted *p* Values for Multiple Comparisons in Heterogeneous ITT Effects

Table B1. Adjusted q Values for Heterogeneous Effects—Standardized ZAT Score

Subgroup	Coefficient	Std. Error	Unadjusted *p* value	Adjusted *q* value
Boys	0.131	0.086	0.136	0.385
Girls	0.072	0.076	0.342	0.545
Below median ZAT at baseline	0.012	0.077	0.873	0.913
Above median ZAT at baseline	0.174	0.097	0.078	0.385
Below median EGRA at baseline	0.051	0.068	0.451	0.639
Above median EGRA at baseline	0.128	0.099	0.202	0.472
Below median EGMA at baseline	0.026	0.083	0.755	0.913
Above median EGMA at baseline	0.160	0.089	0.078	0.385
Below median Oral Vocab at baseline	0.011	0.080	0.890	0.913
Above median Oral Vocab at baseline	0.179	0.077	0.023	0.385
Caregiver not schooled	0.011	0.090	0.901	0.913
Caregiver attended school	0.116	0.073	0.119	0.385
Household non-poor	0.078	0.069	0.262	0.495
Household very poor	0.079	0.085	0.353	0.545
Katete	0.231	0.176	0.222	0.472
Sinda	0.129	0.074	0.099	0.385
Petauke	0.011	0.096	0.913	0.913

Table B2. Adjusted *q* Values for Heterogeneous Effects—Standardized EGRA Score

Subgroup	Coefficient	Std. Error	Unadjusted *p* value	Adjusted *q* value
Boys	0.277	0.078	0.001	0.003
Girls	0.277	0.115	0.018	0.035
Below median ZAT at baseline	0.185	0.080	0.025	0.039
Above median ZAT at baseline	0.347	0.111	0.003	0.008
Below median EGRA at baseline	0.118	0.080	0.145	0.165
Above median EGRA at baseline	0.444	0.115	0.000	0.002
Below median EGMA at baseline	0.081	0.080	0.319	0.319
Above median EGMA at baseline	0.471	0.105	0.000	0.001
Below median Oral Vocab at baseline	0.187	0.092	0.047	0.066
Above median Oral Vocab at baseline	0.344	0.095	0.001	0.002
Caregiver not schooled	0.222	0.097	0.025	0.039
Caregiver attended school	0.253	0.094	0.009	0.022
Household non-poor	0.363	0.093	0.000	0.002
Household very poor	0.145	0.081	0.076	0.100
Katete	0.284	0.213	0.214	0.228
Sinda	0.222	0.131	0.108	0.131
Petauke	0.279	0.108	0.014	0.030

Table B3. Adjusted *q* Values for Heterogeneous Effects—Standardized EGMA Score

Subgroup	Coefficient	Std. error	Unadjusted *p* value	Adjusted *q* value
Boys	0.200	0.075	0.009	0.138
Girls	−0.017	0.081	0.829	0.940
Below median ZAT at baseline	0.053	0.072	0.465	0.639
Above median ZAT at baseline	0.122	0.097	0.214	0.639
Below median EGRA at baseline	0.054	0.078	0.488	0.639

Above median EGRA at baseline	0.105	0.085	0.221	0.639
Below median EGMA at baseline	0.000	0.079	0.996	0.996
Above median EGMA at baseline	0.185	0.075	0.016	0.138
Below median Oral Vocab at baseline	0.029	0.065	0.663	0.805
Above median Oral Vocab at baseline	0.174	0.085	0.045	0.256
Caregiver not schooled	0.012	0.091	0.893	0.949
Caregiver attended school	0.088	0.075	0.246	0.639
Household non-poor	0.085	0.079	0.288	0.639
Household very poor	0.070	0.084	0.408	0.639
Katete	0.172	0.179	0.361	0.639
Sinda	0.064	0.085	0.458	0.639
Petauke	0.066	0.093	0.482	0.639

Table B4. Adjusted *q* Values for Heterogeneous Effects—Standardized Oral Vocab. Score

Subgroup	Coefficient	Std. Error	Unadjusted *p* value	Adjusted *q* value
Boys	0.146	0.077	0.064	0.909
Girls	−0.048	0.070	0.496	0.909
Below median ZAT at baseline	0.061	0.081	0.454	0.909
Above median ZAT at baseline	0.021	0.083	0.796	0.909
Below median EGRA at baseline	0.068	0.077	0.377	0.909
Above median EGRA at baseline	0.028	0.081	0.728	0.909
Below median EGMA at baseline	0.007	0.072	0.923	0.923
Above median EGMA at baseline	0.101	0.074	0.176	0.909
Below median Oral Vocab at baseline	0.093	0.078	0.239	0.909
Above median Oral Vocab at baseline	0.010	0.070	0.881	0.923
Caregiver not schooled	−0.022	0.089	0.802	0.909

Caregiver attended school	0.077	0.071	0.282	0.909
Household non-poor	0.059	0.068	0.395	0.909
Household very poor	0.030	0.089	0.739	0.909
Katete	0.048	0.183	0.801	0.909
Sinda	0.058	0.096	0.555	0.909
Petauke	0.034	0.080	0.673	0.909

About the American Institutes for Research®

Established in 1946, the American Institutes for Research® (AIR®) is a nonpartisan, not-for-profit institution that conducts behavioral and social science research and delivers technical assistance both domestically and internationally in the areas of education, health, and the workforce. AIR's work is driven by its mission to generate and use rigorous evidence that contributes to a better, more equitable world. With headquarters in Arlington, Virginia, AIR has offices across the U.S. and abroad. For more information, visit AIR.ORG.

AIR® Headquarters
1400 Crystal Drive, 10th Floor
Arlington, VA 22202-3289
+1.202.403.5000 | **AIR.ORG**

28848_09/25