Mixed-Methods Cluster-Randomized Controlled Trial of Impact Network's eSchool 360 Model in Rural Zambia

Endline Report

Thomas de Hoop, Hannah Ring, Garima Siwach, Paula Dias, Gelson Tembo, Adam Taube, Torben Behmer

JANUARY 2025

American Institutes for Research® | AIR.ORG

AIR®
Advancing Evidence.
Improving Lives.

Contents

Tables

Figures

Acknowledgements

The authors would like to thank the AIR Opportunity Fund for funding this study. We would especially like to thank Reshma Patel at Impact Network, Michael J. Weiss at MDRC, and Jordan Rickles at the American Institutes for Research for their valuable comments and feedback on this paper. 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.

Executive Summary

Background

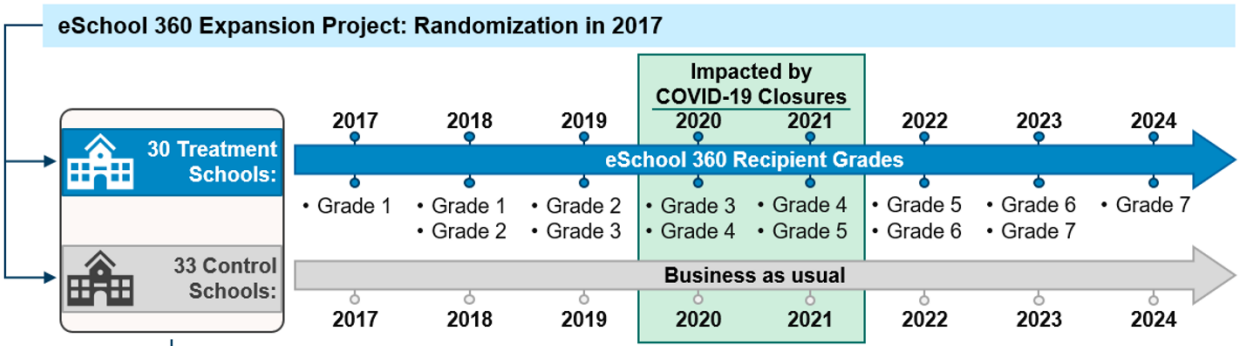
Since 2017, Impact Network has invested in improving the quality of primary education in 30 community primary schools (and 5 government schools) in Zambia's Eastern Province using its eSchool 360 model and with support from American Institutes for Research (AIR). Zambia has a large, autonomous community schooling system that expanded during the past 25 years to increase education access in remote areas. Impact Network's eSchool 360 model is a technology-aided instruction program that incorporates four high-impact interventions that offer important complementarities following the upgrading of community school infrastructure: (a) standardized e-learning curriculum and pedagogy, (b) ongoing teacher training and professional development, (c) community ownership through setting up parent–teacher associations, and (d) free primary education.

Methods

To determine program impacts, AIR designed and conducted a mixed methods cluster-randomized controlled trial (c-RCT) of the eSchool 360 model. We randomly assigned the program across 30 treatment and 33 control schools (business as usual community schools). Among treatment schools, the intervention did not cover all grades (Impact Network typically operates in all grades in the absence of an evaluation); rather, it was intended to be delivered to two cohorts of students: students in Grade 1 in 2017 and 2018, then to the same students when they progressed to Grade 2 the next year, Grade 3 the subsequent year, and all the way through primary school up to Grade 7. The intent of this design was to deliver the targeted eSchool 360 program to the same student cohorts throughout their entire time in primary school. Impact Network finished implementing the programming funded by the initial grant in 2024, during which time the eSchool 360 intervention targeted Grade 7 students—the final year of primary schooling. Exhibit E1 summarizes the project rollout and timeline.[1]

[1] Impact Network received a follow-up grant from the AIR Opportunity Fund for additional programming as well.

Exhibit E1. Summary of Project Rollout & Timeline



We compared all children eligible to enroll in Grade 1 who lived near the Impact Network schools during the baseline survey with children eligible to enroll in Grade 1 who lived near the control schools to estimate the effect of the opportunity to enroll in an Impact Network school (i.e. the intention-to-treat (ITT) effect). The primary outcomes included scores on the early grade reading assessment (EGRA), early grade mathematics assessment (EGMA), the Zambian achievement test (ZAT), and oral vocabulary assessments. In addition, we estimated impacts on school attendance and enrollment, parental perceptions of school and education quality, aspirations for children's education, and views on early marriage, child cognitive development, and household food security as secondary outcomes. We complemented the impact evaluation with a process evaluation using qualitative methods to assess the fidelity of implementation of the program.

To estimate ITT effects, we randomly sampled 30 households from a census-generated sample frame for each of the 30 treatment and 33 control schools. We then used an ANCOVA model to estimate ITT effects 14 months after the start of the program during the midline survey and 4.5 years after the start of the program during the endline survey. Exhibit E2 presents a consort flow chart that describes the sampling strategy, starting with the eligibility criteria, followed by the random assignment, the census to determine the eligible population, and the survey and assessment data we collected during baseline, midline (14 months after the start of the program implementation) and endline (4.5 years after the start of the program implementation).

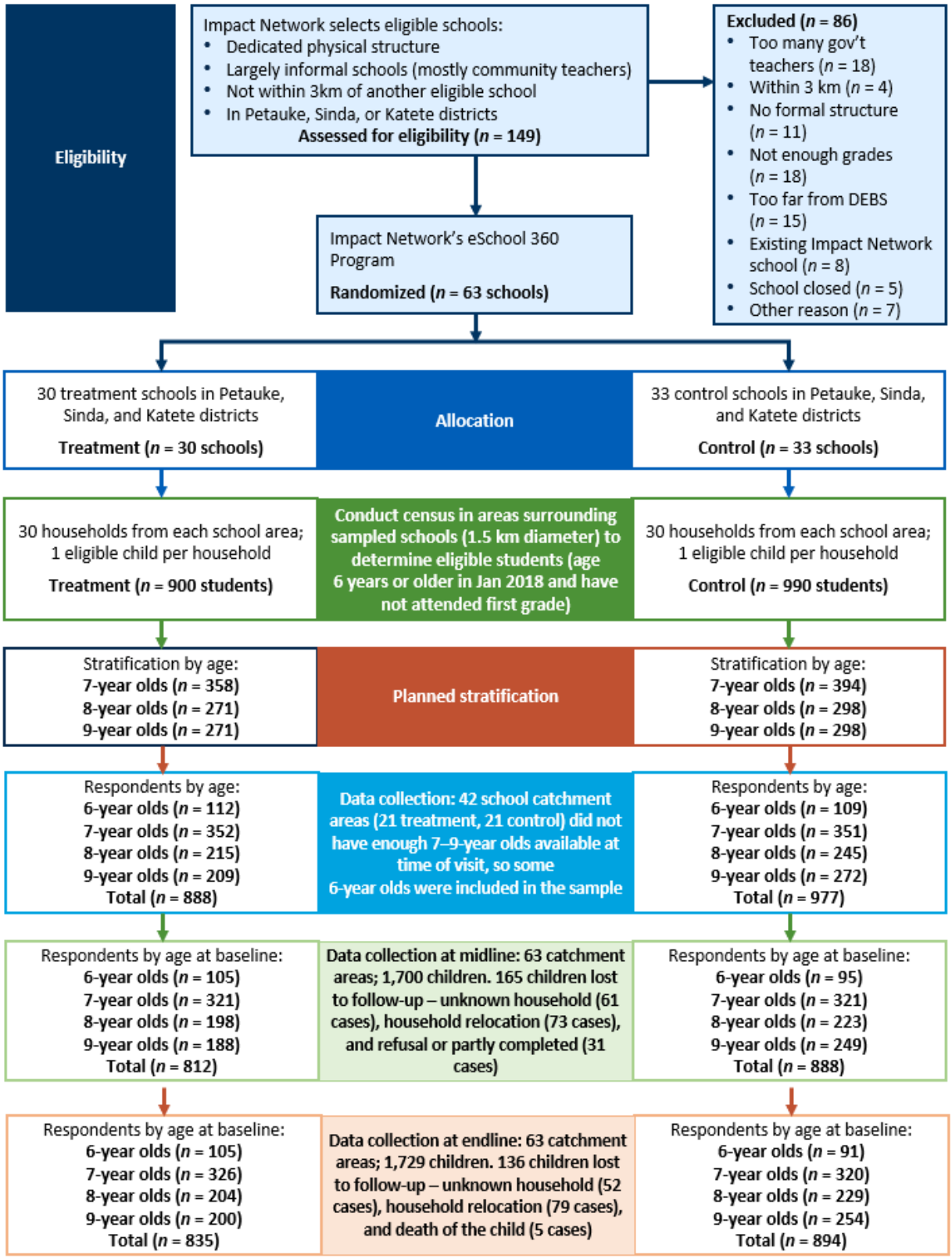
We also estimated two types of treatment effects on the treated (TOT) (effects for children who enrolled and attended in Impact Network schools) by using the program assignment as an instrumental variable. The first TOT examined effects for children who enrolled in Impact Network schools. The second TOT examined effects for children who self-reported attending Impact Network schools at least three times in the week before the survey.

Midline Results

The results of the study suggest substantial enrollment and learning gains for children 14 months after program start. Access to Impact Network schools increased primary school enrollment by 8 percentage points, EGRA scores by 0.40 standard deviations, EGMA scores by 0.21 standard deviations, ZAT scores by 0.15 standard deviations, and oral vocabulary scores by 0.25 standard deviations for all children eligible to enroll in first grade (de Hoop et al., 2023). These effects were larger for students who enrolled in and regularly attended Impact Network schools with effect sizes of up to 0.83 standard deviations on EGRA scores for students who attended Impact Network schools at least three times in the week before the survey (de Hoop et al., 2023). Despite the positive effects, children residing in Impact Network catchment areas scored an average of only 10.9% correct on EGRA assessments and 24.3% correct on EGMA assessments. The children did score an average of 56.9% on ZAT, however.

Process evaluation findings indicated that the high fidelity of implementation was instrumental in achieving these results, which is consistent with recent evidence from impact evaluations showing that technology-aided instruction is most effective when it aims to improve instruction as in structured pedagogy programs (Muralidharan et al., 2019; Rodriguez-Segura, 2022).

Exhibit E2. Consort Flow Chart



Endline Results

This report shows follow-up impact evaluation results more than four and a half years after the start of the program and slightly more than one year after the restart of primary schooling following the COVID-19 pandemic in Zambia. In this way, the study examines whether learning gains driven by improvements in the quality of education were sustainable after long school closures and a post-COVID-19 return to school in rural Zambia. We also present findings of a second round of qualitative process evaluation data to examine the fidelity of program implementation as well as a cost-effectiveness analysis.

The results continue to show substantive and statistically significant impacts on EGRA scores, but on average we no longer find statistically significant effects on EGMA, ZAT, or oral vocabulary scores. We find positive and statistically significant impacts of 0.26 standard deviations on EGRA. The point estimates for the other assessments are all positive, but not statistically significant. The relatively high quality of education in Impact Network schools likely contributed to the positive impacts on literacy as shown by substantial and positive impacts on parents' perceptions of the quality of education. It is, however, also possible that students in treatment schools had higher EGRA scores because they received more schooling over the course of the evaluation as they started primary school earlier, on average.

While we did not find positive average effects on EGMA, ZAT, and oral vocabulary scores, the study did show positive and statistically significant impacts on EGMA scores for boys and children who had higher than median EGMA scores during the baseline. We also find larger effects on EGRA scores for these sub-groups. Impacts on literacy assessments were also larger for students who scored higher than the median on literacy assessments during the baseline, but effects remained statistically significant for students who scored lower than the median EGRA score during the baseline. The positive effects on EGMA, ZAT, and oral vocabulary scores were not statistically significant for any of the subgroups after applying corrections for multiple comparisons.

While all point estimates were positive, we did not find statistically significant average effects on EGMA, ZAT, and oral vocabulary scores, likely due to a combination of ceiling effects, COVID-19 school closures, and dosage effects (only 49% of the treatment group received at least 1 year of the eSchool 360 program; 20% received more than 4 years of the program). Children in the treatment and control group both made significant gains in ZAT and oral vocabulary assessment suggesting that ceiling effects likely limited the potential to achieve positive effects on these assessments. The COVID-19 school closures also likely limited positive program effects on EGMA outcomes, similar to broader findings in the literature on the decline in math performance in OECD countries after COVID-19 (OECD, 2023).

We also did not find longer-term impacts on school enrollment and attendance, possibly because of the school closures during the COVID-19 pandemic.[2] The midline results showed positive program effects on school enrollment and attendance, but we found larger increases in school enrollment in the control group in between the midline and the endline survey. While treatment students were 8 percentage-points more likely to be enrolled in school at midline, the control group had caught up by endline. We did not find any statistically significant differences in enrollment rates by treatment status during the endline survey.

Despite larger gains in school enrollment between the midline and the endline survey, the control group was less likely to be enrolled in grades 5 or 6 compared to the treatment group, as the treatment group had already attained higher enrollment rates by midline. Children in the treatment group were 9 percentage points more likely to enroll in grade 5 or higher compared to children in the control group. The longer duration in school may have contributed to positive impacts on learning outcomes. This difference in higher grade enrollment was observed among both boys and girls, despite a lower likelihood for boys to be enrolled in schools, overall. At endline, 21% of boys in the treatment group and 10% of boys in the control group were attending grades 5 or 6. In both the treatment and the control group, girls were almost 13 percentage points more likely to attend grades 5 or 6.

Findings from a comparison with government schools revealed that Impact Network students continued to score higher on all assessments than students in government schools in Grade 5. Overall, Impact Network students who were enrolled in Grade 5 scored 8 percentage points higher than students in government schools in EGRA and 5 percentage points higher in EGMA scores. These findings indicate that the eSchool 360 program may have the potential to generate significant learning gains when the program would get scaled up with high fidelity of implementation as part of the Zambian national education system.

Process evaluation findings indicate that the eSchool 360 program was implemented with a high fidelity of implementation. School surveys indicated that teachers in Impact Network schools were more likely than teachers in control schools to have received professional development trainings in the past year. Qualitative data indicated that training was carried out with the intended frequency, and respondents perceived the trainings as engaging and responsive to the needs of staff members and teachers. However, some respondents noted

[2] While social distancing rules for the treatment and the control schools were the same during the COVID-19 pandemic, it is possible that the implementation of the social distancing rules was different in treatment and control schools. We do not find evidence for such differences in school-level survey data, but it remains difficult to completely rule out such differences because of potential recall bias.

challenges related to transportation and teachers' ability to absorb training materials. Weekly classroom observations and coaching sessions were largely implemented as planned, despite transportation challenges of Teacher Supervisors. Respondents indicated that coaching sessions were particularly effective in building the capacity of teachers, though some felt that the frequency of coaching was excessive.

Qualitative data revealed that the tablets and projectors distributed by the Impact Network program were used in classroom instruction to near unanimously positive reception. However, despite regular maintenance checks, technology usage in the classrooms occasionally faced challenges because of malfunctions of projectors and solar panels. Common problems included projects with burnt bulbs, tablets that were insufficiently charged, and a long wait for repair parts for solar panels. Further, some students and parents mentioned lack of seating as a challenge in the classrooms.

Teachers, students, and parents all reported positive perceptions of how students were treated in Impact Network program schools. The positive perceptions of Impact Network schools' classroom management practices extended to gender equity, with students reporting that boys and girls were treated equally. Further, parents reported that Impact Network takes child protection issues seriously, and children consistently reported knowing how to approach their teachers if they face a problem at home or at school. While teachers received training on child protection, both Impact Network staff and teachers themselves said they need more training and support on how to handle more complex child protection challenges.

Teacher pay and teacher working conditions were important concerns for respondents. While teachers are paid regularly and on time, the amount was considered nominal. Teachers, Teacher Supervisors, and Impact Network staff all agreed that low salaries were a demotivating factor for teachers.

Perceptions of parents and Impact Network staff indicated that the eSchool 360 program was more effective in improving literacy than government schools. In line with the quantitative findings showing larger effects on literacy outcomes, the process evaluation did not uncover as much information about the contributions of the eSchool 360 program to mathematics as to literacy outcomes.

While the process evaluation indicates that the program was implemented with high fidelity, the relatively small percentage of students receiving the full dosage of the program likely led to reduced program impacts. Because of the design of this study in which the program followed the same two cohorts of students rather than schoolwide implementation, 49% of the treatment group received at least one year of the eSchool 360 program and 20% received the

program for more than four years. Our findings suggest that those students who received more than four years of the program did achieve positive and statistically significant impacts on reading, while impacts on mathematics continue to show positive point estimates that are statistically insignificant for the full sample (the effects are statistically significant for boys and children who scored higher on EGMA during the baseline). Some caution is required in interpreting these findings because of methodological reasons, however. The estimate of correlation between the treatment assignment and receiving four and a half years of the program was statistically weak, which creates challenges in interpreting the impact estimates for the students who received the full dosage of the program (among others because of limited statistical power).

Costs and Cost-Effectiveness of the program

After four years of implementation, estimates of the cumulative average costs of the eSchool 360 model range from $55.00 to $113.00 per student, while the average costs in control schools was $5.90 per student, depending on varying assumptions. The cost estimates for the eSchool 360 model include all expenditures made by the school, as well as expenditures made by Impact Network in planning, delivering, and monitoring the implementation of the eSchool 360 program under this project. Cost estimates for control schools include all school-level expenditures only. The costs and cost-effectiveness estimates vary depending on whether they include the costs of the teacher salaries (which are also included in government schools), and on whether they include Impact Network's operational costs in Zambia. While school expenditures are lower in Impact Network schools than other community and government schools, the costs of the eSchool 360 program increase considerably when including the costs of the Impact Network' operations in Zambia.

The cost-effectiveness analysis at endline suggests that the program generates between 0.35 and 1.40 standard deviations in EGRA outcomes and 0.11 to 0.25 standard deviations in EGMA outcomes for every additional $100 spent.

Scaling back costs to midline where Treatment schools were 14 months into program implementation, we find substantially larger cost-effectiveness of the program, owing to both larger impacts and lower costs at midline. Estimates suggest that every $100 spent on the program resulted in an additional 1.38 to 3.30 SD gain in EGRA and 0.73 to 1.62 SD in EGMA.

The costs and cost-effectiveness estimates are challenging to interpret and compare with other education programs in Africa because of differences in context and methodology. Existing cost-effectiveness analyses of education programs in international development tend to use inconsistent methods for estimating the costs and cost-effectiveness of education programs (Evans & Popova, 2016). For example, some studies include the costs of operating a non-

governmental organization, while other studies only include the costs of the implementation of the educational intervention. Our cost estimation is more conservative than most estimates in the literature, because we consider costs borne by schools as well as costs realized by the program implementer(s).

Future Research

While the eSchool 360 program showed positive effects on literacy outcomes, the findings suggest that the longer-term program effects and the cost-effectiveness of the program, while still positive, were less positive than at midline. School closures due to COVID-19 and the partial dosage of the intervention for the majority of the students likely limited the longer-term program effects. In addition, it remains uncertain how much children will benefit from the program if they do not enroll in secondary school or if the quality of secondary education is relatively low.

To generate additional impacts on literacy outcomes, Impact Network has introduced an intervention for early grade literacy based on the use of phonetic charts (the "Read Smart" program) in Impact Network schools. Preliminary findings suggest that this intervention may generate additional positive impacts on literacy outcomes. However, the program has not yet been implemented in government schools and the research on the effectiveness of the intervention is only preliminary, without a rigorous impact evaluation. This suggests that additional research is required to assess the impact and cost-effectiveness of the "Read Smart" program in community and in government schools.

Impact Network also started a partnership with Promoting Equality in African Schools, which aims to improve the quality of secondary education in Zambia. While the eSchool 360 model has positive impacts on literacy, it is uncertain how much students will benefit from these learning gains in the absence of high-quality secondary education. Because the quality of secondary education in Zambia is relatively low, it is critical to conduct additional research on the impact of innovations to improve the quality of secondary education.

It is also important to design and implement additional research to examine how to scale up the program in a cost-effective manner as part of the Zambian national education system. While the eSchool 360 program had substantial impacts on literacy and numeracy during the midline survey, the impacts during the endline were smaller. In addition, the cost-effectiveness analysis suggests that Impact Network operations carry substantial costs that are not sustainable when aiming to expand the program to a larger number of schools. The educational costs of the eSchool 360 program are smaller than the educational costs of community schools. However, the program has substantial indirect costs, which may increase further when expanding the program to a larger number of schools. This finding suggests that more research

is needed to determine whether the eSchool 360 program can maintain the same or larger impacts, while reducing indirect costs.

Finally, it is important to examine why the Impact Network program had larger and statistically significant effects on EGMA outcomes for boys than for girls. Boys are more likely to drop out of school than girls because of cattle herding, but the eSchool 360 program seems to have larger positive effects on learning outcomes for boys. It is possible that gender norms influence the heterogeneous effects. Future research can examine the role of gender norms and other potential explanations for why boys seem to generate larger benefits from the eSchool 360 model than girls.

Recommendations

We provide various recommendations to the AIR Opportunity Fund, international donors, the Zambian Government and Impact Network based on the midline and endline findings. These recommendations include suggestions on how to scale the program, and specific implementation features. When providing recommendations based on the endline findings, it is important to consider the context of the COVID-19 pandemic and the associated school closures. The program impacts four and a half years after the start of the program implementation were likely smaller because of increased drop-out after COVID-19 and the associated school closures. For this reason, we also consider the possibility that the midline impact estimates are more externally valid for the Zambian context after the COVID-19 pandemic than the endline findings. At the same time, it remains important to consider the endline findings in the recommendations. While the endline results may not be fully externally valid, they provide relevant information about the potential longer-term impacts of the eSchool 360 program. The midline results only provide information about the program impacts 14 months after the start of the program implementation.

Recommendations to scale up the eSchool 360 program

We recommend the AIR Opportunity Fund to invest in Impact Network to expand the program to grade 1 and grade 2 of the control schools: The midline and endline findings indicate that upgrading the infrastructure and implementing the eSchool 360 model in treatment and control schools will bring substantial educational benefits to those schools and enable Impact Network to continue expanding its primary school program in a cost-effective manner. The endline findings show that the eSchool 360 model improves literacy outcomes and mathematics outcomes for boys and children with higher EGMA scores during the baseline. Most of these impacts were generated during the first two years of the program implementation during which the program showed larger impacts on literacy and mathematics outcomes than during the

endline survey. These findings suggest that expanding the program to Grade 1 and 2 of the control schools will generate substantial benefits.

While it is possible that the program would have generated larger impacts on students in grade 5 and 6 in the absence of the COVID-19 pandemic, current evidence is insufficient to recommend expanding the program to all grades in control schools. The program effects and cost-effectiveness reduced considerably between the midline and the endline survey. While the COVID-19 pandemic and associated disruptions in schooling may have contributed to reducing the impact of the program, it is uncertain whether expanding the program to all grades in control schools would result in sufficient benefits to justify a large investment, considering that the costs would increase considerably when expanding the program to all grades.

We recommend the AIR Opportunity Fund to examine opportunities for integrating literacy elements of the eSchool 360 model as part of Zambian government schools: The midline and endline results show substantial positive impacts of the eSchool 360 model on literacy outcomes, especially in grades 1 and 2. In addition, Impact Network has invested in a new curriculum based on phonetic charts (the "Read Smart" program) that could generate additional impacts on literacy outcomes. However, at this moment only community schools implement the Read Smart program. Expanding the eSchool 360 program will require integrating its literacy components into government schools, and additional research on the impact and cost-effectiveness of the "Read Smart" program in community and government schools.

We recommend the AIR Opportunity Fund, the Zambian government and international donors to look for additional opportunities to improve the quality of secondary education for community school students who enroll in secondary schools in the Impact Network catchment areas: While the eSchool 360 model has positive impacts on learning outcomes in primary school, the long-term benefits remain uncertain if the quality of secondary education remains relatively low. Low-quality secondary education may result in the fade-out of improvements in learning outcomes in primary schools after students enroll in secondary school. Strengthening and empowering school leaders to create a secondary school environment that enables high-quality learning could also incentivize secondary school participation and bring sustainable learning benefits to students who enroll in secondary school. This suggests that additional investments are needed to pilot and assess the effectiveness of innovations to improve the quality of secondary education in rural Zambia.

Recommendations to improve implementation of the eSchool 360 program

Based on the process evaluation findings, we suggest that Impact Network consider the following recommendations to improve its programming:

- **Device maintenance and charging** – Despite regular maintenance checks, Impact Network schools faced challenges related to malfunction of projectors and solar panels. To minimize these challenges and improve the reliability of technological tools at Impact Network schools, Impact Network should consider having a stockpile of commonly replaced parts for projectors and solar panels, and more training for teachers and school staff on proper use of these devices.
- **Analog teaching materials** - Given the potential for failure and/or malfunction of technological devices as well as the demand for at-home learning, Impact Network should consider if more textbooks could be provided to supplement the existing printed materials provided to teachers (e.g.: backup lesson plans) and ensure that teachers can continue to deliver content in the case of technological issues.
- **Seating** - When budgeting, Impact Network should consider allocating funds to provide additional classroom seating (benches or desks) to ensure all children have their own seat. Many respondents noted inadequate classroom seating, with some saying that the lack of seats is especially problematic for the type of group work incorporated in both Catch-Up classes and the e-school 360 model.
- **Coaching** – Although coaching was generally considered effective in supporting teachers, some teachers noted that the frequency of coaching/observation sessions was sometimes excessive. In addition, Teacher Supervisors reported difficulties in the logistics of traveling to schools. Impact Network should consider developing a progressive observation schedule where the coaching frequency can decrease to biweekly sessions for teachers with a certain length of service who are performing well. Also, Impact Network should consider making the coaching rubric available to teachers so they can track progress and areas of improvement.
- **Additional support to the Lusangazi district** – Given the challenges with student attendance and parental engagement in parent-teacher associations in Lusangazi, we recommend Impact Network provides additional support to the Lusangazi parent-teacher association to encourage parental participation and highlight the importance of regular student attendance. Lusangazi district was made independent from Petauke district in 2018, and therefore likely does not yet have all of the capacity of an independent district.
- **Support for more complex child protection challenges** - Interviews with Impact Network teachers showed that they need more support to safely resolve more complex child protection challenges. We suggest Impact Network provides additional training and written guidance on how to resolve more complex child protection issues and consider additional ways to ensure that teachers feel both empowered and safe as the work to resolve these issues.

- **Teacher salaries** - We suggest that Impact Network regularly reassesses and potentially recalibrates teacher salaries based on required credentials and the government pay scale. At the time of data collection, teachers in treatment schools expressed considerable frustration with their compensation and reported it was much less than what government teachers were paid.[3]

The eSchool 360 model has considerable potential to generate positive impacts on learning outcomes at scale when the government would scale-up the program. Incorporating the recommendations to scale up and improve implementation of the program may generate even larger impacts at a lower cost.

[3] It is important to recognize, however, that community school teachers in control schools are volunteers.

1. Introduction

An increasing body of evidence indicates that COVID-19-induced school closures led to significant learning losses globally (Goldhaber et al., 2023; Schuurman et al., 2023) that tended to increase inequality globally and within the United States (Jack & Oster, 2023). Some of the largest studies from the United States indicate that school closures and remote and hybrid instruction led to large learning losses (Goldhaber et al., 2023). Similarly, a study from the Netherlands shows large learning losses caused by school closures (Haellermans et al., 2022). Data from OECD countries further suggest that the decline in education performance was steeper for math performance (OECD, 2023). As in the United States and Europe, schools in sub-Saharan Africa closed for a long period of time during the pandemic with children in some countries losing ten percent or more of the total time they were expected to spend in-person at school over the course of their lives (Evans et al., 2021) and two-thirds of an academic year of schooling in 2020 (UNESCO, 2021). However, there is only limited rigorous evidence on the effects of the pandemic on learning outcomes in sub-Saharan Africa. This raises questions about the magnitude of learning losses in sub-Saharan Africa, the sustainability of those learning losses, and how existing and new programs may have mitigated some of the negative effects of the pandemic on learning losses.

Simulations from early in the pandemic suggested large learning losses in sub-Saharan Africa, but empirical studies show mixed evidence for this hypothesis. For example, Angrist et al. (2020) predicted that students from Africa would lose approximately 2.8 years of schooling because of COVID-19 induced school closures across the continent. However, empirical evidence from Uganda suggests that learning losses were very limited, possibly because of the low quality of education before the pandemic (Sandefur, 2022). The latter finding is aligned with the idea of a learning crisis in sub-Saharan Africa. While student enrollment in sub-Saharan Africa increased tremendously, learning gains have not kept up with increases in school enrollment (Angrist et al., 2020; Nakamura et al., 2023; World Bank, 2018).

To mitigate learning losses during the pandemic, some implementers focused on using phone tutorials and radio education with mixed evidence for success. Program in Botswana and Nepal sent SMS messages with mathematics problems and combined those with follow-up calls to let NGO staff help the students with the problems. Rigorous studies of these programs found positive effects on mathematics outcomes ranging from 0.16 to 0.29 standard deviations (Angrist et al., 2022; Radhakrishnan et al., 2021). A third study examining a 30-minute phone mentoring session with a student volunteer from a local university found larger effects of 0.56 standard deviations on mathematics and 0.66 standard deviations on literacy outcomes (Hassan et al., 2021). However, two studies examining the effects of accountability checks or tutoring

calls in Kenya (Schueler & Rodriguez-Segura, 2021) and radio instruction in Sierra Leone (Crawfurd et al., 2023) did not find positive effects on learning outcomes.

In this report, we contribute to the literature by examining whether learning gains driven by improvements in the quality of education from before the pandemic were sustainable after long school closures and a post-COVID-19 return to school in rural Zambia. To examine this research question, we estimate four and a half year impacts of the eSchool 360 model, a multi-faceted technology-aided instruction program implemented by Impact Network in Zambia, based on a five-year cluster-randomized controlled trial. We combine the analyses on the impact of the program on educational outcomes with a longitudinal process evaluation and a cost-effectiveness analysis to guide decisions on whether and how to scale up the program as part of the national education system in Zambia.

As discussed in our impact evaluation 14 months after the start of the implementation before COVID-19, the program represents a promising approach to improving educational outcomes by incorporating three potentially high-impact components that could create important synergies: a standardized e-learning curriculum and pedagogy, ongoing teacher training and professional development, and community ownership (de Hoop et al., 2023). Each component could, on its own, have positive impacts on student outcomes by engaging the three main factors in the education system—students, teachers, and parents. Combining these components into a single program may be particularly effective by aligning all three factors toward improving the educational outcomes of students. Earlier research has suggested that these complementarities may be substantial, with even higher impacts from educational technology programs that include a strong focus on pedagogical practices (Muralidharan, Singh, & Ganimian, 2019).

The mixed-methods cluster-randomized controlled trial of the eSchool 360 program comprises three main evaluation components: an impact evaluation of the eSchool 360 model, an analysis of the cost-effectiveness of the eSchool 360 model, and a process evaluation of the expansion of the eSchool 360 model. To determine the impact of the program, we used a cluster-RCT in which 63 eligible schools were randomly assigned either to receive Impact Network's eSchool 360 program (30 treatment schools) or not to receive the program (33 control schools). The population consisted of children ages 6–9 who lived in close proximity to the 63 selected community schools across four districts—Petauke, Lusangazi, Sinda, and Katete—in Zambia's Eastern Province during the baseline survey. The primary cognitive skills outcomes were aggregate scores on the early grade reading assessment (EGRA), early grade mathematics assessment (EGMA), the Zambian Achievement Test (ZAT), and Oral Vocabulary assessments.

In this report, we present longer-term program effects four and a half years after the start of the program and slightly more than one year after the restart of schooling in Zambia after the pandemic using longitudinal survey data. We also present findings of a second round of qualitative data to examine the fidelity of program implementation as well as a cost-effectiveness analysis. Previously, we produced an inception report that details the design of the study (De Hoop et al., 2017), a baseline report that presents the quantitative data collection, analysis, and findings at baseline (de Hoop, Brudevold-Newman, & Davis, 2018), and a midline report and a peer-reviewed paper that discuss the quantitative data collection, analysis, and findings 14 months after the start of the program implementation (de Hoop et al., 2022; de Hoop et al., 2023).

The rest of this report is structured as follows. We describe the research questions (Section 2) followed by a description of the background and the eSchool 360 model (Section 3). Then, we present the theory of change of the program (Section 4) followed by the study design and the methodology (Section 5). We present the results in Section 6 and conclude the report with recommendations for scale-up and for improving implementation (Sections 7 and 8).

2. Research Questions

Next, we present research questions related to the impact evaluation, cost-effectiveness analysis, and process evaluation components of this study. Each part of the evaluation is designed to answer different but complementary questions.

2.1. Impact Evaluation

a. What is the effect of the eSchool 360 program on students' numeracy, preliteracy, and literacy skills?
b. Do students enrolled in the eSchool 360 program improve in numeracy and literacy skills over the course of this study (four-and-a-half year)?
c. Does the eSchool 360 program increase attendance and enrollment?
d. Does the eSchool 360 program lead to an improved perception of school and education quality among students, teachers, and parents?
e. Does the eSchool 360 program improve parental and children's aspirations?

2.2. Cost-Effectiveness

a. How cost-effective is the eSchool 360 program in improving literacy outcomes?
b. How cost-effective is the eSchool 360 program in improving numeracy outcomes?

2.3. Process Evaluation

a. Was the eSchool 360 program implemented as designed? If not, why was it not implemented as designed, what were the challenges to implementing it as designed, and how was it implemented?
b. How did the eSchool 360 program implementation vary by geography, culture, and time of year?
c. Did perceptions of the quality of teachers differ among students, parents, Teacher Supervisors, and teachers? If yes, how?

3. Background and eSchool 360 Model

This section presents the background of Zambia's education system and existing evidence on the impact of technology-in-education and structured pedagogy programming in sub-Saharan Africa. The section heavily borrows from the discussion of the literature in the paper describing the results of the midline survey and assessment results (de Hoop et al., 2023).

As discussed in de Hoop et al. (2023), Zambia has a large, autonomous community schooling system that expanded considerably, especially over the last two decades, to increase education access in remote rural areas. Impact Network's eSchool 360 model focuses on improving the quality of education of community schools using a technology-aided instruction program that incorporates four potentially high-impact interventions that can create synergies following the upgrading of community school infrastructure: (a) standardized e-learning curriculum and pedagogy, (b) ongoing teacher training and professional development, (c) community ownership through recruitment of local teachers and setting up parent–teacher associations, and (d) free primary education.

The core of the eSchool 360 model is e-learning technology whereby tablets and projectors, provided by IN's partner, *Mwabu*, are loaded with lesson plans for teachers and interactive lessons for students, approved by the Zambian government, and delivered in the local language. IN supplies electricity via solar power, and provides teachers with training and professional development. Before the midline of this study (pre-2019), Impact Network recruited local, untrained teachers who received weekly training focused on using the technology and enhancing teachers' pedagogical skills (de Hoop et al., 2023). However, by 2022, Impact Network schools recruited more fully qualified teachers following new regulations from the Zambian government.

Integrating technology-aided instruction through a standardized e-learning curriculum and pedagogy could improve the quality of education, but merely providing technology is insufficient to improve learning outcomes (Valk, Rashid, and Elder. 2010; Snilstveit et al., 2016; Rodriguez-Segura, 2021). A systematic review by Rodriguez-Segura (2021) suggests that technology in education programming is most effective when it centers around self-led learning or aims to improve instruction, for example through structured pedagogy approaches that provide teachers with structured teaching guides and continuous feedback. In line with this hypothesis, Muralidharan, Singh, and Ganimian (2019) show that using technology to individualize content to students through the Mindspark program in India improved both mathematics and reading scores at a relatively low cost. Integrating technology into education is unlikely to lead to positive effects on learning outcomes when the technology is not integrated into the curriculum and when teachers do not believe in using technology, however (Tauson & Stannard, 2018; Ale, Loh, and Chib, 2017; Valiente, 2010; Barrera-Osorio & Linden, 2009; Rodriguez-Segura, 2021).

Systematic reviews and impact evaluations also suggest that multifaceted education programs using structured pedagogy approaches are effective in improving learning levels in rural areas in sub-Saharan Africa (Glewwe & Muralidharan, 2016; Snilstveit et al., 2016; Eble et al., 2021; Eble & Escueta, 2022). Multi-faceted programs with several distinct interventions (including para teachers delivering after-school supplementary classes, scripted lesson plans, and frequent monitoring focusing on improving teacher practice) have large effects on literacy and numeracy outcomes among primary school children in rural poor areas in the Gambia (Eble et al., 2021; Eble & Escueta, 2022). Integrating technology into multifaceted programs may generate even larger learning gains for students. For example, a program that combines highly detailed lesson guides to teachers using tablets with standardized systems for daily teacher monitoring and feedback, school construction, and financial management in Kenya shows large effects on test scores after two years (Gray-Lobe et al., 2022). The midline results of the impact evaluation of the eSchool 360 program show similar results demonstrating substantial enrollment and learning gains for children 14 months after program start (de Hoop et al., 2023).

4. Theory of Change

The theory of change (see Figure 1, which we largely adapted from de Hoop et al. (2023)) is consistent with a structured pedagogy approach, which provides lesson plans and training for teachers with technology serving to strengthen instruction. Structured pedagogy programs focus on the inclusion of structured teacher guides, classroom observations, and feedback to teachers, and can lead to meaningful impacts on learning outcomes, caused by improvements

in teacher knowledge and practices in the classroom (Evans & Acosta, 2021; Piper et al., 2016; Gray-Lobe et al., 2022). The technology component of the eSchool 360 program can increase the effectiveness of structured pedagogy approaches by supporting teachers to apply activity-based learning using a curriculum approved by the Zambian government. In this way, the theory of change is consistent with a systematic review indicating that more or better resources improve student achievement only if they result in changes in children's daily experiences at school (Ganimian & Murnane, 2016).

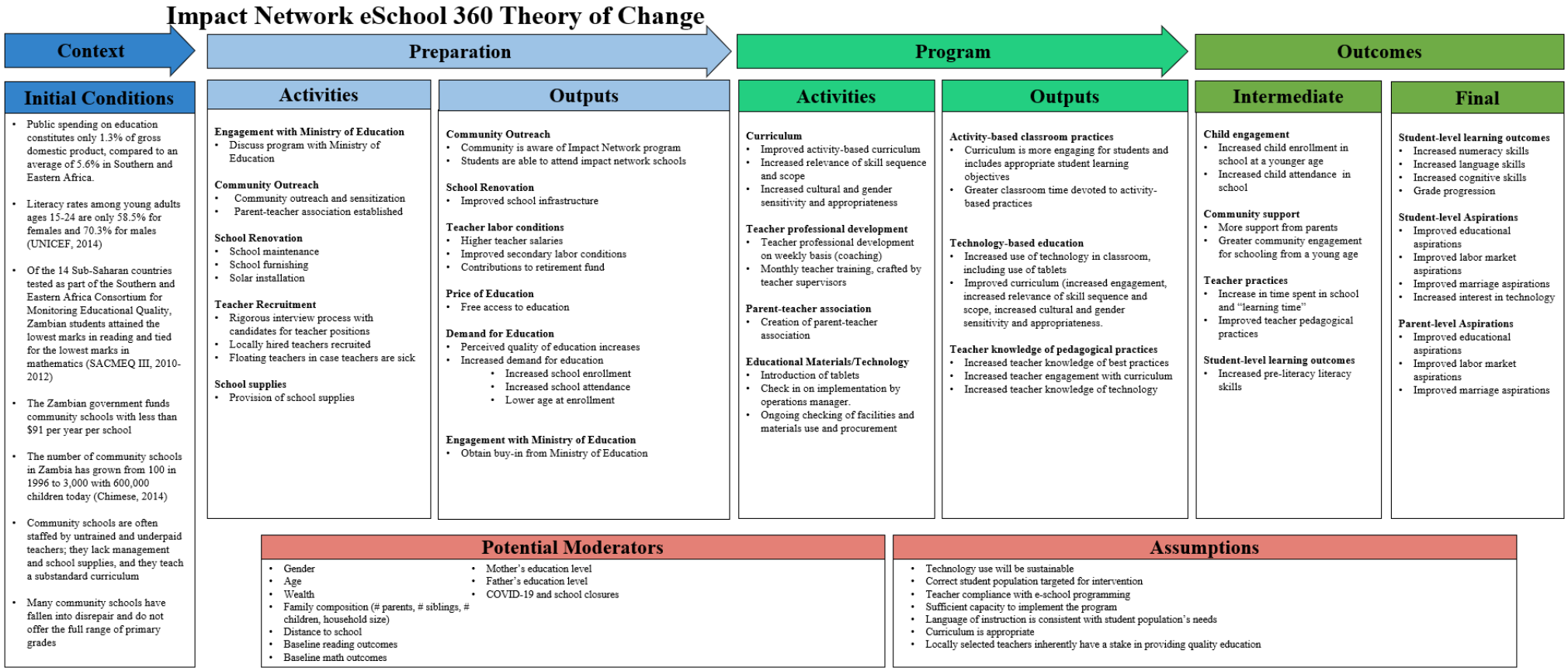
The program may lead to improvements in learning outcomes through various mechanisms improving the quality of the supply of education and increasing the demand for education. First, teacher professional development may lead to improvements in knowledge and practices of teachers, which may result in improvements in the quality of education—for example, through the integration of activity-based learning methods and improvements in the curriculum. These improvements, in turn, may lead to improvements in preliteracy, early grade reading, and early grade mathematics outcomes. Second, the incorporation of technology to aid instruction may increase the teaching effectiveness of teachers since pre-prepared lessons are sequenced and built into the tablets which are regularly checked and monitored by operations management. Third, the infrastructure improvements in the community school and the free primary education may lead to increased demand for education, which may result in increased education enrollment and attendance as well as early enrollment of Zambian students (lower age at enrollment in schools). These improvements in school attendance and enrollment may then result in an increase in the time spent on education, which may lead to improvements in learning outcomes. We present these pathways in Figure 1, which displays the theory of change we used to generate these hypotheses.

The program may also result in improvements in the aspirations of students and parents. Improvements in the quality of education may increase expectations for students' futures. These increased expectations may lead to higher aspirations in the domains of education, labor market, and family outcomes. For example, parents may increase their expectations of the likelihood that their children will study beyond senior grade level, which may result in increased expectations of the returns to education, and increases in expectations of the child's marriage prospects and an increase in the age at marriage. However, the midline results did not show evidence for positive impacts on the aspirations of students and parents.

Importantly, the validity of theory of change may have changed because of the COVID-19-induced school closures. For this reason, we include COVID-19 and the associated school closures as moderators in the theory of change.

The mixed-methods cluster-randomized controlled trial of the eSchool 360 program estimates the impact of the bundle of activities on educational outcomes. The evaluation is, however, not able to examine the impacts of individual program components. The latter would require an experimental or quasi-experimental evaluation with multiple treatment arms. This was not feasible to implement because of the limited number of treatment schools. Having multiple treatment arms with the current sample size would result in a significant reduction in statistical power.

Figure 1. Theory of Change



5. Study Design and Methods

5.1. Sampling

AIR and Impact Network designed the expansion of the eSchool 360 model to include a rigorous mixed-methods cluster-RCT to assess the impact of the program on learning outcomes. As discussed in de Hoop et al. (2023), we randomly assigned the program among schools that satisfied Impact Network's geographic, infrastructure, and organizational structure eligibility criteria for the eSchool 360 expansion. AIR and Impact Network jointly determined the geographic criteria based on Impact Network's goal of introducing community schools across Petauke, Sinda, and Katete in Zambia's Eastern Province.[4] Impact Network aimed to select schools with a dedicated physical structure, that were largely informal, and that had more community teachers than government teachers. We excluded pairs of eligible schools that operated within three kilometers of one another to reduce bias from spillovers or contamination. Impact Network first consulted with local Zambian government officials to obtain a list of all community schools in the region. Of the 149 community schools that were identified, 63 met all eligibility criteria.[5] Impact Network staff then visited each of the 63 schools to obtain information on the eligibility criteria.

To design a mixed-methods cluster-RCT, we closely consulted with Impact Network and Zambian government officials for the randomization. For the randomization, AIR and IN worked with Zambian government officials to obtain letters of approval for random assignment of the eSchool 360 program to 30 treatment schools and 33 control schools. In May 2017, Zambian Ministry of Education officials implemented the randomization, with AIR staff ensuring the integrity of the randomization process. Local representatives of the Ministry of Education implemented the randomization separately in each of the three districts (thus stratifying the randomization by district) with oversight from the research team in May 2017. We assigned five schools in Katete, nine in Sinda, and 16 in Petauke to the treatment group. The remaining 33 schools were assigned to the control group (the group not receiving the eSchool 360 model). Among treatment schools, the intervention did not cover all grades; rather, it was intended to

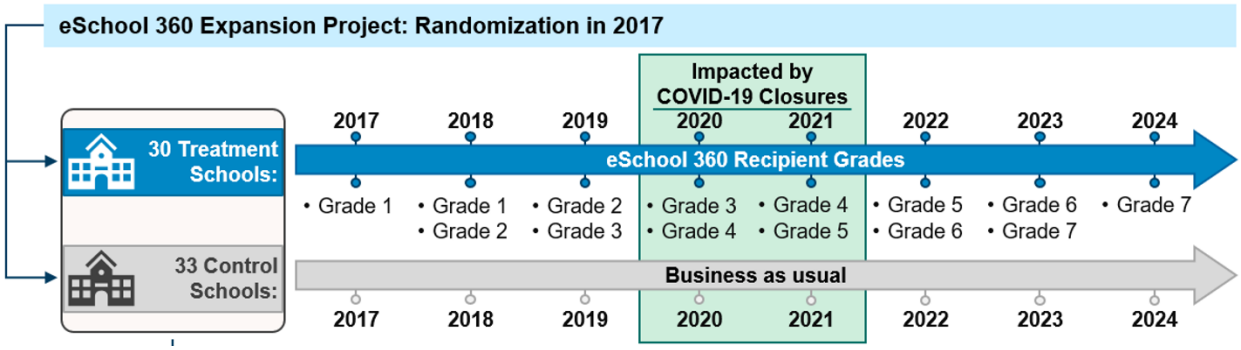
[4] Lusangazi district was made independent from Petauke district in 2018.

[5] Impact Network originally identified 64 eligible schools. However, further analysis suggested that the 64th school was too far away from the Impact Network offices to be considered for the program. This school had been assigned to the treatment group; we replaced that treatment school with a randomly selected school from the control group in the same district.

be delivered to two cohorts of students: students in Grade 1 in 2017 and 2018, then to the same students when they progressed to Grade 2 the next year and so on.

The intent of this design was to deliver the targeted eSchool 360 program to the same student cohorts throughout their entire time in primary school, while maintaining a counterfactual for the evaluation. Because Impact Network implementation of the eSchool 360 model influenced school enrollment and attendance, changes in enrollment could result in changes in the classroom composition. Therefore, the evaluation sampled children who were not yet enrolled in school during the baseline survey rather than sampling in-school students. This approach helped alleviate concerns about potential bias in the impact estimates. Impact Network finished implementing the programming funded by the initial grant in 2024, during which time the eSchool 360 intervention targeted Grade 7 students—the final year of primary schooling. Figure 2 summarizes the project rollout and timeline.

Figure 2. Summary of Project Rollout & Timeline



This study estimates longer-term program impacts for children who were eligible to enroll in first grade in 2017 and 2018, and who lived near one of the 63 schools when the eSchool 360 program started. We collected endline data in mid-2022, thus estimating four-and-a-half year impacts (starting from 2018). We focused on children eligible to enroll in first grade in 2017 and 2018 because the introduction of the eSchool 360 model was designed to expand to an additional grade each year of the model's operation. In the first year, only the first-grade cohort received the full program package. This expanded to Grades 1 and 2 in the second year, and so on. The study followed each of the sampled children, regardless of where and when they enroll in school, to estimate (1) Intent to Treat (ITT) effects of the model, which identifies the effect of the opportunity to enroll in an Impact Network school; and (2) Treatment effects on the Treated (TOT), which identify the effect of attending an Impact Network school on those who

enroll and those who actively participate in the program (i.e., regularly attend Impact Network schools).[6]

For the ITT analysis, we compared all eligible children who lived near the 30 Impact Network schools with eligible children who lived near the 33 control schools, thus including all children regardless of whether they enrolled in a school. To identify the eligible sample of children, we conducted a census in the areas surrounding the sample schools to identify all households with children eligible to enroll in first grade in January 2018; that is, children ages 6 years or older in January 2018 who did not attend first grade or higher in the prior school year.

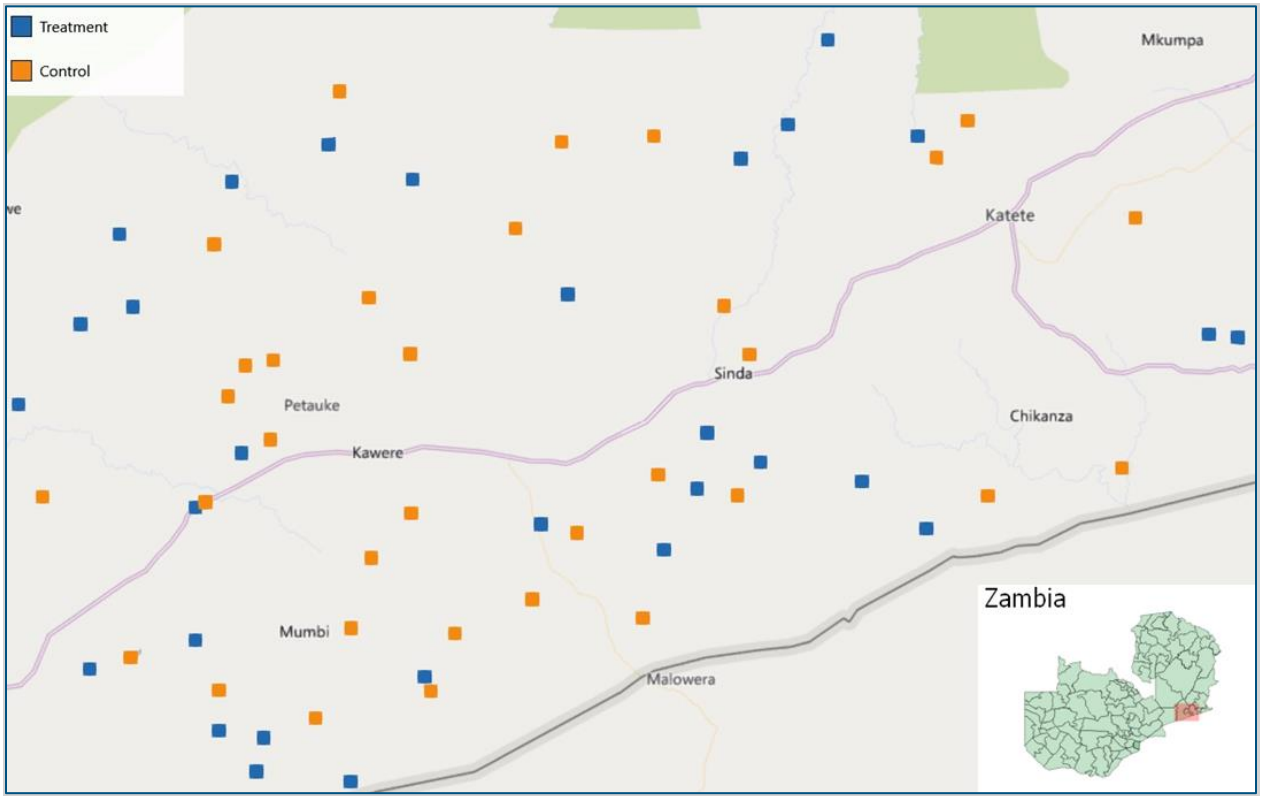
It is important to identify ITT effects because the announcement of the treatment schools and the construction of the additional infrastructure involved with the implementation of the eSchool 360 model are visible to the beneficiaries. The randomization of schools to treatment occurred nearly seven months before the evaluation cohort was first admitted to the sample schools in January 2018. This visibility may result in a different composition of students attending the treatment schools relative to the control schools. The introduction of Impact Network schools may have induced some children to enroll in school (who wouldn't have enrolled in school in the absence of the program). Such compositional changes could result in a bias in the impact estimator that compares students enrolled in the Impact Network schools with students enrolled in other community schools. Comparing the outcomes (e.g., test scores) of all eligible children allows for capturing any effects on inducing children to attend school.

To enable the estimation of ITT effects, we randomly sampled 30 households from the census-generated sample frame for each of the sample schools. For households with more than one eligible child, we selected the oldest child for inclusion in the sample. To increase statistical power, we oversampled 8-year-olds and 9-year-olds based on descriptive statistics about the age distribution from the Eastern Province (Ministry of General Education, 2016). These statistics indicated that 8-year-olds and 9-year-olds were more likely than 7-year-olds and 6-year-olds to be enrolled in first grade (Ministry of General Education, 2016). For our sampling strategy, we started by randomly selecting households with 7-year-olds, 8-year-olds, and 9-year-olds within a distance of 1.5 kilometers from each of the schools. However, we slightly adjusted the sampling strategy because 42 catchment areas (a catchment area is the area around a

[6] The Impact Network program only operates in the grade cohorts that started in the first two years of the program implementation. Because we use panel data, students who repeat their grade or only enroll in Impact Network schools after two years do no longer receive the eSchool 360 curriculum. They benefit from the infrastructure of the schools but less from the teaching methods. We examine this in more detail in the section describing the results in which we also examine treatment effects on the treated for those students who have always enrolled and attended Impact Network grades.

particular school) did not have enough 7- to 9-year olds. The sample was adjusted by randomly sampling additional 6-year olds, followed by sequentially expanding the radius of the circle by an additional 0.5 kilometers until a sufficient sample was identified. This procedure was implemented consistently across treatment and control school-catchment areas.[7] The final sample at baseline included 888 children ages six to nine from 30 treatment areas, and 977 children ages six to nine from 33 control areas. Figure 3 displays a map of the treatment and control areas.

Figure 3. Map of Sample Schools



5.2. Quantitative Data Collection

The Zambian data collection firm Palm Associates collected baseline (early 2017), midline (early 2019), and endline (mid 2022) data using enumerators who speak the local language of the areas included in the study and who are familiar with the assessments that we are

[7] The initial distance stemmed from conversations with local experts who suggested that children eligible for first grade generally do not walk more than 1.5 kilometers to school. However, more recent information suggests that children sometimes walk further.

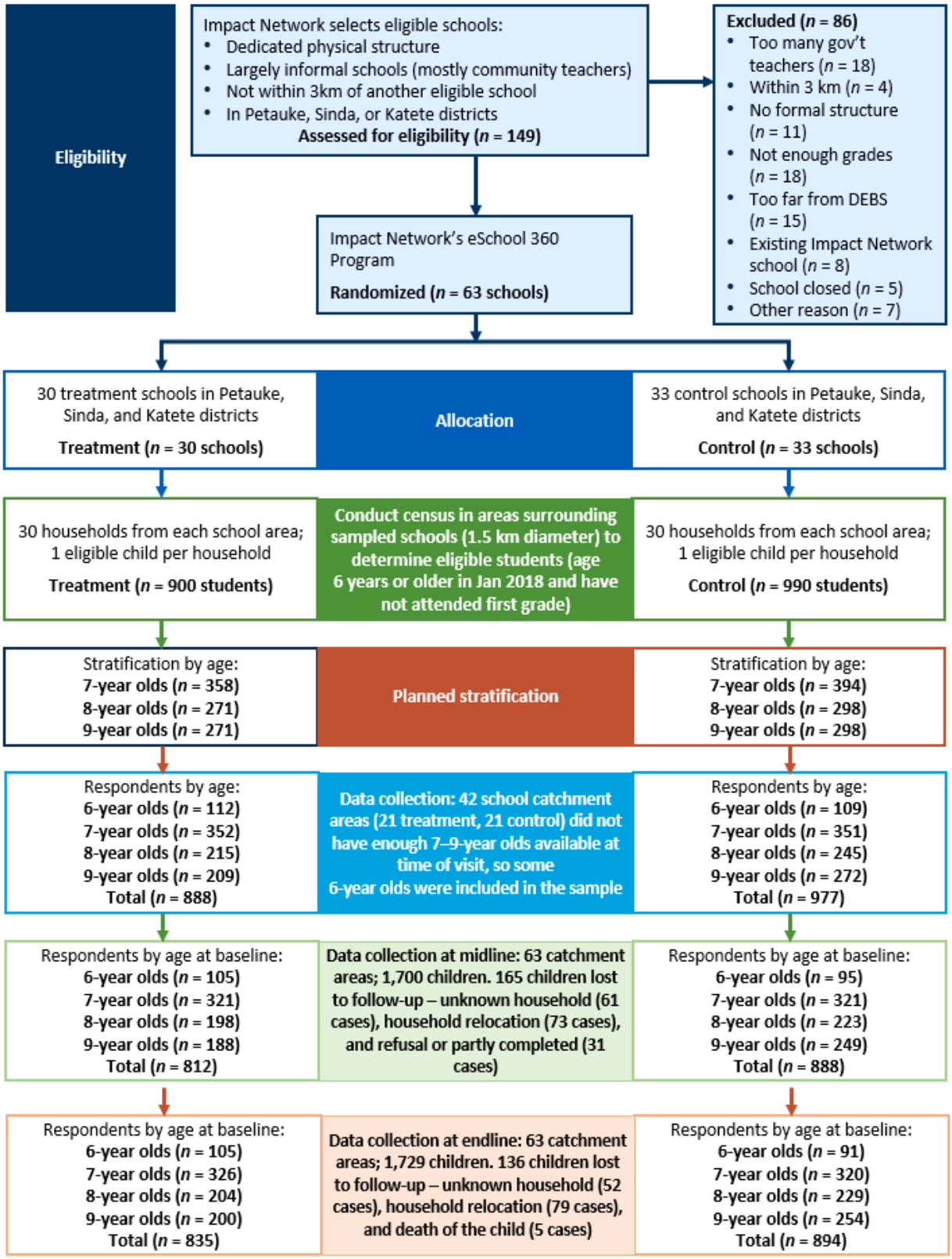
implementing. The enumerators used tablets with SurveyCTO software to conduct the baseline, midline, and endline data collection, thus improving the quality of data collection, minimizing the need to clean data, and eliminating data entry. In addition, the tablets allow for the collection of GPS data, which enables the tracking of households from baseline and midline to endline data collection in the field.

The primary achievement indicators for numeracy, preliteracy, and literacy come from the EGRA, EGMA, and ZAT, which were already adapted and used in the same region as this study. In addition, we collected oral vocabulary assessment data on the students. These instruments have been translated and validated in the context of Zambia. We collected assessment data in primarily the Nyanja and Chewa dialects. Appendix C includes a detailed overview of the outcome variables. Further, we collected enrollment and attendance data during the endline household survey as well as parent perceptions of schools, teachers, their children's education, education expenditures, aspirations, and food security. Finally, we collected data on parents' aspirations about children's education and early marriage. The survey also incorporated moderating variables at different levels, including – (1) Student-level—gender and age; and (2) Household-level—distance from school, poverty level, parents' education level, and household size.

Figure 4 presents the consort flow chart that describes the sampling strategy, starting with the eligibility criteria, followed by the random assignment, the census, the planned stratification by age, and the practical decisions that we made during data collection. The consort flowchart also includes information about the attrition rates during the midline and the endline survey.

During the endline survey, we were able to track and administer surveys with 93% of children (N=1,729) of the 1,865 children in the baseline sample, which is higher than during the midline survey. Attrition was higher in Control areas, where we were unable to follow up with 9.0% of children (N=88), as opposed to 5.5% in Treatment areas (N=48). Reasons for attrition included – unknown or unidentified household in the community (52 cases), household relocation to a distant place (79 cases), and death of the child (5 cases).

Figure 4. Consort Flow Chart



5.3 Cluster-RCT Analysis

We used an ANCOVA model to estimate the ITT effect four and a half years after the start of the program. The ANCOVA approach uses a regression specification that includes the baseline measures of outcome variables as an additional explanatory variable. The evaluation design, with random assignment of schools to treatment, provides an unconfounded measure of the direct effect of the Impact Network intervention on student outcomes. Thus, the ITT effect of treatment on outcome Y_i can be estimated using the regression specification:

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

Here IN_i is an indicator variable equal to 1 if individual *i* resides in the catchment area of an Impact Network school and equal to 0 otherwise, S_i is a vector of dummy variables for the district strata, Y_{i2018} is a baseline value of the outcome of interest, C_i is a vector of other control variables, and ϵ_i is a conditionally mean-zero error term. We used ordinary least squares regression analysis with cluster-robust standard errors clustered at the school catchment area level to account for potential correlation in outcomes within a school catchment area.[8]

We registered the study protocol in the registry for international development impact evaluations at https://ridie.3ieimpact.org/index.php?r=site/page&view=about after the baseline survey and present more details of the pre-analysis plan in an inception report (de Hoop et al., 2017). The analysis plan pre-specified the use of EGRA, EGMA, ZAT, and oral vocabulary scores as primary outcome measures. The pre-analysis plan also pre-specified OLS regression analyses to estimate impacts on school attendance and enrollment, perceptions of school and education quality, and parents' and children's' aspirations each of which we present in this study. Further, it includes a section on the use of factor analysis to understand what latent constructs the assessments are tapping into. We decided to not implement the factor analysis and instead limit our analysis to analyses of EGRA and EGMA sub-scores, while applying corrections for multiple comparisons. The baseline report included a detailed overview of the construction of outcome variables (De Hoop et al., 2018).

We applied corrections for multiple comparisons to multiple outcome measures within the same outcome domain using the Benjamini-Hochberg procedure, as employed by Banerjee et al. (2015). Two of the outcome measures for the impact analyses are organized into two

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

domains: primary EGRA outcomes and primary EGMA outcomes. The EGRA and EGMA subtask outcomes are related within the domains, and therefore, the results are corrected for multiple comparisons. We pre-specified the corrections for multiple comparisons when we registered the study. In addition, we used randomization inference to test whether the ITT effects on primary outcomes were robust to any variations in the data that may have arisen from randomization.[9] Randomization inference tests the significance of the estimate by calculating the proportion of alternative realizations of the regression coefficient that are further away from zero than the estimated regression coefficient (Heß, 2017; Young, 2019).

We also tested whether the program had a differential impact by age; gender; region; socioeconomic status; caregiver's education; and baseline EGRA, EGMA, and ZAT attainment. For noncategorical moderators (baseline test scores), we constructed a binary variable equal to 1 for individuals with a value greater than the median and equal to 0 for those with a value less than the median. We then interacted the binary variable with treatment to test for whether the treatment was statistically different for groups with high and low levels of baseline performance. For socioeconomic status, we constructed an asset index using the calculated values from the first principal component of a list of assets as recommended by Filmer and Pritchett (2001).

We estimated two types of TOT effects, which reflect either the effects of actively participating or ever enrolling in an Impact Network school. To estimate TOT effects for children who actively participated in the program we first used the program assignment as an instrumental variable to predict enrollment in an Impact Network school over the last year. Second, we reported TOT for children who self-reported attending Impact Network schools at least three times in the week before the survey to estimate the impact of consistent school attendance using a two-stage least squares regression analysis.

To address concerns regarding composition effects and to enable the estimation of ITT effects, we randomly sampled 30 households from the census-generated sample frame for each of the sample schools. The visibility of the infrastructure of the Impact Network schools may result in a different composition of students attending the treatment schools relative to the control schools—for example, by influencing school enrollment and attendance. Such changes in the composition could result in a bias in the impact estimator that compares students enrolled in the Impact Network schools with students enrolled in other community schools because of differences in either observable (e.g., age, gender, parents' education level) or unobservable

[9] We did not pre-register the use of randomization inference.

(e.g., motivation, noncognitive skills) characteristics. For households with more than one eligible child, we selected the oldest child for inclusion in the sample. The introduction of Impact Network schools may have induced some eligible children to enroll in school (who would not have enrolled in school absent IN). Comparing the outcomes (e.g., test scores) of all eligible children allows for capturing any effects on inducing children to attend school.

To increase statistical power, we oversampled 8-year-olds and 9-year-olds from our Census-generated sample based on descriptive statistics about the age distribution from the Eastern Province (Ministry of General Education, 2016). These descriptive statistics indicated that 8-year-olds and 9-year-olds are more likely than 7-year-olds and 6-year-olds to be enrolled in first grade (Ministry of General Education, 2016). For our sampling strategy, we began by randomly selecting households with 7-year-olds, 8-year-olds, and 9-year-olds within a distance of 1.5 kilometers from each of the schools. However, we had to slightly adjust the sampling strategy because 42 catchment areas did not have sufficient numbers of 7- to 9-year-old children.[10] The sample was adjusted by randomly sampling additional 6-year-old children, followed by sequentially expanding the radius of the circle by an additional 0.5 kilometers until a sufficient sample was identified.

5.4 Cost-Effectiveness Analysis

To determine the cost-effectiveness of the program we combined the benefits estimated in our impact analyses with the collated costs from a school survey and information from Impact Network to carry out a cost-effectiveness analysis.

We collected cost data from two sources – (1) we worked with Impact Network to collect all expense reports from project implementation over 2017 through 2023. (2) we collected cost data directly from treatment and control schools during the endline survey, which included annual expenditures the schools made from government funds and from community funds. For treatment schools, we then added the school-reported expenditures and expenditures borne by Impact Network in carrying out the program activities, including retrofitting school infrastructure, maintenance of solar panels, teacher professional development and salaries, and ongoing expenditures on monitoring and implementation. For control schools, we considered all expenditures reported by schools, irrespective of the source of those expenditures.

We collected all costs in Zambian Kwacha, then converted to USD using the market exchange rates. Since program costs were realized over multiple years across the lifespan of the

[10] A "catchment area" is defined as the geographic area served by a particular school.

intervention, these costs need to be stated in base-year prices, after adjusting for inflation. We applied the USD inflation rate using the Consumer Price Index (CPI) to convert all costs to base year prices (2017 USD). We decided to account for USD inflation instead of Zambian Khwacha, because all program costs for Impact Network were funded in USD. We then estimated the present discounted value for future streams of costs (borne from 2018 through 2022) applying a discount rate of 10% (Dhaliwal et al., 2013). We then reapplied the USD inflation rate to indicate costs in end-year USD prices – which was 2019 for midline and 2022 for endline. Since we collected school-reported annual expenditure at endline only, we applied similar annual costs to the years prior to endline before adjusting for inflation as explained above.

We then estimated separate cost-effectiveness ratios, as the additional SD gains in reading and mathematics for each $100 spent on the program, at midline (14 months into the program) and at endline (four-and-a-half years into the program). To estimate these cost-effectiveness ratios, we used different scenarios with different assumptions, which we describe in more detail in our section describing the cost-effectiveness of the program.

5.5 Comparing Learning Outcomes with Government School Students

In addition to the cluster-RCT, we compared the EGRA, EGMA, and Zambian achievement test outcomes of children enrolled in Impact Network schools with children enrolled in five government schools in the same school catchment areas. We selected these government schools from a sample of government schools in the three districts. While the comparison with government schools did not enable us to estimate the impact of enrolling in Impact Network schools versus enrolling in government schools, the comparison serves as a useful benchmark to assess whether children enrolled in Impact Network schools learn more than, the same amount as, or less than children enrolled in government schools. In this way, the comparison can help to assess the feasibility of scaling up the program as part of the Zambian national education system.

5.6. Process Evaluation Study Design

AIR also conducted a process evaluation to assess the fidelity of implementation of the eSchool 360 program. The process evaluation complements the impact evaluation by enabling us to explore informal patterns and unanticipated interactions in addition to formal activities and anticipated outcomes, while also giving us flexibility to explore unforeseen areas of interest. Incorporating a qualitative process evaluation into a larger impact evaluation allows us to investigate why we do or do not see impacts and to explore nuances and variations in our sample, particularly with regard to sensitive or complex topics. Further, this component of the study provides rich contextual information to complement the numerical findings of the impact evaluation while also allowing us to triangulate and explain quantitative findings.

Qualitative Data Collection & Analysis

We used three primary approaches to collect two rounds of qualitative process evaluation data: 1) key informant interviews (KIIs) with teachers, Teacher Supervisors, and IN program staff; 2) focus group discussions (FGDs) with students and parents; and 3) classroom observations.

KIIs—KIIs with teachers, Teacher Supervisors, and IN staff are an important element of our evaluation design because they shed light on perceptions of the quality of the eSchool 360 model both in terms of how the technology and teacher trainings have affected classroom instruction and how the program engages with school communities. Throughout KIIs, respondents discussed the perceived impact of the program, and we explored respondent perceptions of the program's activities. These findings allow us to compare the planned interventions with the actual implementation to see how the program affects teachers' behaviors and whether students are learning the intended content.

FGDs—FGDs provide a context in which participants feel comfortable and empowered to discuss the evaluation topics with their peers and the carefully trained facilitators. We created a social dynamic that encourages participants to reflect upon their opinions and experiences and express them verbally. FGDs provided an opportunity to gain a deeper understanding of the factors enabling or inhibiting implementation of the eSchool 360 curriculum in the classroom. From FGDs, we investigated experiences that parents and PTA members had when communicating with IN staff, generally engaging with the program, and using the school in the evenings and on weekends. With students, we discussed the techniques that teachers used in the classroom as well as the friendliness of the school environment.

Classroom Observations[11]***—***We conducted formal classroom observations at a total of 13 treatment schools across the districts. Classroom observations enabled us to examine the extent to which teachers at IN schools were effectively using e-learning technology and applying pedagogical techniques gained during eSchool 360 teacher trainings. Observations also noted the eSchool 360 environment, particularly with respect to its child-friendliness. We conducted classroom observations in one randomly selected academic class at each of the selected schools using a piloted and validated instrument.

KIIs, FGDs, and classroom observations took place in treatment schools in the districts of Katete, Petauke, Sinda and Lusungazi. In each district, the research team worked with IN staff to select one or two schools to visit. The selection was made based on a list of observable characteristics including school size, distance from district center, and the distribution of

[11] Only 2 of the 13 classes observed did use the tablets and projectors in that particular day's lesson, suggesting that we need to exercise some caution in drawing conclusions about the use of technology based on the observations.

high/low performing schools based on student learning outcomes from prior years. The research team worked with IN leadership to identify relevant IN staff, Teacher Supervisors, and teachers to interview for the KIIs. IN leadership and school staff assisted in the recruitment of parents and students for the FGDs. IN staff and teachers selected students from Grade 5 to ensure variation in gender (similar number of boys and girls) and academic performance (including high and low performing students). Further, the IN team selected parents from Grade 5 students to ensure the inclusion of some PTA members in FGDs. Table 1 provides a summary of the qualitative data collection approaches and respondents.

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

AIR partnered with Palm Associates to conduct qualitative data collection. Two experienced enumerators were recruited by Palm Associates to carry out interviews, focus groups, and classroom observations. Researchers spoke the Nyanja and Nsenga dialects, and qualitative instruments were translated into these languages during a three-day training. The researchers digitally recorded all KIIs and FGDs on portable digital recorders. They downloaded recordings to field laptops each day, renamed according to an anonymized code system held in an encrypted spreadsheet and copied to external media for backup. At the end of each day, the field researchers transcribed the handwritten field recording sheets to word processed documents, translating the material as necessary.

Researchers used audio recordings to supplement and validate the written transcriptions and translations. They assigned all transcriptions new names according to the code system in order to ensure data and informant confidentiality. The data collection supervisor ensured that the work was carried out in accordance with research plans, on schedule, ethically, safely, and completely. The supervisor also reviewed transcribed interviews to ensure high standards of data quality.

All data from KIIs and FGDs were coded and analyzed using the NVivo qualitative software program. Our team created a preliminary coding outline and structure based on the research questions, interview protocols, and memos of ideas that emerged during data collection. This coding outline served as a tool to organize and subsequently analyze information gathered in the interviews and focus groups. We created a list of definitions for the codes to ensure that coders categorized data using the same standards. After inputting the raw data into NVivo, coders selected a sample of interviews to double-code in order to ensure interrater reliability. Using these coded data, the qualitative team applied grounded theory to identify themes, categories, and theories that emerged from the data and that confirmed or refuted the researchers' initial impressions. In other words, rather than basing the analysis on a hypothesis, the researchers created concepts and categories based on the data, refining the concepts throughout the process to ultimately inform the overall findings. During this process of data reduction, researchers characterized the prevalence of responses, examined differences among groups, and identified key findings and themes related to the research questions.

Additionally, where applicable, the qualitative team also provided insights from the qualitative findings into the perceived impacts of the program activities.

School-Level Quantitative Data Collection & Analysis

At endline, we conducted school surveys with the head teacher or the longest serving teacher in the school to estimate program impacts and costs at the school-level. We collected data on school characteristics including school size, grade level, student and teacher enrollment; implementation of the Catch-Up program (this program implements the "Teaching at the Right Level" methodology, which groups children based on their learning level, and was scaled up in all community schools in Zambia in between the midline and the endline survey); teacher professional development; and student assessments. Enumerators also reported data on school environment based on their observations, including condition of school buildings, grounds, availability of water, sanitation, and hygiene services, and other physical infrastructure. We collected data on school expenditures over the last school year, which informed our cost-effectiveness analysis. School expenditures were collected based on school records, where available, and head teachers' estimates in cases where written records were unavailable. We collected data on expenditures made from funds received from the community as well as government.

We estimated impacts at the school-level by regressing implementation characteristics on the treatment assignment using bivariate regressions. We collected data from 62 of the 63 schools because the data collection team was not able to find the 63rd school during the fieldwork (it is likely the school no longer exists based on more recent information from Impact Network). For

the missing school, we imputed cost information based on the average costs in the same district, for schools assigned to the same Treatment status.

6. Results

6.1. Attrition and Program Take-Up

***Endline Attrition*—**The endline survey was delayed due to COVID-19. While originally planned for 2020, we eventually collected data in 2022. This resulted in a longer than planned gap between baseline and endline, which may have increased the likelihood of attrition.[12] However, we were able to track and administer surveys with 93% of children (N=1,729) of the 1,865 children in the baseline sample. We thus achieved a lower attrition rate than during the midline survey.

For unidentified households, we used a combination of three approaches interchangeably to pinpoint the location of the household: (1) a snowballing tracking approach that involved finding households by asking other members of their community where they were; (2) a GIS approach that involved using GPS coordinates from baseline; and (3) an approach that relied on using telephone numbers collected at baseline and midline to contact missing households. Combining these approaches helped to reduce the attrition rate. We succeeded in tracking 93% of the households we interviewed during the baseline data collection.

We found no difference in baseline test scores between children in the treatment and control groups who remained in the study. That is, we did not find evidence for differential attrition, thereby preserving the study's internal validity created through randomization. We also did not find statistically significant differences in baseline values of EGRA, EGMA, ZAT, and Oral Vocabulary scores between children who remained and children who withdrew from the sample. At the same time, it remains important to note that attrition was slightly higher in control areas (at 9.0%) compared to treatment areas (at 5.4%). This small difference is statistically significant but unlikely to impact the results of the analysis. Table 2 below shows our test for differential attrition.

[12] This larger gap may also have contributed to the ceiling effects for the ZAT and the oral vocabulary test.

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.

6.2. Baseline Equivalence

At endline, we were able to follow up with 1,729 children out of 1,865 children, achieving a lower attrition rate than during the midline survey. We did not find statistically significant differences in the large majority of background characteristics across households residing in IN catchment areas and control households. Child and household-level characteristics, including district of residence, household perceptions of poverty, and household distance from school were similar across treatment and control households. Overall, almost half of the sample was age 8 years or older during the baseline survey, and slightly more than half of the sample was male. Most of the sample resided in Petauke[13] District (compared with Katete or Sinda) and considered itself to be "very poor" or "poor". More than half of the sample reported that the child caregiver had attended school, and the average distance from school was approximately 0.9 kilometers. We report descriptive statistics on sample background characteristics of the households included in the endline survey by treatment and control areas in Table 3.

Although we found some statistically significant differences between treatment and control households, none of the differences were larger than 0.25 standard deviations. A study by Imbens (2015) indicates that statistically significant differences that are smaller than 0.25 standard deviations are unlikely to cause bias. However, we did observe some small but statistically significant differences when comparing sample means. For example, children in treatment households were less likely to be aged 8 years or older (48%) during the baseline

[13] This sample includes respondents from Lusangazi district. Lusangazi district was made independent from Petauke district in 2018.

compared with control households (54%). Children in the control group also scored higher on baseline EGMA scores than children in the treatment group. All regressions estimating the ITT and TOT effects control for variables that showed some imbalance at baseline to generate more precise estimates. Specifically, we control for household distance from school, baseline EGMA score, and child age at baseline.

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)		
EGMA (proportion correct)	894	0.078	835	0.060	0.018**	0.170
	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<0.05.

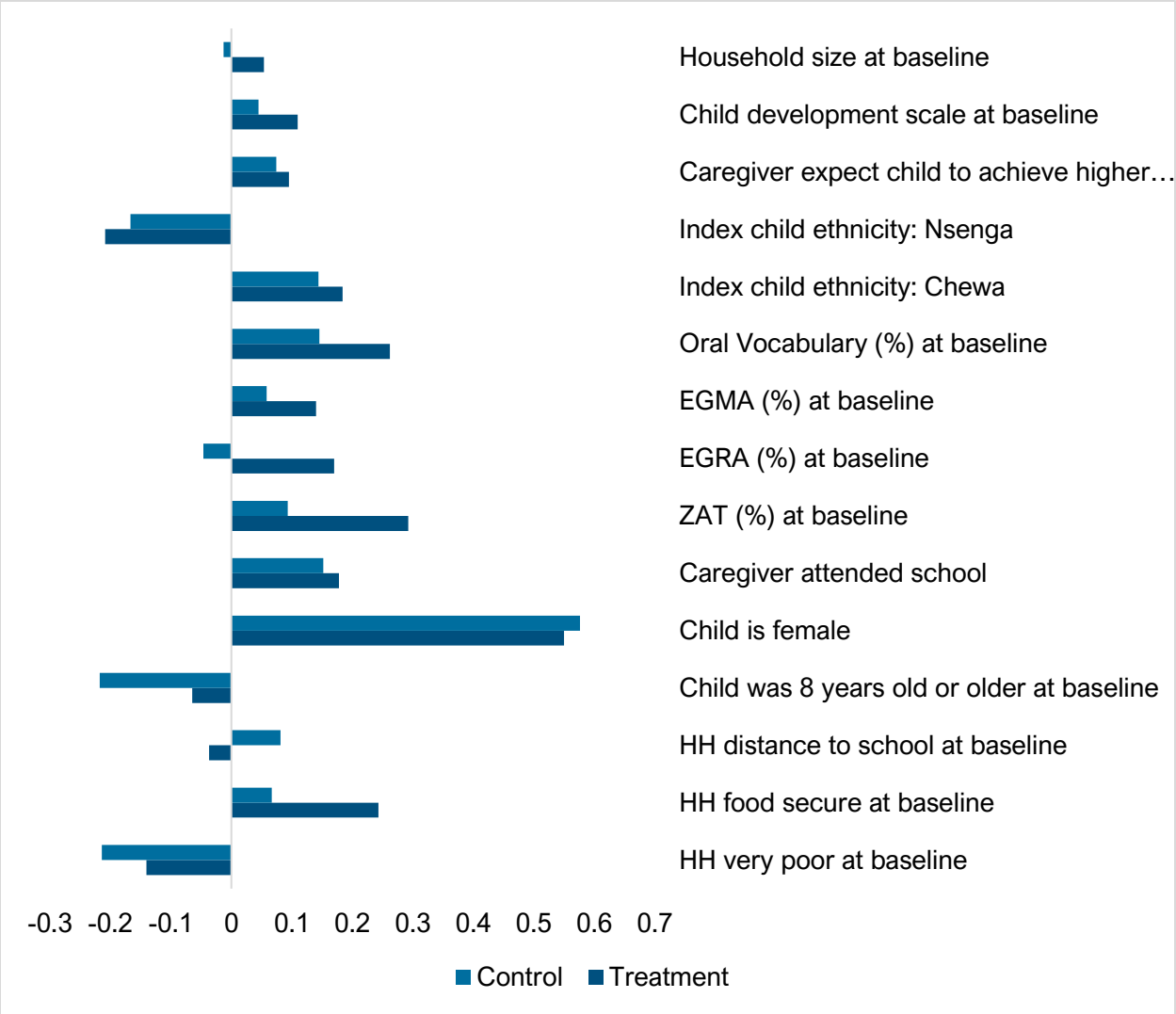
6.3. Sample Characteristics by School Enrolment

Program take-up, as measured through enrolment in IN schools, is one of the main channels for any impacts of the eSchool 360 program we observed during the midline survey. Before diving into the impact estimates therefore, it is important to understand the levels of school enrolment, and the sample characteristics of children who enroll in schools in treatment and control areas.

At endline, 67% of treatment children and 68% of control children reported attending school (though not necessarily in an Impact Network grade) in the current year or in the previous year. Thus, there was no observed difference in school enrolment between the treatment and control group at endline, which is different from the midline when 62% of treatment children and 55% of control children had enrolled in school. However, since treatment children started schooling earlier than control children (as shown by the midline enrolment rates), they were enrolled in higher grades at endline. Among those who attended school in the current year at endline, 42% of children in the treatment group were in grade 5 or higher, compared to 25% of children in the control group.

Overall, we find similar differences in enrollees and non-enrollees across the treatment and the control group. For example, across both groups, children who were enrolled in school were less likely to be very poor, older, and belong to the Nsenga ethnicity, than children who were not enrolled in school. Enrollees were also more likely to be from households that were food secure, and had higher scores in baseline assessments (ZAT, EGRA, EGMA, and Oral Vocabulary) compared to non-enrollees. Perhaps the most important difference between the two groups relates to the sex of the child. Enrollees were substantially more likely to be female than male (despite 47% of the total sample being female). Qualitative and quantitative data indicate that boys were more likely to be involved in farming and cattle herding, at both midline and endline, than girls, which likely contributed to higher drop-out rates (we did not find lower enrollment rates for boys during the baseline). We report these results in Figure 5.

Figure 5. Standardized Mean Differences Between Enrollees and Non-enrollees



6.4. Years of Schooling and Treatment Intensity

Understanding program dosage and treatment intensity is important to contextualize the evaluation findings. Because the program was only implemented in two cohorts, students who repeated classes or enrolled in school later did not benefit fully from the eSchool 360 program. In fact, only 49 percent of the sampled children in the treatment schools ever enrolled in an Impact Network school, and 20% of the children in the treatment group received more than 4 years of the program.

While enrollment rates at endline were higher than during the midline, enrollment increases since the midline were larger for the control group. The control group had lower enrollment rates at midline. As shown in Table 4, 39% of the children in the treatment group and 37% of

the children in the control group were not enrolled during the endline. Only 25% of the treatment group received the eSchool 360 program at endline (that is, they were in grades 5 or 6 and attended treatment schools). Among the same sample, 43% of the children in the treatment group was not enrolled in school at midline, compared with 50% of the children in the control group. 35% of the children in the treatment group received the eSchool 360 program at midline (that is, they were in grades 2 or 3 and attending treatment schools). These numbers suggest that a relatively small percentage of children in the treatment group actually received the eSchool 360 program, and an even smaller percentage received the program for the full four years. The expected effect size of the program reduces because of the smaller number of children participating in the program, which also reduces the statistical power of the study.

Table 4. Grade Distribution (sample with information at midline and endline)

Grades	Treatment=1		Treatment=0	
	Midline	**Endline**	**Midline**	**Endline**
Grade 1	21%	6%	24%	7%
Grade 2	33%	6%	21%	9%
Grade 3	2%	12%	4%	14%
Grade 4	0%	12%	0%	18%
Grade 5	0%	24%	0%	14%
Grade 6	0%	1%	0%	2%
Grade 7	0%	0%	0%	0%
Not currently enrolled	43%	39%	50%	37%
Total	778	778	848	848

Note. Cohorts receiving the eSchool 360 program are highlighted in table cells.

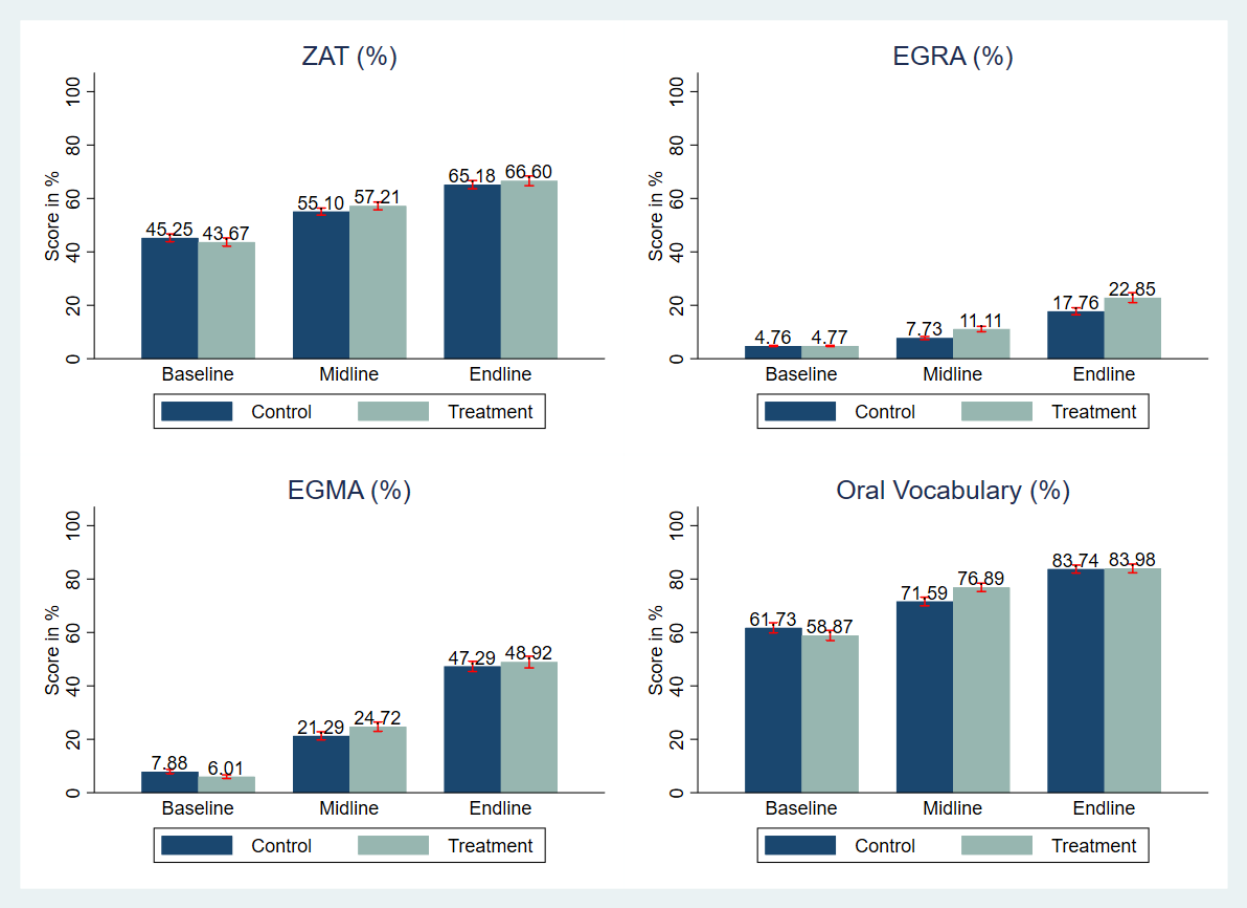
6.5. Impact Evaluation Results

Primary Outcomes—ZAT, Reading, Mathematics, and Oral Vocabulary

Intent-to-Treat Impacts on Primary Outcomes

Initial comparisons in the test scores between the treatment group and the control provide indications for positive program effects on literacy scores, but we found only small differences in mathematics scores, and differences in ZAT and oral vocabulary scores were largely absent 4.5 years after the start of the program. Figure 6 shows the mean percentage scores (as % correct) across treatment and control groups in the two rounds of testing along with 95% confidence intervals.

Figure 6. 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.

Results from the formal ANCOVA framework showed positive point estimates for all primary test scores, but we only found consistently statistically significant effects on literacy scores (see Table 5). The findings indicate that residing in IN catchment areas (treatment assignment) led to a 0.26 standard deviation or 5.1 percentage points increase in EGRA scores ($p < 0.01$). All coefficients for the other test scores suggest a positive change compared with baseline but we find no statistically significant ITT effects.

These results suggest that by 2022, the differences observed at midline had narrowed significantly between the treatment and control groups except for literacy. Various reasons may have contributed to the lack of impacts on non-EGRA outcomes including – 1) ceiling effects for some outcomes, especially the Zambian Achievement Test and Oral Vocabulary test scores [see

Figure 5]; and 2) COVID-19 school closures and lack of treatment compliance. Finally, while we find large positive effects on EGRA, the average EGRA scores were still relatively low in both treatment and control groups (less than 25%), and effect sizes were smaller than during the midline assessment. Since the ITT estimates compare students who were eligible to enroll in school, it is also important to remember that 39% of treatment students and 37% of control students were not actually enrolled in school at endline, which may drive the low average performance.

Table 5. 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

Clustered standard errors in parentheses. * p<0.1 ** p<0.05; *** p<0.01.
Estimations control for district fixed effects, baseline outcomes, age at baseline, distance from school at baseline, and baseline EGMA score.

ITT impacts on EGRA subscores show consistent findings, we find large and statistically significant impacts (over 0.20 standard deviations) on 6 out of 8 EGRA subscales (Table 6). While we also see positive point estimates for 8 of the 10 EGMA subscales (Table7), only three of these were statistically significant (pattern completion, addition questions 2, and subtraction questions 2). Addition questions 2 and subtraction questions 2 have a relatively higher difficulty level compared with the other subtasks, indicating some evidence that Impact Network may improve performance on more challenging EGMA tasks. However, as shown in Tables A1 and A2 in the Appendix, these results were not robust to correcting for multiple comparisons using the Benjamini-Hoschberg method (as shown in the adjusted q-value).

Table 6. 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

Clustered standard errors in parentheses. * $p<0.1$ ** $p<0.05$; *** $p<0.01$. 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.

Table 7. 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

Clustered standard errors in parentheses. * $p<0.1$ ** $p<0.05$; *** $p<0.01$. 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.

Heterogeneous Effects

We estimated treatment effects by subgroups, and found evidence for some variation in the effects by gender and baseline assessment scores (Figures 7 through 10). Children who scored above median on ZAT, oral vocabulary or EGMA at baseline saw positive and statistically significant impacts on ZAT and EGMA (in addition to EGRA). One potential reason for the larger effects on students with higher baseline scores is that they were more likely to return to school after school closures due to COVID-19. Analyses of enrollment rates suggest that children with higher assessment scores at baseline were statistically significantly more likely to continue enrolling in school after the pandemic. These higher enrollment rates may have contributed to positive impacts driven by the higher quality of education in the Impact Network schools.

Additionally, we find meaningful and statistically significant treatment effects for boys, of 0.28 SD on EGRA, 0.20 SD on EGMA, and 0.15 SD on oral vocabulary scores. Because boys are not overrepresented in the subsample with above median baseline scores, these results suggest that the heterogeneous impacts by sex are distinct from the heterogeneous impacts by baseline assessment scores. For example, while 55% of the full sample is male, 49% of the subsample scoring above median EGMA score at baseline is male.

The heterogeneous effects by gender are critical and somewhat puzzling considering the low enrollment and achievement rates among boys in the Zambian context. Higher drop-out rates likely decreased effect sizes for boys. A potential reason for the larger effects for boys is that the activity-based learning style was more suitable for boys than for girls in later Grades. It is possible, for example, that boys volunteered to present in front of the classroom more often. Alternatively, it is possible that the boys who remained enrolled in school were the most committed to education. However, this mechanism would normally not be different in the control group, limiting the explanatory power of this mechanism. While we find substantial heterogeneous effects by gender, we do find that the statistical significance of impacts on boys is not robust to correction of standard errors for multiple hypotheses, as shown in Appendix B.

One potential reason for the larger effects on students with higher baseline scores is that they were more likely to return to school after the school closures due to COVID-19. Analyses of enrollment rates suggest that children with higher assessment scores at baseline were statistically significantly more likely to continue enrolling in school after the pandemic. These higher enrollment rates may have contributed to positive impacts driven by the higher quality of education in the Impact Network schools. We did not find strong evidence for other heterogeneities in the impact estimates. In general, we need to exercise some caution in interpreting these results, however, because we may not have adequate statistical power to detect small heterogeneities in the effect sizes with sufficient precision.

Figure 7. Heterogeneous Effects – ZAT

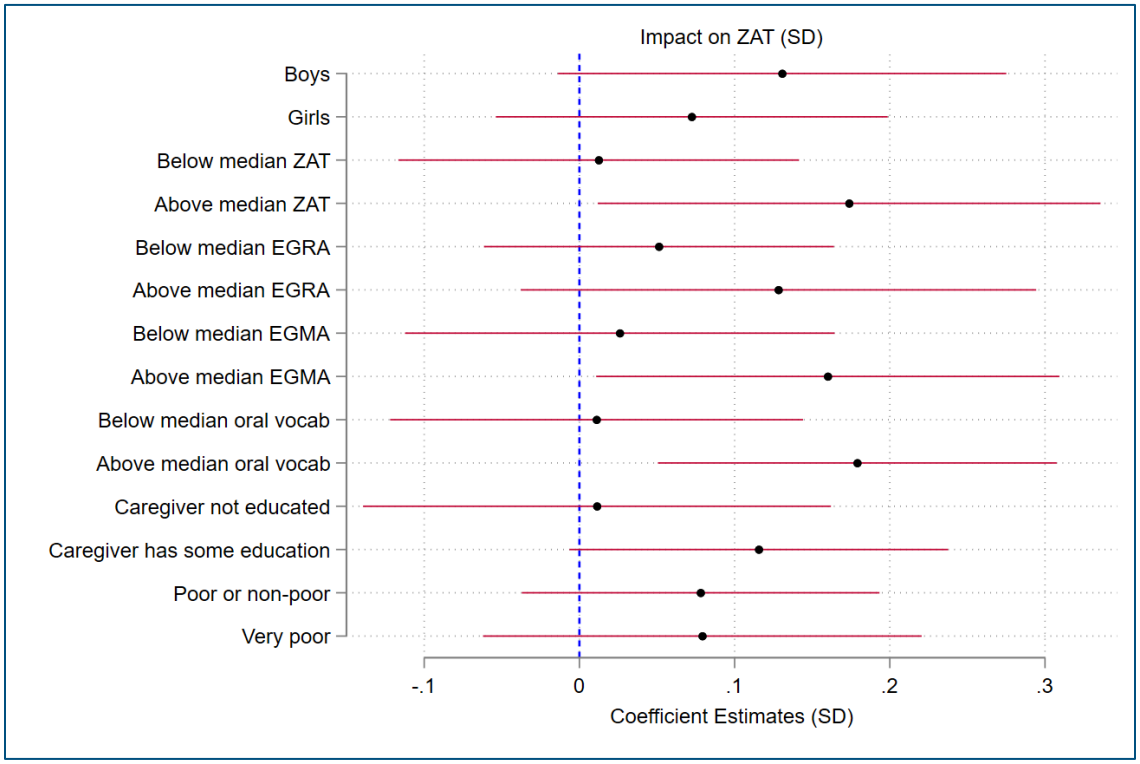


Figure 8. Heterogeneous Effects – EGRA

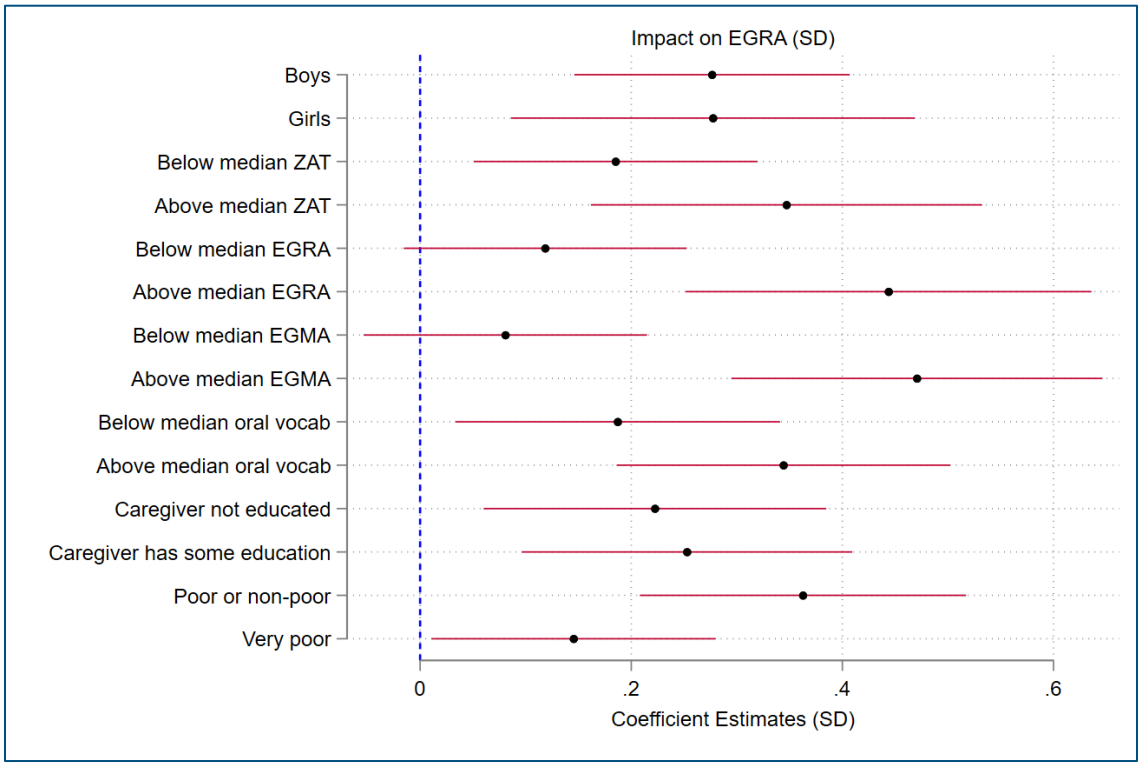


Figure 9. Heterogeneous Effects – EGMA

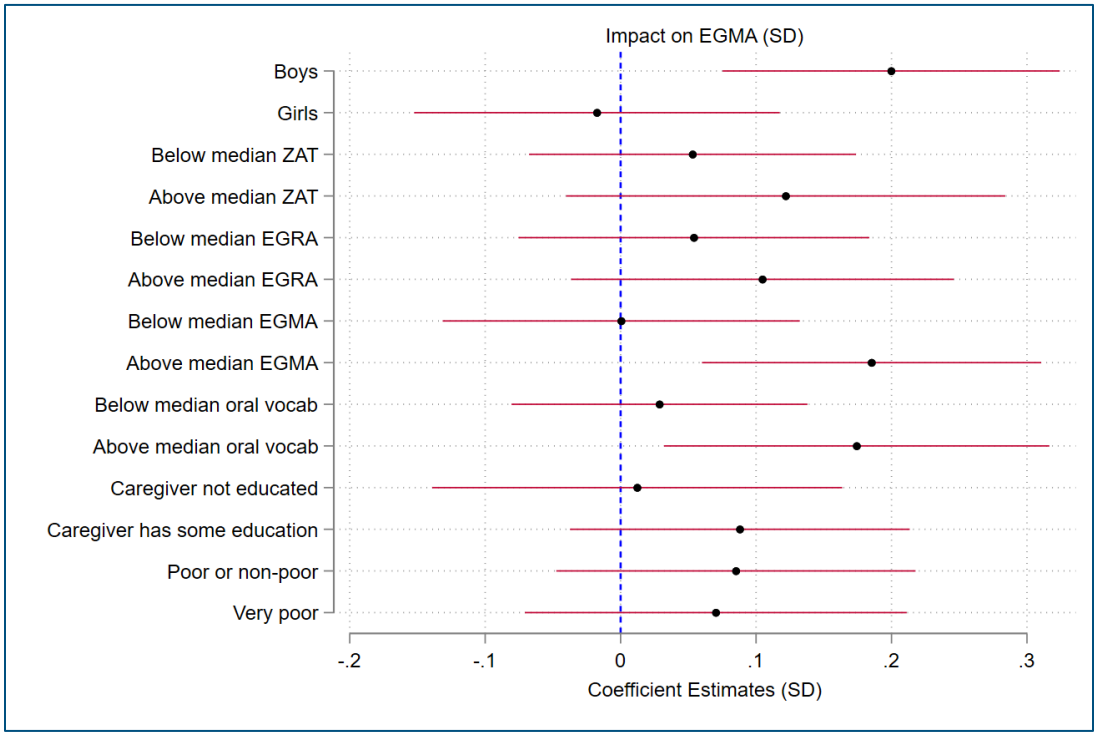
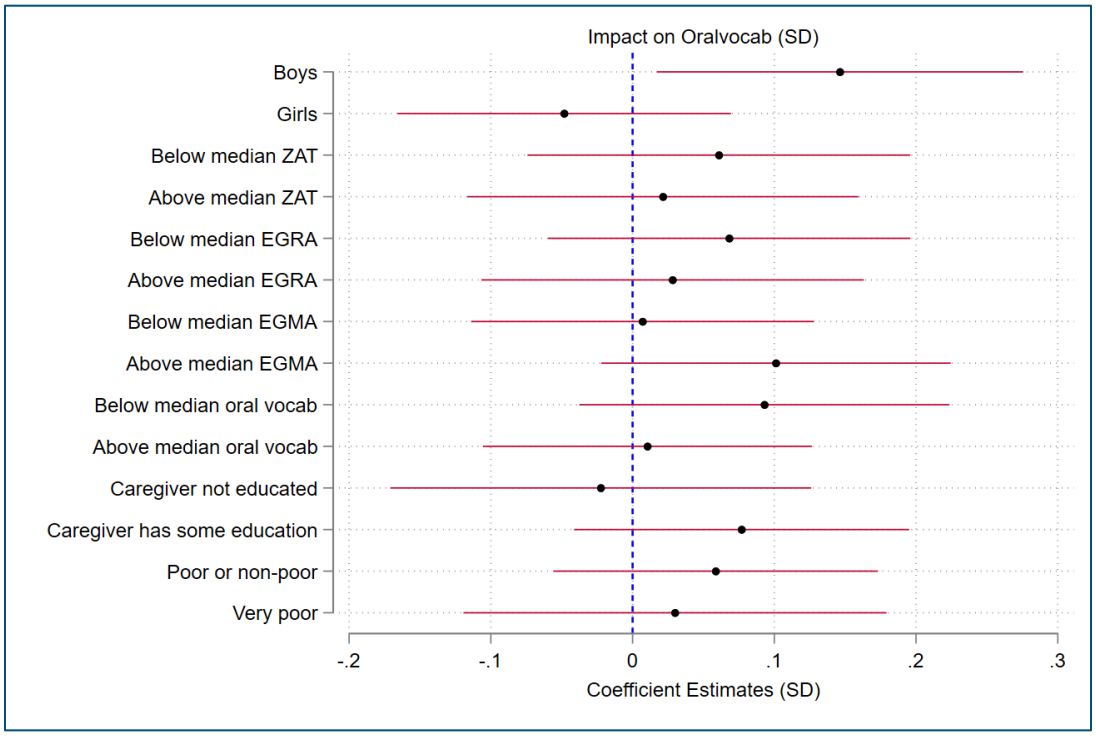


Figure 10. Heterogeneous Effects – Oral Vocabulary



Treatment Effects on the Treated

As expected, TOT effects of attending schools at least three times in the week before the survey are substantially larger than ITT effects, though impact estimates for non-EGRA assessments continue to be statistically insignificant (Table 8). Specifically, attending an Impact Network school for at least three days in the past week improved EGRA outcomes by 10 percentage points or 0.51 standard deviation. It is possible that the impacts on mathematics outcomes are not statistically significant because of a reduction in statistical power. The point estimate of 0.16 standard deviations is meaningful, but an increase in the standard error relative to the ITT estimations may have contributed to a reduction in statistical significance.

Results from an estimation of the impact of having ever enrolled in an IN school over the year prior to the survey show similar results demonstrating positive effects on EGRA but statistically insignificant effects on other outcomes. Enrollment in an IN school at endline increased EGRA scores by 8.1 percentage points or 0.42 standard deviations. We show all TOT effects formalized in an instrumental variable framework in Table 9, where the first column shows the relationship between IN school assignment and IN school enrollment; and the remaining columns show second-stage estimates reflecting the treatment effects on the treated.[14]

One important reason for the smaller intention-to-treat impacts is the relatively low dosage of the program among children in the treatment group. As shown in Section 6.4, only 49 percent of the sampled children in the treatment schools ever enrolled in an Impact Network school. 20% of the children in the treatment group received 4 years of the program (see Table 3).

[14] Although not shown here, we also estimated ToT effects of attending an Impact Network grade at endline (that is, child enrolled in an Impact Network school in grade 5 or 6 at endline). The estimates are larger in magnitude, but there is no difference in the statistical significance of estimates, likely because of lower power. Being enrolled in an Impact Network grade led to 1.1 SD higher EGRA scores ($p<0.01$). Other estimates were statistically insignificant, but positive: they suggest an increase in the ZAT score of 0.36 SD, in the EGMA score of 0.34 SD, and an increase in the oral vocabulary score of 0.20 SD.

Table 8. 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

Clustered standard errors in parentheses. * $p<0.1$ ** $p<0.05$; *** $p<0.01$

Estimations control for district fixed effects, baseline outcomes, age at baseline, distance from school at baseline, and baseline EGMA score.

Table 9. 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

Clustered standard errors in parentheses. * $p<0.1$ ** $p<0.05$; *** $p<0.01$.

Estimations control for district fixed effects, baseline outcomes, age at baseline, distance from school at baseline, and baseline EGMA score

Qualitative reports of perceived effects on student learning

While quantitative data only showed consistent positive effects on literacy outcomes, qualitative data indicate that respondents consistently perceived positive effects of the Impact Network program on student learning. One possible reason for the more positive perceived effects discussed in the qualitative data analysis is that parents compared Impact Network schools with government schools and not necessarily with other community schools. Here, we discuss perceived effects of the Impact Network program on students' performance, literacy, and numeracy, based on qualitative data.

Parents generally believed that Impact Network students outperformed students in government schools, and stated that Impact Network students were sharper and ahead of students in government schools. For example, one parent from Chatata declared, "*I think the children here at Impact Network schools are well ahead of the other schools in terms of how the children learn,*" while a parent from Lenda noted, "*The pupils [at the Impact Network school] are far much better than the pupils at the other schools… ever since Impact came we have seen massive improvement in the way the pupils learn.*" This view was supported by Impact Network staff who alluded to data showing that Impact Network students outperformed those in government schools.

Impact Network staff pointed to the structure of the Impact Network program, with its use of technology, continuous improvement and supervision for teachers, and high levels of teacher attendance, as ways in which the Impact Network program improved learning for students. For example, an operations manager noted, "*Students who are under Impact Network do much better than other students in other grades. Maybe it is because of the technologies that we use and teachers are always there to teach.*" Further, another Impact Network staff member explained that the qualification and continuous improvement of teachers had improved the quality of education at Impact Network schools.

> *The programs and also the work that we do… has really improved the quality of education in the schools where we operate. One it is because we use qualified teachers unlike in other community schools there is just volunteers… And we train them, we observe lessons to make sure that whatever the teacher is teaching is in line with the curriculum and also provide immediate feedback and support to our teachers.*

In addition, some Impact Network teachers and staff members noted that Impact Network schools provided a conducive learning environment for students by prohibiting physical punishment and ensuring that students had basic learning materials. As a teacher from Lenda stated,

We are looking at the rules which the Impact Network is giving us as teachers on how to handle Impact Network learners... So they are handled in a way that, punishment we are even guided in which punishment to give them, there is no beating, so they have … It is a conducive environment for them.

Perceived effects on literacy and numeracy

The qualitative data showed a widespread perception among parents, teachers, and Impact Network staff that Impact Network had improved pupils' reading skills, with many respondents noting that children who were part of the Impact Network program learned to read better and at earlier ages than those who attended government schools. A parent from Petauke, for example, stated, "*This [Impact Network] school is very good because children at this school are able to read at a young age they can come home and be able to read what they have learned, so this school is very good and the teachers are very good as well*". Impact Network teachers pointed to the quality of teaching and continuous monitoring of Impact Network schools as reasons for improved performance in literacy. For instance, a teacher from Petauke explained,

[Impact Network] teachers are committed in such a way that whatever project is introduced teachers implement it together with the children… our Impact Network students are doing very well like in literacy levels compared to… government learners, because our children for the grade fives most of them were able to read [in Cinyanja] by grade one term three… and by the time they were reaching grade four, most of them were able to read English. Teaching in government schools they are not monitoring regularly but for Impact Network we are monitored weekly.

Qualitative results did not indicate perceived effects of the Impact Network program on numeracy. Numeracy was considered less challenging to teach than literacy according to Impact Network teachers. However, this result is consistent with the quantitative analyses showing positive point estimates but statistically insignificant effects on mathematics (though the statistical insignificance may be related to limited statistical power).

Intermediate Outcomes – Mechanisms of Impact

Intent-to-Treat Impacts on Schooling Outcomes

While treatment students were more likely to attend higher grades than control students, we did not find statistically significant effects on student enrolment and attendance in school. The midline study showed that treatment students were 8 percentage point more likely to enroll in school 14 months after program start, indicating that treatment students started primary school earlier than control students. In line with those findings, the endline data showed that treatment students were 12 percentage points more likely to attend grade 5 or higher. Table 10

shows these results. While not shown in the table, this difference between treatment and control groups was seen among both boys and girls, though boys were substantially less likely to attend grades 5 or 6 compared to girls. At endline, 21% of boys in the treatment group and 10% of boys in the control group were attending grades 5 or 6. In both the treatment and the control groups, girls were almost 13 percentage points more likely to attend grades 5 or 6.

Table 10. ITT Impacts on Schooling Outcomes

	Index child is enrolled at any school (last year or this year)	Weekly attendance (number of days attended)	Grade 5 or 6 by Endline
Treatment	-0.019	-0.019	0.118***
	(0.027)	(0.144)	(0.024)
Comparison Mean	0.679	2.621	0.162
N	1729	1729	1616

Clustered standard errors in parentheses. * $p<0.1$ ** $p<0.05$; *** $p<0.01$
Estimations control for district fixed effects, age at baseline, distance from school at baseline, and baseline EGMA score.

Bivariate correlations suggest that school enrollment, weekly attendance, and attending grade 5 or higher by endline were strongly predictive of higher scores on all four assessments, suggesting that attending higher grades (which also implies more years of schooling) may have contributed to improved learning outcomes. While we need to exercise caution in interpreting the results causally, children in grade 5 or higher at endline scored more than one standard deviation higher than other children on three of the four assessments (the differences for oral vocabulary were smaller and not statistically significant).

Perceived Effects on Teacher & Student Attendance

Qualitative data suggests that the Impact Network program encourages attendance for both teachers and students. The results on the perceived effects on student attendance require some caution in interpreting, however, because they are inconsistent with the quantitative results demonstrating statistically insignificant effects on school attendance. It is again possible that the qualitative results indicate more favorable impacts because respondents compare educational outcomes with government schools.

Teacher attendance

respondents agreed that Impact Network teachers attended work regularly and that there was no seasonal variation in teacher attendance. Parents and Impact Network staff mentioned that teachers lived either next to or in the same villages as schools, which contributed to the high

rate of attendance. The Impact Network program actively monitored teacher attendance through requiring teachers to sign in daily in the school register and report absences to the Teacher Supervisors, who conducted periodic monitoring through school visits. Teachers faced discipline if they did attend work regularly and report and justify their absences. Teacher attendance data were entered into the CommCare system on a weekly basis for analysis.[15] According to Impact Network staff, teachers were only absent for reasons such as illness and family emergencies, and School Support Officers (SSO) were expected to fill in for teacher absences.

This aligns with the experiences of parents, many of whom reported that their children had never been without a teacher. For instance, a parent from Chamboo school in Katete noted, "*The teachers come because I have never heard that a grade 5 pupils has missed class. They are always at school*." In addition, a parent in Chatata school in Sinda explained that substitutes were readily available in case of teacher absences. "*The Impact teachers are always in school and whenever they are sick immediately they submit a report to their office, and a replacement teacher is sent to come and teach.*"

Student attendance

Impact Network schools track student attendance daily through a register, and attendance data is entered into the CommCare system on a weekly basis. Based on an analysis of attendance trends, Impact Network staff flag absent students for follow up visits.

Respondents across schools and districts tended to agree that the Impact Network program encouraged student attendance. Most parents reported that their children attended school regularly. However, students and Impact Network staff stated that some children did not attend school regularly. According to Impact Network staff, approximately half of the students were absent more than a few days each term, but chronic absenteeism numbers were low. That some students attended school irregularly is consistent with the quantitative endline data, showing statistically insignificant effects on school attendance.

Barriers to student attendance included distance from school, illness, funerals, and children being asked to do other work, such as farming and cattle herding. According to key informants, attendance varied by seasonality and gender. Impact Network staff stated that there was higher absenteeism in the rainy season due to farming and transportation difficulties. Boys tended to be more absent than girls due to their involvement in cattle herding. A teacher from

[15] We provide more details on the CommCare system in our process evaluation of the SimPrints program (Dias et al., 2023).

Manjazi summarized the main causes for absenteeism in his school, noting the seasonal variation.

> *Children might miss school sometimes, it may be because of illness, natural disasters like rain, funeral, sometimes like in our case here when it is rainy season their parents engage their children in farming activities.*

While most parents stated that their children attended school regularly, several mentioned Impact Network's attendance tracking and follow up system as a motivator to student attendance. These parents noted that follow ups encouraged them to send their children to school regularly, which they saw as a positive outcome for children. A parent from Lusangazi illustrated this perception in the following way:

> *Impact makes sure they make follow ups on children they notice are not going to school. So us parents, we know that if we stop our children from going to school, we are going to be pursued by those from Impact. This scares those parents that have this habit, and this is good because I did not go to school, so I want my children to learn.*

This perception from parents, along with reports from Impact Network staff, suggest that the attendance tracking and follow up system is effective in encouraging parents to send their children to school. We provide more details on these perceptions in the process evaluation of the SimPrints program (Dias et al., 2023).

Intent-to-Treat Impacts on Caregiver Satisfaction, Child Development & Parent Engagement

We assessed program impacts on three intermediate outcomes related to child growth and development, and parent and child engagement in school. First, we estimated impacts on child development, which we measured by using an index constructed based on answers of parents about the behavior of their children. Second, we examined impacts on parents' perceptions of school and education quality by estimating impacts on 4-point Likert-scale questions related to parents' general perceptions on the quality of education as well as to more specific attitudes related to Impact Network's activity-based curriculum, the use of technology in the classroom, and teachers' pedagogical practices. Finally, we examined program impacts on children's level of engagement in the classroom, the time devoted to activity-based learning activities, the use of technology in the classroom, and children's interaction with teachers.

While we did not find statistically significant effects on the child development scale, parents in the treatment group reported statistically significantly higher levels of satisfaction with school (0.20 points higher on satisfaction scale), and higher levels of perception about child's engagement in school (0.60 points higher on perception/engagement scale). Table 11 shows these estimates.

Table 11. ITT Impacts on Caregiver Satisfaction, Child Development & Parent Engagement

	Child development scale	Scale for caregiver's satisfaction on teachers, staff and school structure	Caregiver perception of child engagement scale
Treatment	0.026	0.197**	0.595***
	(0.093)	(0.081)	(0.220)
Comparison Mean	9.634	2.364	8.341
N	1729	1055	1055

Clustered standard errors in parentheses. * $p<0.1$ ** $p<0.05$; *** $p<0.01$
Estimations control for district fixed effects, age at baseline, distance from school at baseline, and baseline EGMA score.

These findings combined with the statistically insignificant effects on school attendance and enrolment suggest that higher quality of education in Impact Network schools likely contributed to improvements in literacy. One additional factor that may have contributed to higher quality of education is the more consistent implementation of the Catch-Up program in the Impact Network schools. We present more details on this mechanism in the discussion describing school-level impacts.

iSchool tablets

Qualitative Data on Satisfaction with School and Education Quality

Respondents across the board (including students, teachers, and parents) voiced positive perceptions of education quality in Impact Network schools. Students generally cited the quality of Impact Network teachers, adding that teachers stopped to explain materials and checked student understanding before moving on. Teachers generally focused their remarks about education quality on the design of the Impact Network program and quality of the school environment. Teachers mentioned that Impact Network had well designed lessons and that the school environment was conducive to learning. For instance, one teacher stated, "*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 made a remark about how Impact Network's teacher training and rules created a conducive learning environment for children.

> *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.*

Many teachers interviewed believed that Impact Network students learned fast and performed well, especially in comparison to students in other schools.

Most parents voiced positive perceptions of Impact Network schools and education quality. When asked why they had chosen to send their children to Impact Network schools, many parents stated that Impact Network schools were of higher quality and more effective at teaching students to read than other local schools. For example, one parent explained that she chose Impact Network schools because of how quickly students learned to read there. "*Why we chose to bring our children to this school is because we saw that at impact schools children are able to read just after a short period of time."* Further, another parent believed that the quality of education at Impact Network schools was better than that of other local schools. "*For me I think I can compare them on the part of the education that they get. I think the children here at Impact Network schools are well ahead of the other schools in terms of how the children learn."* This view was widespread among parents interviewed for this study.

School environment/infrastructure (including cleanliness, safety, and school climate).

Despite positive perceptions of education quality and learning, Impact Network school environments present challenges for safe learning. PTAs reported good collaboration with Impact Network to maintain schools, and students themselves helped keep schools clean: *"We are asked that when we are done sweeping, we enter the classroom and those who haven't swept, they can't enter not until they sweep"* (Katete student). Parents, too, came to school occasionally to help with repairs. Despite students and parents' efforts, however, significant challenges with respect to the school environment persisted. For one, some students did not feel safe at school or traveling to and from school. A Katete student said, *"when you have an argument with someone you can be afraid to come to school because you can be beaten up by them"* while a Sinda student shared, *"We are scared that there might be thieves lurking or many other things you might find in the road."* Some students traveled long distances to school, and the remote rural setting also proved difficult for some teachers who had trouble acclimating to their new environment and reported clashes between the school and the local community.

The more widespread concern about the school environment pertained to infrastructure: students and teachers alike reported that there is not enough space in classrooms and students were squeezed together on benches. A teacher in Sinda said,

> *"One [challenge] is the state in which the classes are because there are potholes there, so the children sometimes hurt themselves. Another thing is lack of sitting benches, it is*

first come first serve basis and so some would fight for the seating and the teacher needs to be close by so that the fights are controlled."

Other common infrastructure-related complaints pertained to toilets (not enough latrines, no separate latrines for teachers and learners); the lack of a functioning borehole; leaky roofs; and unclean classrooms.

Parents' Perceptions of School Engagement

Most parents stated that Impact Network had made them more aware of the importance of supporting their children's education. Parents described supporting their children's education by encouraging their children to go to school and buying uniforms and school materials (while also noting that they did not pay fees for children under the Impact Network program). Parents also reported reviewing report cards and checking their students' notebooks. This statement from a parent from Kachewele summarizes how the Impact Network program raised awareness about the importance of parental involvement and support of children's education.

> *We did not know that a child needs to be helped where the teacher cannot help. Then we were just being dormant if the child leaves their books, then that's it. But now even us parents we were given pencils and books so that we can be able to take part and help our children so it has made us parents to know that we need to be close to our children and help them because in the past we did not pay attention.*

Many parents described using report cards to have discussions with their child about school performance. For instance, a parent from Petauke explained that report cards were useful because they included data about student attendance and performance, which helped him encourage his children to pay attention in class.

> *[The report] is very helpful to us, because in that report we get to know how many times the child has been absent from school, it could be twice or three times in the whole term. They say those days the child missed, this and that lesson has passed them. We tell the child see this and that passed you and that is why you have been found on this number position in class so you need to pay attention in class.*

Some parents also reported communicating with teachers about student attendance and student performance, while a few also mentioned directly helping children with homework. For example, a parent from Chamboo school in Katete noted, "*I look through my child's books when they come from school we check their books so that we can check their work and help them with corrections, in any subject like math, that's how I assist."* Further, a parent from Lenda noted, "*When I get the report forms I am able to see why my child is performing badly and I can talk to the teachers about it.*"

PTA members reported helping schools with construction of teacher housing and toilets, as well as overseeing meal preparation. In some schools, PTAs seemed to play an important role in mediating between the school and the community, including in follow ups about attendance and in child protection cases (we present more details in the section on child protection). A parent from Chamboo school described the PTA's involvement with attendance follow ups in the following way: "*When letters are written to the parents of the children that don't attend school, when those parents and children come to the school to speak to the teachers the PTA is present during those talks.*"

While program staff noted that parent communication was a strength of the Impact Network program (with school support managers and Teacher Supervisors being tasked with facilitating relationships with parents) they also pointed to challenges in engaging with parents. Some teachers stated that parents did not value education very much, and therefore did not participate/get involved with their children's education enough. Program staff noted that some teachers believed that local communities were involved with witchcraft, which made them afraid to discipline children for fear of reprisals. An Impact Network program staff member stated,

> *I think the other challenge is… these communities teachers are living in the communities near the schools so they somehow know people who are into witchcraft so once they learn that this child is coming from a mother or a father or a parent who is into witchcraft it's very hard for the teachers to discipline those leaners because they are scared of the witchcraft issues.*

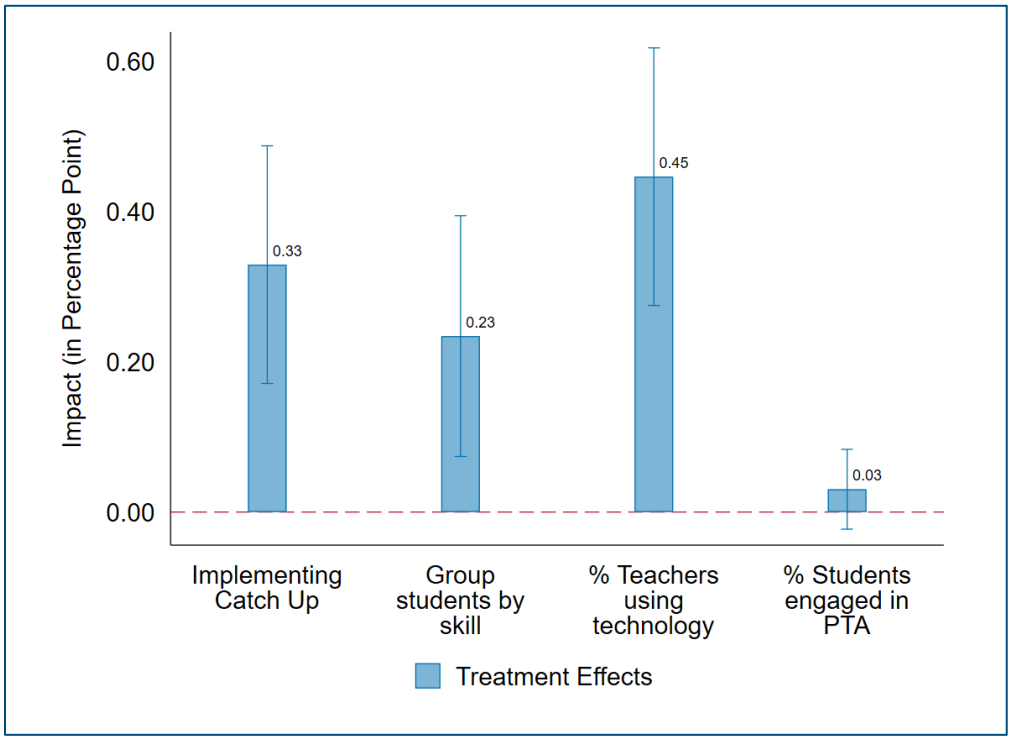
This view was only mentioned by a few respondents, suggesting either that this challenge is not widespread or that respondents felt uncomfortable bringing up this topic.

School Surveys on Teaching Quality & School Infrastructure

School surveys with 62 schools indicated substantial differences in teacher training, school infrastructure, and implementation of the Catch-Up program. Almost all of Impact Network schools (97%) were implementing Catch-Up by endline, compared with 64% of Control schools. Aligned with Catch-Up practices, Impact Network schools were more likely to group students by skill level. This is important because Banerjee et al. (2016) found strong evidence that implementing the Teaching at the Right level methodology (which guides the Catch-Up program) led to significant learning gains in India, which was confirmed in a meta-analysis of the program showing large and consistent treatment effects when they are implemented with high fidelity (Angrist & Meager, 2023). This suggests that high quality implementation of the Catch-Up program may have resulted in larger impacts of the eSchool 360 program. Additionally, Impact Network teachers were 45 percentage-point more likely to know and use

technology compared with control school teachers. We did, however, not find any differences in the likelihood of PTA engagement, however. Figure 11 shows these findings.

Figure 11. Differences in School-Level Teaching Practices



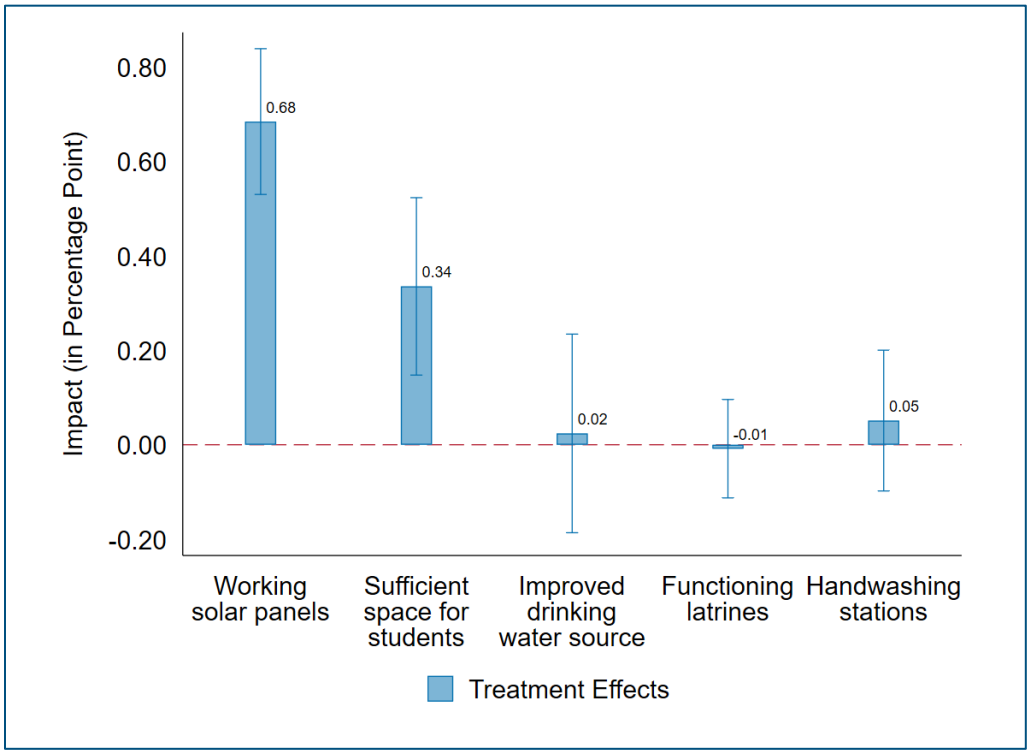
Teachers, students, and Impact Network staff confirmed that Catch-Up classes in literacy and numeracy took place twice a week during 40-minute sessions. SSOs assessed students each term to create groups for Catch-Up classes based on learning ability rather than age or grade. Impact Network staff said that Catch Up classes were well aligned to the Impact Network curriculum in terms of their use of play-based, interactive learning approaches. According to one Impact Network respondent, *"CatchUp reinforces [active learning strategies]. It's nice that it comes from the MoE, [it's] endorsed by the MoE. [It's] not some crazy thing [Impact Network] are just doing…play-based learning is not just a weird Impact Network thing."* Most teachers agreed that Catch-Up classes complement the Impact Network approach, and that there is no tension between the two.

Catch-Up classes appear to have run smoothly, except for a minority of students who felt the content was too easy (a Sinda teacher said, *"Some learners find it boring because they already know how to read, so to them it is boring"*) and some Impact Network staff mentioned that they had to abbreviate other lessons to make time for Catch-Up classes. Physical space for Catch-Up classes also presented a challenge for some teachers, as it does for regular classes. Teachers made comments like, *"They need to write on the ground, so the classroom becomes small. And*

again, challenging, because you have desks in there, you have the space itself it's not enough, because every learner needs to solve those challenges on their own" (Sinda teacher) and *"I think the biggest challenge is that we normally share the room. You'll find that the other teacher is there, and another teacher is there using the same room, so it's quite a challenge"* (Katete teacher).

Other school-level data also indicate that Impact Network schools delivered higher quality education than control schools. Enumerator observations in treatment and control schools revealed that treatment schools were significantly more likely to have working solar panels, and sufficient and well-lit spaces for students, though we did not find statistically significant differences in other factors like availability of drinking water, latrines, and handwashing stations. Availability of latrines was nearly universal, with 94% of both treatment and control schools having functioning latrines for all children. On the other hand, only 17% of treatment and 12% of control schools had handwashing stations with soap and water, 45% of treatment and 42% of control schools had improved drinking water sources on ground (without the need to walk longer distances); and 52% of treatment and 18% of control schools had sufficient spaces for students. We present these results in Figure 12.

Figure 12. Differences in School Infrastructure



Secondary Outcomes – Aspirations & Expenditures

Treatment parents spent significantly less (50 Kwacha or 2 USD) on schooling expenses, compared with Control parents, consistent with the Impact Network model of free education.[16]

We found no evidence for effects of the eSchool 360 model on parental aspirations. Specifically, we did not find statistically significant effects on caregiver's aspirations about child's school (whether child would study beyond senior secondary level), preferred age at marriage, or expected bride prices for their daughters. Table 12 shows these findings.

Table 12. ITT Impacts on Secondary Outcomes

	Age at which caregiver wishes child would marry	Caregiver wants child to achieve college and higher education level	Amount caregiver hopes to receive in bride price (female students)	Amount spent on school expenditures, last two years
Treatment	-0.033	0.054	-990.943	-50.571***
	(0.295)	(0.035)	(650.227)	(15.735)
Comparison Mean	25.611	0.438	5018.299	179.484
N	1680	1623	702	1505

Clustered standard errors in parentheses. * $p<0.1$ ** $p<0.05$; *** $p<0.01$
Estimations control for district fixed effects, age at baseline, distance from school at baseline, and baseline EGMA score.

6.6. The Influence of COVID-19

The COVID-19 pandemic likely contributed to the smaller impacts during the endline evaluation than during the midline evaluation. This section presents some details on school closures associated with COVID-19 and the way these school closures may have influenced the results. The Zambian government closed learning institutions for the first time in March 2020 after which students returned to school in September 2020. However, the government continued directing school closures frequently after this time. As a result, schools were not able to open for longer periods of time until mid-2021, slightly more than one year before the start of our endline survey.

Survey data did not suggest significant differences by treatment status in the impact of COVID-19 on school enrolment. While marginally higher proportion of parents in the control group indicated that the pandemic impacted their decision to enrol their child at school (32% in control versus 28% in treatment group), the difference was not statistically significant. Further,

[16] The Zambian government did, however, introduce free primary education for all schools in January 2022.

when probed about how the pandemic impacted their child's schooling, most parents (68%) in both treatment and control groups reported that their child resumed school after a temporary break; and there was no difference in the proportion of parents who reported that they delayed or pulled enrolment because of the pandemic (see Table 13).[17]

While most parents reported that their child resumed school after a temporary break, quantitative results suggest that students with lower assessment scores during the baseline were statistically significantly less likely to continue their education after the reopening of schools than students with higher assessment scores. It is likely that the COVID-19 pandemic contributed to this difference. This result also may explain the larger intention-to-treat effects for students who had above-median assessment scores during the baseline.

Table 13. COVID-19 Impacts

	Decision to enroll child at school affected by COVID-19	Delayed or stopped school because of COVID-19
Treatment	-0.038	-0.016
	(0.026)	(0.016)
Comparison Mean	0.318	0.100
N	1724	1724

Clustered standard errors in parentheses. * p<0.1 ** p<0.05; *** p<0.01

Estimations control for district fixed effects, age at baseline, distance from school at baseline, and baseline EGMA score.

When schools reopened, Impact Network put in place several measures to keep students safe and mitigate learning loss. Larger classes were split in half, either alternating days or having shortened class periods all within the same day. Impact Network created and distributed cloth masks, delivered sanitizer to classrooms for both hands and equipment, and increased emphasis on health and safety at schools. The Operations Coordinator visited schools to ensure teachers and students were wearing masks, washing hands regularly, and trying to maintain social distance. Lastly, the Catch-Up program was modified by Impact Network schools to attempt to better address learning loss during COVID-related school closures.

[17] While social distancing rules for the treatment and the control schools were the same during the COVID-19 pandemic, it is possible that the implementation of the social distancing rules was different in treatment and control schools. We do not find evidence for such differences in school-level survey data, but it remains difficult to completely rule out such differences because of potential recall bias.

While schools were closed due to COVID-19, Impact Network teachers were tasked with visiting students regularly and assigning "homework packs" to help children continue to learn at home. Teachers sometimes met with small groups of students together and counseled them on how to protect themselves from COVID-19. Teachers reported having received targets for numbers of children to see each day: *"…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"* (Petauke teacher). Students and parents confirmed that these visits took place and were useful, with one parent from Katete saying, *"[Teachers] would go through homes and give the children homework to do and then they would come again the following week to check upon the work that was given to them the previous week."* In some cases, village headmen even helped distribute homework assignments to children.

In response to the COVID-19 pandemic, Impact Network schools also gave masks to students, encouraged handwashing (including providing wash buckets and soap where needed), and tried to implement social distancing. Teachers agreed that limited space and seating arrangements made social distancing challenging: *"The most challenging is social distance since we do not have enough furniture, so we do not observe social distance"* (Petauke teacher). Beyond the space constraints, teachers mentioned other resistance to COVID-19 precautions from both students and parents. According to a teacher from Petauke, *"These who come in the afternoon don't wear masks, and they don't even observe…. even parents themselves they don't observe social distance or if you tell them we are living in the era of corona, they will tell you there is no corona."* Other teachers echoed these sentiments, maintaining that adherence to COVID-19 protocols was inconsistent despite their best efforts.

Impact Network also worked with local radio stations (Mphangwe radio, Breeze FM, and Radio Explorer) to broadcast lessons during school closures. While some students and parents felt the radio lessons were useful (a Petauke parent said, *"Because even when schools were closed, the children used to learn from the radios [so] they were not far behind"*), not all learners listened to the radio or had access to radios. Respondents also noted that poor signal strength also interfered with radio lessons. The radio education program had some similarities with the program implemented by Rising Academies in Sierra Leone, which did not show positive impacts on learning outcomes (Crawfurd et al., 2023). Both treatment and control group children had access to radio education.

While school closures undoubtedly interfered with student learning, qualitative data indicated that teachers and Impact Network staff were most concerned about students who didn't re-enroll when schools reopened and felt that most learners largely managed to catch up. Respondents reported that students got limited academic help from their family members during school closures (some teachers complained that parents did not encourage or remind

students to complete their assignments, and a Sinda PTA member said, *"It was difficult for the children to learn on their own because you find that some parents don't even know how to read"*), but that visits from teachers and homework packs helped students continue learning at home. To this end, one parent from Katete said, *"For me I think they were able to catch up because they never stopped learning even during that time."* A PTA member from Petauke cautioned, however, that learning at home was not the same as learning at school: *"I don't think they used to learn 100% [during COVID closures] they were a bit disturbed as compared to when they were at school."* Impact Network staff had the greatest concerns were about older students ("*The kids in grade 6 are most impacted since they have the least time to catch up on what they missed. Younger grades still have more time for the program to continue working on them")* and of course those who did not re-enroll when schools opened.

In addition to the learning implications of COVID-19 school closures, the pandemic had other negative repercussions for students and their families. Some students reported being sad to be home from school and feeling like they were missing out. A student from Katete told us, "*I felt bad because I felt we were missing out on many lessons, and we wouldn't be able to learn them"* while one from Petauke said, *"We were not feeling good because when schools were closed, we began to forget what we had learned."* Parents also shared that COVID-19 compromised their livelihoods, such as this one from Sinda: *"During COVID times it was very difficult to find money because the jobs that we used to do to find money had stopped for a while."* Another parent from Sinda shared, *"For me I am a garden farmer, so I used to sell tomatoes and so during the COVID times it was difficult to go to town and sell tomatoes, so I lost some income in that way."* Lastly, some teachers felt that while most students returned when schools reopened, some of them demonstrated less interest and commitment to their education. A teacher from Kanchwele said the following: "*I think some they lost interest so when we reopened, they stopped. And then others, maybe the reason why they are absent is because the parents they involve them in chores."* Other teachers made similar comments about learners losing momentum for their schooling due to prolonged school closures.

6.7. Comparison with Government Schools

To examine differences in learning outcomes between Impact Network and government schools, we also collected assessment data among students in Grade 5 and 6 in government schools. The comparison with assessment data from government schools does not allow for establishing causality. However, the data do allow for a descriptive comparison between learning outcomes in Impact Network and government schools, which generates lessons for the potential scale-up of the program to government schools. We present the results of the comparison with government schools in Figure 13.

The results show that learning outcomes are statistically significantly higher among students in Grade 5 in Impact Network schools than among students in Grade 5 in government schools. Assessment scores are higher for EGRA, EGMA, ZAT and oral vocabulary assessments. In addition, students in community schools in the control group tend to score higher than students in government schools. It is possible that the higher quality of education in Impact Network schools may have contributed to the higher assessment scores. However, it is also possible that students in government schools are less likely to drop out before Grade 5, in which case the higher test scores in Impact Network and other community schools may be a result of a comparison between higher performing students in community schools with students with average skills in government schools.

Figure 13. Comparison with Government Schools

6.8. Cost-Effectiveness

We estimated program costs following the implementer perspective, including expenditures incurred to deliver program activities. We considered three different cost estimates, based on

assumptions ranging from least to most conservative. In all three estimates, we included expenditures reported by treatment and control schools on teacher professional development, staff salaries, classroom improvement, other infrastructure needs, technology, and miscellaneous expenses. Additionally, for treatment schools, we added expenditures incurred by Impact Network on running the eSchool 360 program under this project. However, several expenditures incurred by Impact Network would also have been incurred by district educational boards for overseeing education and infrastructure in control schools. Therefore, we started by identifying line items that included these common expenditures (for example, books and stationery, teacher supervision, etc.). In Scenario 1 of the cost estimation, we excluded these potentially overlapping costs. In Scenario 2, we further excluded teacher salaries paid by Impact Network, to assess the program's cost-effectiveness when the government would scale-up the program (where salaries would be paid by the government). Finally, in Scenario 3, we followed a conservative approach and added up all costs incurred by Impact Network on running this project, and do not remove any expenditures that could potentially overlap with control schools. This last scenario is the most conservative and based on our observations the least likely scenario. However, we include this scenario to provide a complete picture of the potential cost-effectiveness of the program.

Table 14 presents average cost per student in 2022 USD, from both sources – amount spent by schools from government and community funds, and amount spent by Impact Network. As shown, control schools have higher expenses from government and community funds, while these expenditure buckets were partially or mostly borne by Impact Network in treatment schools.

Table 14. Average cost per student over 5.75 years

Average costs per student over 5.75 years	Treatment	Control
Amount spent by schools from government and community funds		
Teacher prof. development	0.5	2.1
Support staff salaries	2.1	2.4
Classroom improvement	0.6	7.4
Other infrastructure improvement	0.6	1.9
Technology for teaching	0.2	0.1
Other miscellaneous	1.2	2.4
Total school spending	**5.2**	**16.3**
Amount spent by Impact Network		
Variable project costs including small assets, supplies, equipment etc.	49.4	0
Teachers	20.8	0
All NGO and project staff	108	0

Costs reported in 2022 USD

Findings suggest that after four years of implementation, the cumulative average costs of the eSchool 360 model ranged from roughly $55 to $113 per student, depending on varying assumptions. When including costs incurred by Impact Network, we added all operating expenditures related to small assets, teachers and supervision, setting up the school infrastructure and solar panels, and expenses related to local NGO staffing, like Teacher Supervisors, operations officers and specialists. Because we estimate per capita school expenditure as school expenses divided by total students enrolled in the school, we use ToT estimates to compute the cost-effectiveness of the program. When estimating costs per capita, we divided each cost bucket by the target number of schools and students that particular cost bucket would apply to. For example, Impact Network's data manager and data monitoring officer monitor operations across 50 schools.[18] Therefore, we divided those costs among student populations from 50 schools (instead of 33 schools as under this project).

Our estimates suggest that for every $100, the program generates 0.35 SD (under the most conservative scenario) to 1.40 SD (under the least conservative scenario) gains in literacy, by endline. These estimates are based on five years of program expenditure, with all costs

[18] This number is based on information received from Impact Network, on how it would theoretically divide costs across schools (the denominators for different variable costs).

reflected in 2022 USD, when the endline assessments were conducted. Figure 14 shows additional standard deviation gains in EGRA at endline, per $100 spent on the program.

For every $100, the program generates 0.11 SD (under the most conservative scenario) to 0.25 SD (under the least conservative scenario) gains in numeracy, by endline. As expected, the confidence intervals for EGMA outcomes suggest that the numeracy gains are not statistically different from 0, in line with the ToT findings above. Figure 15 shows additional SD gains in EGMA at endline, per $100 spent on the program.

Figure 14. Average SD Increase in EGRA per 100 USD Spent by Endline

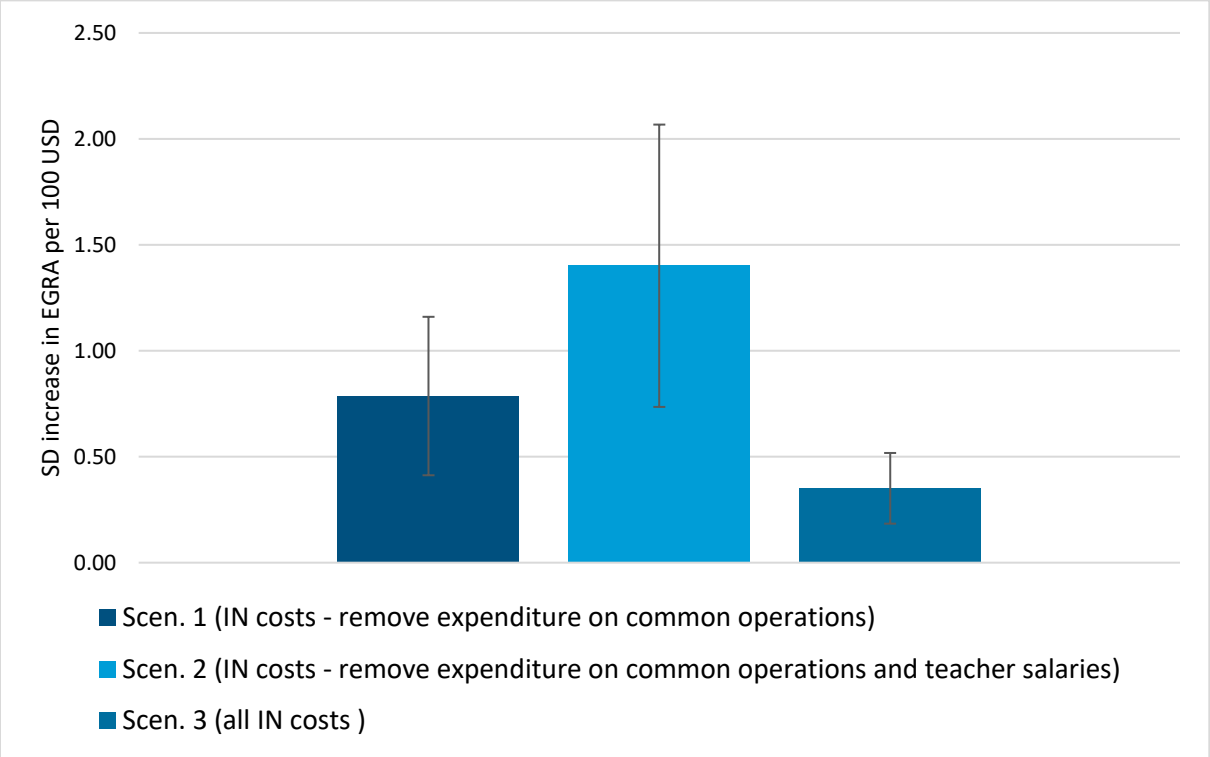
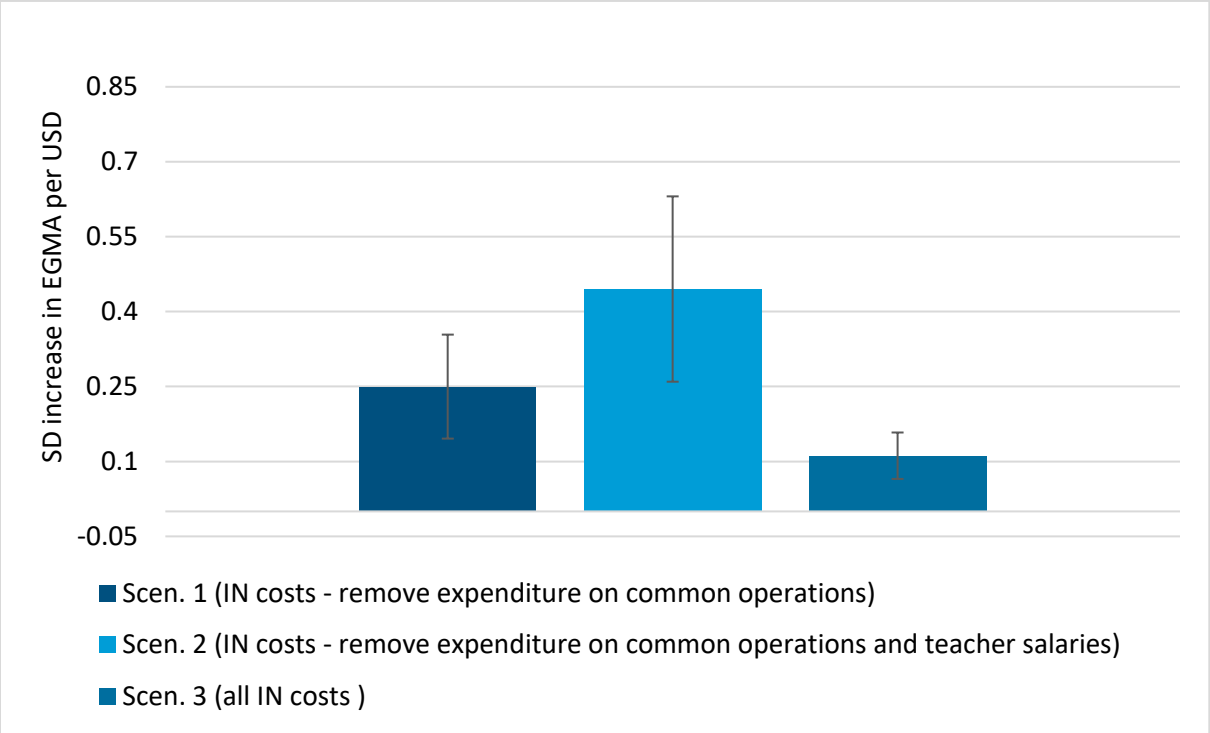


Figure 15. Average SD Increase in EGMA per 100 USD Spent by Endline



Scaling back costs to midline where Treatment schools were 14 months into program implementation, we find substantially larger cost-effectiveness of the program, owing to both larger impacts and lower costs at midline. Estimates shown in Figures 16 and 17 indicate that every $100 spent on the program resulted in an additional 1.38 to 3.30 SD gain in EGRA and 0.73 to 1.62 SD gain in EGMA. All estimates are statistically significant at 95% level of confidence.

Figure 16. Average SD Increase in EGRA per 100 USD Spent by Midline

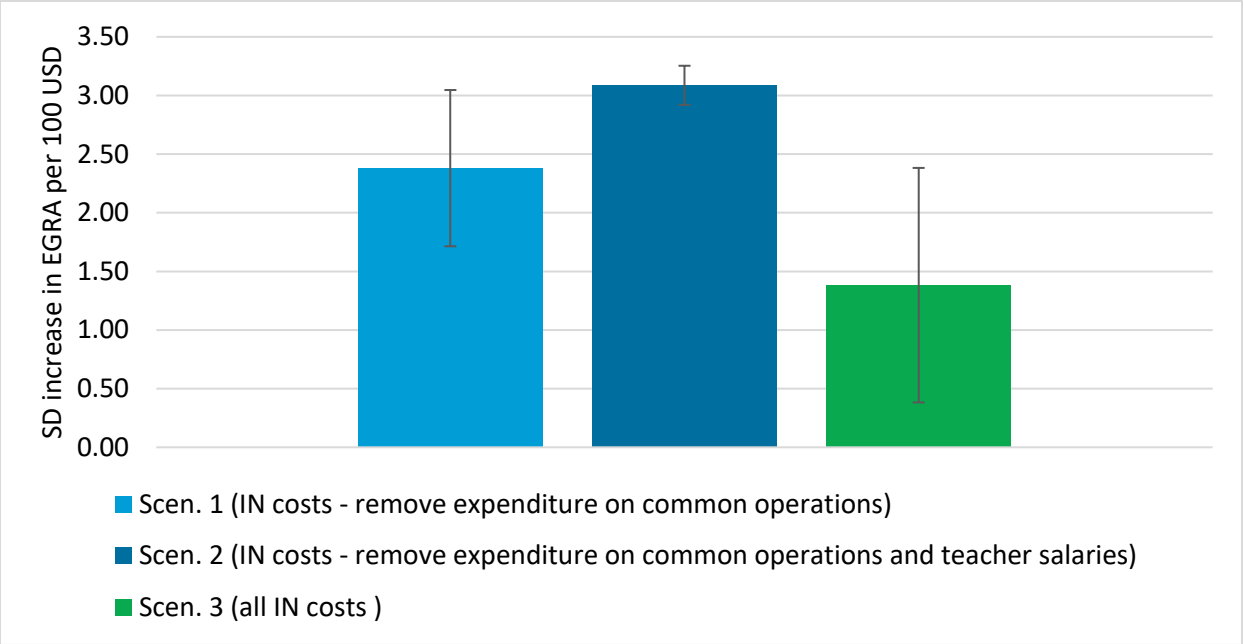
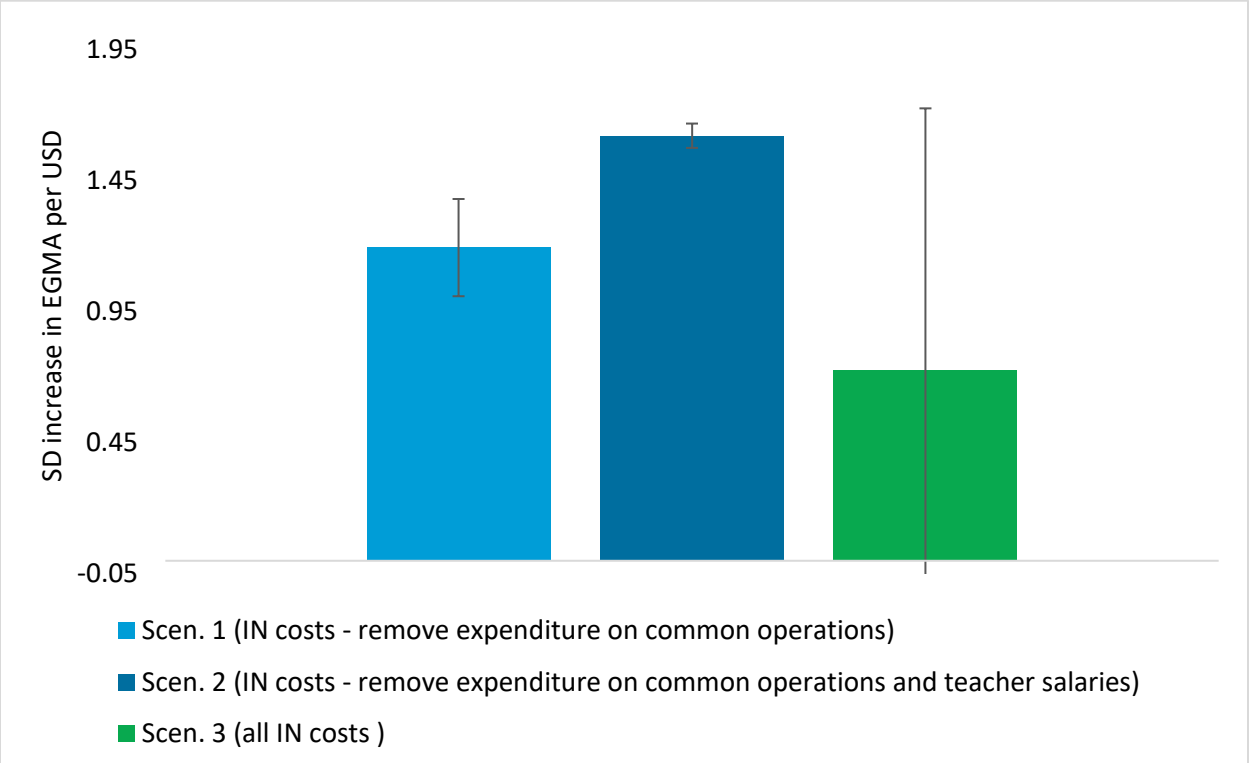


Figure 17. Average SD Increase in EGMA per 100 USD Spent by Midline



6.9. Process Evaluation Results

The IN program featured design elements intended to promote student enrollment and learning outcomes. These elements included:

- Low-cost of attendance at program schools
- Using an e-learning pedagogy with curriculum in line with the MoE's standards
- Using tablets and projectors to deliver content
- Providing electricity to schools via solar panels
- Providing classroom materials/supplies
- Providing teacher training and professional development through coaching
- Creating community ownership (through PTA engagement)

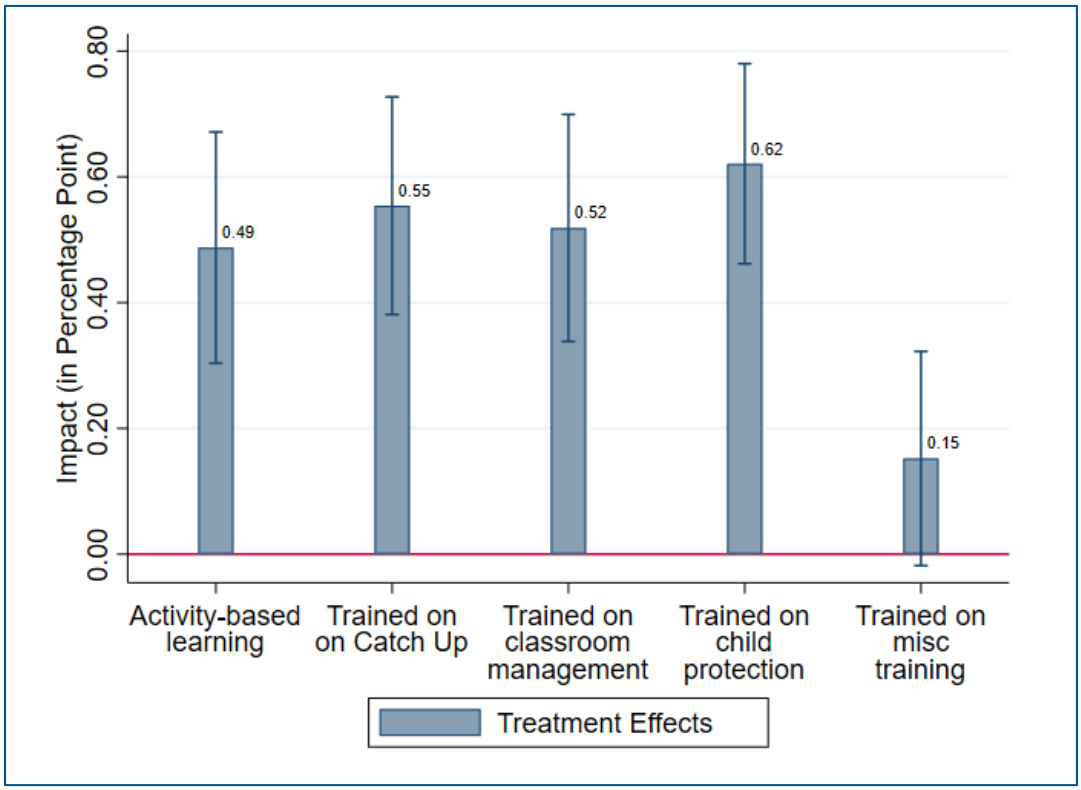
This section discusses the extent to which the IN program was implemented as designed, and in cases where it was not, explores the challenges that prevented the intended implementation.

Fidelity of Implementation

Findings from the school survey suggest that Impact Network teachers were substantially more likely to receive trainings on activity-based learning, Catch-Up, classroom management, and other miscellaneous topics, compared with control school teachers. An emphasis on ongoing and continued teacher professional development is an integral component of the eSchool 360 program, as shown in the school survey data. Figure 18 shows these findings. These responses relate to teacher trainings for teachers teaching grade 5 in 2022 – where the eSchool 360 program was implemented.

Further, 85% of treatment schools reported that their grade 5 teachers sometimes or often use technology in classrooms, compared with just 7% in the control group.

Figure 18. Differences in Teacher Professional Development



Endline KIIs, FGDs, and direct observations revealed that the IN program was implemented with a high level of fidelity overall. These findings can be separated into implementation of interventions associated with training and professional development for teachers and IN staff, classroom instruction, and child protection. The subsections below explore each of these implementation areas.

Teacher training

Impact Network's teacher training model was implemented as designed, with approximately three trainings a term. This includes a 5-day immersive teacher conference in the beginning of the year, two teacher training days in the middle of the term, and a wrap-up session at the end. The 5-day conference covered lesson planning and pedagogy, operational topics such as use and maintenance of solar panels and tablets, and human resources. The teacher training days covered teaching and learning as well as challenges and ways to address them. The wrap up session included a data collection exercise where teachers brought exam scores and attendance records to be entered into the monitoring data for the program. According to program staff, training topics were informed by an assessment of teacher's needs from classroom observations and solicited teacher input. For example, Impact Network staff designed a child

protection training to address the increase in child protection needs since schools reopened after pandemic closures. An Impact Network program staff member described the iterative process of selecting training topics in the following way:

> *We hold a number of [teacher] trainings, mostly especially from the academics part, we look at the lesson observations that are being done in schools and then once we look at those we evaluate to say which areas are the teachers really struggling with so if we discover that a teacher is really struggling with… classroom management then we bring in a training piece on how they would manage their classrooms… So I think we've done trainings on classroom management, on child protection and safeguarding, on some challenges they face with different subjects, and how they would address them during teaching.*

Teachers consistently reported participating in these three trainings, indicating that the trainings were implemented in a consistent fashion. In addition, Impact Network staff and teachers noted that there were refresher courses for new teachers in the organization. Teachers generally agreed that trainings were sufficient and effective in preparing them to teach with technology in the classroom. They particularly liked that the training was hands-on and informed by the challenges faced by teachers in the classroom. For instance, one teacher described the usefulness of the training in the following way:

> *They told us to bring some challenging topics in computer studies so we are told to discuss in groups on how best we can teach those lessons. So, we had to get the projector, and then we project. Maybe some of them they chose technology studies which was on text alignment and then we could [set up the] projector [and] see how best they can teach the lesson. So due to that I feel like I was helped such that in technology I was told how to go about those lessons.*

Teacher Supervisors and other Impact Network program staff stated that the training was effective, with many citing that they observed teachers implementing practices learned from the training in the classroom. For example, an Impact Network key informant explained, "*I am very much confident [in the training] because I have seen that most of them they apply whatever information that they have learnt from the training that we have delivered in their respective classrooms.*" Further, a program staff member who started as a teacher noted that the trainings were also useful for teachers' professional development more broadly. He explained, "*Teacher trainings… have helped our teachers. Let me not say our teaching but also us develop our career. And so yeah, it helps us develop our career and also do our responsibilities effectively."*

However, some respondents suggested that the training may have been too complex for some teachers. In particular, a few teachers expressed that the training was too long and covered too much content. For instance, a teacher stated,

> *There is too much work [in the training]. When we go for training at times the time that we have been there we take a lot of time because if it can be done for a short time for example it started at 8 hours or we end at 12 hours it's enough but most of the times we take long so at least the content must be less, at least less content so that someone is able to grasp or understand the content itself.*

Similarly, some Impact Network staff reported that teachers were not able to follow what they learned in trainings. A staff member declared, "*This time around we do not have major challenges [with the training], except it just gets to be frustrating when you train the teacher on something and then you find them doing something contrarily."* In addition, another staff member noted that while most teachers were able to implement what they learned in trainings, a few didn't due to negligence or laziness. "*So, from the observations that we have been doing we have seen that there were some of the teachers who could not follow what was delivered during the presentations and trainings but most of the teachers do follow. Some of the reason why some do not do that is maybe negligence or lack of …they don't …they are just lazy.*" Nonetheless, this perception does not appear to be widespread among Impact Network staff and teachers.

Program staff members noted that coordinating transportation for teacher trainings was logistically challenging. First, some teachers could not attend trainings because of transportation challenges of rural areas exacerbated by the rainy season. Second, some teachers complained that they did not receive travel reimbursements to attend trainings – this issue seems to be related to the reorganization of trainings after the COVID-19 pandemic. Impact Network staff used to conduct trainings in a central location (Joel School), but during the pandemic they organized trainings at different schools. Teachers who lived less than 15 kilometers away from the training site were no longer reimbursed for transport, which made it difficult for them to attend.[19] Lastly, in some cases, power interruptions or lack of water reportedly disrupted trainings.

Teachers, on the other hand, mentioned that the training would benefit from having additional handouts and materials to take home, providing more snacks, and giving incentives for

[19] Starting in Term 2 of 2023, Impact Network started reimbursing everyone for transport, regardless of travel distance.

participation, especially in light of the transportation challenges mentioned above. One teacher suggested that the training could be improved by, "*Providing handouts for some of the things that they project during the training maybe they just give you a hard copy so that even when you go back you are able to refresh on what you learn*." Another teacher noted that additional snacks and incentives could help motivate teachers, especially in view of the transportation challenges noted above.

> *What we are usually complaining during our teacher training is first the transport refund some teachers are coming from very far, but they're just told to come and attend the training and then when going they don't refund their teachers so that thing I think they need to improve on it so that they can be refunding them. Then the other one is food is not enough during teacher training, they are not giving us enough food during teacher training so these are the things that I can say they need to improve on in the next teacher training.*

In summary, teacher trainings seem to be largely implemented as intended, though respondents noted a few challenges related to transportation and teacher's ability to absorb training materials.

Impact Network staff training

The Impact Network program also included staff and management trainings, which seem to be largely implemented as intended. Program staff described holding a staff/management training once a year at the end of term, as well as management trainings throughout the year (up to six trainings/year). These trainings brought together Teacher Supervisors, Operations Coordinators, and all the management staff aside from the teachers. Topics included human resource topics such as hiring processes and practices, equity, communication among members of staff, and vehicle safety trainings, as well as a discussion of data findings related to enrollment, gender breakdown, operational data (infrastructure, vehicle and fuel usage, consumption survey), and feedback from teachers. Further, staff trainings included guidance on coaching and emerging issues at schools.

Similar to perceptions on teacher training, program staff agreed that staff trainings were engaging and responsive to needs of staff members and teachers. Respondents also mentioned that staff/management trainings allowed the usually dispersed program staff to get to know each other and come together to identify common challenges and brainstorm solutions. Further, some program staff believed that management trainings were particularly impactful because of the trickle-down effect on teachers and schools. As one staff member put it,

If we train our managers on how, maybe, to conduct the coaching session, these will give them the skills on how best they can help a teacher on certain areas where they have faced the challenges, and the teachers really benefit from the trainings and coaching that they receive. So, we start from the training where we train the coaches who are Teacher Supervisors, then in turn the Teacher Supervisors now are able to deliver this to the teachers. So it has a very significant impact on our teachers as well as our Teachers Supervisors, as well as our managers.

For these respondents, management training was effective in improving the quality of coaching received by teachers.

Nevertheless, respondents also pointed to challenges in staff trainings, including the need to build the capacity of program staff (who are generally young and in their first job) to effectively communicate in the workplace and design more interactive workshops/trainings. One respondent suggested developing a standardized template for training with areas for updates/changes as needed. In addition, a few staff members noted that coordination and logistics were difficult for trainings, since participants traveled long distances and faced transportation challenges in the rainy season.

Classroom Observations and Coaching

Impact Network's teacher training also included continuous improvement for teachers through classroom observations and coaching. Teachers were assigned to a Teacher Supervisor, who observed them teaching a lesson on a weekly basis. After the classroom observation, the Teacher Supervisor would conduct a short coaching session to provide feedback on teachers' quality of teaching, including strengths and challenges. Occasionally, education program officers or SSOs accompanied Teacher Supervisors on co-observations and co-coaching sessions and gave guidance on how to improve and deal with challenges encountered in observation and coaching.

According to respondents, classroom observations and coaching sessions were largely implemented as planned. Teachers and program staff across schools and districts reported that Teacher Supervisors conducted classroom observations weekly, in class sessions ranging from 30-50 minutes. Teacher Supervisors reported leading 15-30 min coaching sessions after observing teachers' classrooms. In addition, respondents indicated that Teacher Supervisors kept track of teachers' improvement and progress through a coaching log.

Respondents cited several strengths of the classroom observations and subsequent coaching sessions. Program staff agreed that observations and coaching sessions were particularly effective in building the capacity of teachers, especially because of the one-on-one format and

immediate feedback on teachers' performance. For instance, an Impact Network staff member stated,

> *Sometimes… the teachers just do teacher training, some of them might not like put them into practice. But when they observe through observation or peer teaching, they get like immediate feedback and… they have really improved their skills. And also… one to one coaching [is] one of the good ways of providing feedback, because some of the teachers feel shy to express themselves in a larger group, so during coaching the teachers are able to come out freely and the Teacher Supervisor provides this help to these teachers.*

In addition, program staff noted that classroom observations and coaching incorporated accountability into the program. For example, Teacher Supervisors' coaching visits were sporadic, and teachers did not know when they would be observed. This meant that teachers had more incentives to attend school daily and to be prepared for class.

Teachers generally agreed that observations and coaching sessions were helpful. One teacher described the coaching process in the following way:

> *We discuss on the lesson learning outcomes for that particular topic, and then apart from learning outcomes we go on the… achievable strengths during the lesson that what I as a teacher achieved during the lesson. Then after that we go on to, to the challenges weaknesses I face during the lesson after discussing those challenges we go in how to improve those challenges. If I have a challenge in time management then the Teacher Supervisor will be there advising me on how best I can improve on time management.*

The most cited issues discussed in coaching sessions included learner-centered methodologies, time management, and language of instruction (more specifically, code-switching between English and Nyanja during lessons). For instance, a teacher incorporated feedback from coaching sessions to make her lessons more interactive. As the teacher stated,

> *I remember last time I was teaching about traditional dances in Social Studies… and I was being observed so I was advised to say like if this lesson you had used a role play you involve the learners like maybe to come in front and perform like dances some traditional dances, it can help the learners very much because if they practice that it will be easy for them to remember what they were doing. So at some point I used that… this time we were looking at traditional music so I had to involve the learners so that we could come in front and sing some of the traditional songs they know.*

Most Teacher Supervisors and program staff agreed that teachers were confident to apply coaching feedback to their teaching practices. A Teacher Supervisor noted, "*Yes, [teachers] are very confident [on the implementation of feedback from coaching sessions]… most of them they*

are really appreciating the training that we give them. Maybe they are struggling in one area then they would like to have more information on how best they can do it." However, respondents also cited some challenges to the implementation of the classroom observation/coaching model. Some program staff noted that teachers would sometimes take coaching feedback negatively. As one Teacher Supervisor put it, "*On coaching the challenge which I normally face, is some teachers get it from a negative side, whereby they think that maybe I am getting wrong things from them.*" According to another program staff member, this is because Zambian teachers are not used to the self-reflective model, given that their training promotes a more "checklist" approach to continuous improvement (in other words, teachers want to know what they did wrong and how to fix it in a direct manner).

Further, several program staff mentioned logistical challenges that affected classroom observations and coaching, most of which were related to the general difficulties of working in rural Zambia. Teaching supervisors experienced transportation challenges, especially during the rainy season. They had to cover several schools and travel long distances between the office and the sites they visited. Travel was made more difficult during the rainy season, when bridges are not usable and Teacher Supervisors often have to travel longer routes. Further, if motorbikes broke down, Teacher Supervisors often had no one to help – a situation that was especially challenging for female Teacher Supervisors. In addition, Teacher Supervisors faced challenges related to communication, since many of the Impact Network schools were in areas with no network.

Although teachers were generally satisfied with the coaching model, they did mention suggestions and improvements. Some teachers felt that weekly classroom observations/coaching sessions were excessive and suggested making the observations biweekly to "*give some breathing space*" to teachers. Other suggestions included having Teacher Supervisors jump in and demonstrate teaching techniques while teachers observed. This would require that teachers have a copy of the coaching sheets that teaching supervisors use to monitor progress, so that they can track their own weaknesses and progress.

Classroom Instruction – Technology

Qualitative data revealed that the tablets and projectors distributed by the IN program were used in classroom instruction to near unanimously positive reception. Parents in FGDs described how Impact Network schools do a good job of helping the children learn to use computers, and both parents and teachers agreed that learning to use computers is important as a life skill that could potentially lead to generating income in the future. Both students and teachers confirmed that the projector helped communicate the lesson content, showing it enlarged on the classroom wall for all to see. Teachers specifically praised how tablets

promoted student motivation and understanding, helping them grasp difficult topics through audio-visual aides. According to the Impact Network Director of Academics, the images provided during the tablet lessons enable children to become familiar with words and images of things that are outside their local communities: "*Tablets bring context to some things that are 'other'."* Impact Network staff mentioned some additional benefits of tablets, including that tablets are more durable than textbooks and can easily be updated with the most up-to-date content. The tablets were also reported as encouraging attendance for more remote students. Respondents also confirmed that Impact Network schools consistently had the prescribed number of tablets—seven pupil tablets and two teacher tablets.

In FGDs, parents commonly expressed familiarity with the tablets and projectors used to educate their children. Students reported that the devices were used regularly, with most saying that the tablets were used to deliver lessons on 2 to 3 days of the week, depending on the lesson content. However, according to the Impact Network Director of Academics, using the tablets less than 3 out of 5 days per week means that teachers were not following the prescribed pedagogy and curriculum. Students described how when the tablets are not charged for lessons, their teacher would write on the chalkboard, or they would take turns with the available tablets.

The Impact Network Director of Academics described how the Mwabu iSchool content loaded onto the tablets, "…though advertised as working with national syllabus, it doesn't follow in reality," and is supplemented by other content uploaded by Impact Network designed to follow the MoE's curriculum: "Local languages can be lacking in particular topics or particular languages. English lessons are not for ESL and are instead for students who are native and learning writing. When Impact started, it was 90% Mwabu, and now closer to 50-60%." The rest of lessons are supplemented with Jolly Phonics and other sources, according to the Impact Network Executive Director. Teachers praised that the content was available at various levels, facilitating group learning for students who learn at different paces to be grouped according to their ability and receive appropriate content.

Impact Network staff described how they work with supervisors and teachers on site visits to make sure that the technology was working, as well as to perform regular maintenance checks. These checks serve to minimize downtime on repairs. According to an operations manager in Katete, *"In terms of reliability they are all reliable, because what usually happens, an Op's coordinator when he goes through the schools one of the reports he does is equipment verification. The purpose of that is to check the worthiness of each equipment, so if an equipment is found to be not reliable in that particular week it has to be withdrawn with documentation, and when it comes here in that particular week the technical personnel who is*

in charge of equipment will give us a backup equipment to take back to the school…" However, despite regular maintenance checks, technology usage in the classrooms faced challenges with malfunctions of projectors and solar panels.

Students, teachers, and Impact Network staff mentioned that occasionally the projectors break (usually a burnt bulb) and sometimes the tablets do not have sufficient charge, but that is usually linked back to a solar panel issue rather than the tablets themselves. In particular, the solar panels can be difficult to hastily repair if a part malfunctions because of how remote some schools are. An Operations Manager explained that it can take a long time to requisite the specific parts for solar panels. Beyond repair issues, the battery charge on solar panels was occasionally misused to charge teachers' or community members' personal devices, leaving no power left to charge student tablets for class. One parent from Lusangazi worried that children were overly reliant on the tablets for lessons: *"On my part I think on the issue of textbooks, I have not seen Impact bring in textbooks so that the children can use them maybe because they are very reliant on lesson plans and tablets but with regards to exercises."* Impact Network staff also acknowledged that in reality, the 7 tablets were split across classrooms, leaving too few to be shared feasibly for student group-work.

The Impact Network Director of Academics also mentioned that, "*Teachers are sometimes a bit timid with the equipment because they can get charged for part of the replacement cost; or if a sim card goes missing. They sometimes feel it's better not to use it at all and avoid taking the risk.*" This particular challenge can be mitigated through training efforts with teachers and should abate over time as teachers gain familiarity and proficiency with the technological tools provided to them.

Classroom Instruction – Materials

In terms of the physical learning materials provided to schools, most teachers, parents, and students felt that the items were generally sufficient to enable learning. Parents were appreciative that Impact Network provided these materials and did not ask them to contribute additional supplies at the parents' expense. An education programs officer from Katete listed the supplies delivered at the start of each term but mentioned that there were sometimes shortages: "*Flip Charts, chalk, Manila paper, pens, pencils, books for the learners, mops everything, a lot for school so at times we find that we are faced with the challenge of shortages.*" In these cases, Operations Managers were familiar with processes for requisitioning additional supplies as needed and mentioned that they monitor schools' supply inventories. However, the Impact Network Director of Operations mentioned that it was logistically difficult to distribute materials due to the volume of paperwork that passes from the Teacher Supervisors to the individual teachers. Impact Network staff emphasized timely delivery of

these supplies as another important factor for the program's success, beyond just the quantity of materials delivered.

As referenced above in the section on technological issues, parents, teachers, and students echoed the need for additional textbooks to be provided. Some suggested that added textbooks could serve as a fall back for when the tablets are unusable for reasons like lack of charge or malfunction. Others desired textbooks so students could bring them home as workbooks to continue their studies. One teacher in Katete stated, "*The biggest challenge in the classroom, some learning materials…. learning and teaching materials, like especially text books [are] not available.*" The reliance on technology to deliver the curriculum poses a potential failure point for the program's activities, though teachers were also provided with printed materials of backup lesson plans. Nevertheless, Impact Network staff should consider if more textbooks could be provided to supplement the printed materials and ensure that teachers can continue to deliver content in the case of technological issues.

While the Impact Network program was responsible for distributing chairs, benches, or other seating options for students, beneficiaries mentioned lack of seating as a challenge in the classrooms. One teacher from Petauke reported, "*…we also need more tables and benches because most of our learners cannot fit on the tables that are there…*" Another teacher described how this lack of seating affected group learning – without sufficient furniture, they were unable to effectively seat students together with similar ability levels. Some parents affirmed this seating shortage, saying that many of the benches were in poor condition. The Impact Network program could consider providing added benches to schools to help facilitate large classroom sizes and enable all attending students to feel included.

Other materials distributed by Impact Network included educational wall posters, with most teachers reporting that an adequate number was provided. Classroom observations confirmed that most classrooms had hand-drawn posters in their classroom setup. Students described how the posters depict things like a map of Zambia, common syllables for reinforcing literacy, multiplication tables for reinforcing numeracy, animals, and objects like houses and cars. These posters served to reinforce past lessons and students were pleased with the reminders.

Lastly, when asked for any other areas of difficulty in terms of classroom materials, one parent from Lusangazi mentioned that they would appreciate resources for at home learning. They said, "*Yes I would like to say that they don't have story books that the children can bring with them home. In the past the children used to be given those 'Jelita and Mulenga' books that they brought home and read and were able to learn so with Impact we don't see that happening.*" Reinforcing classroom literacy lessons with at-home reading of storybooks could be an impactful intervention for future cohorts. While workbooks for at-home learning were

sometimes provided, a Teacher Supervisor from Lusangazi mentioned that they were insufficient for the number of students. While this was not a core focus of the program activities, it presents an opportunity to cement the eSchool pedagogy delivered at school with added at-home resources.

Classroom Instruction – Pedagogy

Teachers were trained in specific classroom pedagogy for the eSchool 360 model to make use of the technology provided by Impact Network to deliver the curriculum developed by the Zambian Ministry of Education. The lesson content was well received by parents and students, who highlighted roleplays, singing, and writing as some of their favorite activities. Among the roleplay activities, topics included the consequences of misbehaving, early marriages, child abuse, drug abuse, and how to handle if your parents want you to farm/work rather than be in school. Students highlighted many of these roleplays as relevant and informative. While lessons were broadly received positively by students, teachers mentioned that they sometimes would modify lessons from the tablet version: "*The lessons that are found in the tablet, we need to modify them so that they can suit our learners.*" Impact Network staff were aware of these deviations and said that the changes had mixed results: "*Some teachers are exercising their own judgement based on what's in the curriculum. Sometimes positively (adjusting time spent on certain lessons), sometimes negatively (deciding a lesson isn't part of the national curriculum).*" The content was designed to be flexible, according to Teacher Supervisors, who explained that the tablets and curriculum books provided offer content scaled at various levels of student ability to enable teachers to adjust lesson delivery within the curriculum to meet student needs.

The Impact Network Director of Operations reported that the academics team worked hard to set teachers up for success: "*The academics team has spent a lot of time and energy creating curriculum books that try to make it as easy as possible for these teachers. It lays out that every day they should know what they should be teaching, they know the objectives of what they should be teaching, then you know they need to have lesson plans that align with the that.*" Teachers confirmed these efforts, saying that the curriculum books and term guides lay out which lessons they should be teaching on which days, and that these tools serve as a helpful guide. The main topic where there was reported difficulties in teaching was English literacy – Impact Network staff mentioned that teachers preferred to teach numeracy to literacy, and that the syllabus was demanding of students. Additionally, classroom observation reports suggested that teachers had some difficulties providing examples when teaching integrated science lessons.

In terms of classroom management, teachers, students, and parents all reported positive perceptions of how students were treated in Impact Network program schools. The program instituted a zero-tolerance policy for corporal punishment, so students were given extra cleaning duties as punishments instead and were excited to attend school. Parents praised the roleplay exercises in showing students the consequences of misbehaving and reported positive impacts on attendance as well: "*The reason why children do not miss school is because of the way the teacher handles the student. My child told me that even though he might get some things wrong in class the teacher does not shout at him the teacher just speaks and controls them well.*" The classroom observations confirmed that teachers were generally equipped with good tools for managing disruptions, calling out those who make noise and promptly dealing with those who misbehave.

The positive perceptions of Impact Network schools' classroom management practices extended to gender equity as well. Students reported that boys and girls were treated equally, and one teacher from Katete said, "*We engage girls in the classroom by choosing them as monitresses, and then even as group leaders when we give them group work.*" These efforts to co-mingle students sought to promote gender equity as a societal norm for the younger generation. However, while teachers and students gave qualitative reports of using peer learning in class this term, the classroom observations found only 3 out of 13 teachers successfully utilized peer learning (scoring at least a 3 on the 4-point scale for this metric) on the day of the visit. This suggests that the degree of peer learning may depend on that day's lesson plan or that it was less frequent than teachers led on in qualitative interviews.

Additional evidence from classroom observations suggests that some teachers struggled with modifying lessons to suit learners at different levels (average score of 1.9 out of 4), as well as providing differentiated support (average score of 1.6 out of 4). Another area of difficulty was with teachers not evaluating previously taught lessons nor tracking student understanding (average score of 1.7 out of 4). Ideally, teachers would take notes on student performance in certain lessons to identify those who struggled with the content and develop a support plan to help them catch up. Some classroom observations also identified that some teachers did not add fun and creativity to lessons (average score of 2.0 out of 4), suggesting that additional training could be provided on supporting students to reach lesson objectives. The areas with the greatest observed proficiency were Lesson Structure & Management (average score of 3.2 out of 4) and Content Knowledge & Instruction (average score of 3.5 out of 4), both of which are associated with the curriculum itself, lesson planning, and delivery of that content. These findings suggest that the delivery of the core content is sound, but teachers might benefit from added guidance on strategies for adjusting lessons to situational demands.

If teachers encountered challenges, they were instructed to reach out to their supervisor, who would then contact the broader Impact Network academics team as needed. Operations coordinators performed regular check-ins with program schools to ensure needs were met. There was an established chain of command for resolving challenges, and Impact Network staff seemed confident in that hierarchy's ability to provide teachers support. Some of the pedagogy-related challenges teachers faced included difficulties in using the technology due to unfamiliarity with computers, delivering lessons on construction that "*use a compass and ruler*" (according to the Impact Network Head of Academics), and issues managing large classroom sizes. Additionally, according to the Impact Network Education Programs Officer, teachers also worried about punishing a student whose parent is a practitioner of witchcraft, for fear of retribution: "*…I think the other challenge really is maybe like in these communities' teachers are living in the communities near the schools so they somehow know people who are into witchcraft and everything so once they learn that this child is coming from a mother or a father or a parent who is into witchcraft it's very hard for (laughter) the teachers to discipline those leaners because they are scared of the witchcraft issues.*" Impact Network staff should continue to monitor pedagogical challenges and tailor trainings to meet common issues, alongside ad-hoc support for unique situations.

Data Usage

The Impact Network program utilized the Comm Care software to track data on students and teachers in Impact Network schools. Teachers were responsible for recording student enrollment, attendance, and performance into Comm Care. Performance results were submitted at the end of the term. The Impact Network Director of Academics described how Teacher Supervisors monitored teacher attendance via Comm Care: "*Teachers only take time off during the term if teachers are sick or have an emergency. Attendance is monitored by student supervisor, but it's a challenge. Teachers are required to sign in at the school. Teachers have to document their sick days, report to their supervisors…If School Support Officer [found] teachers not at school, that [was] reported. This can affect contract renewal.*" This monitoring enables the Impact Network program to ensure its teachers are at school as often as possible, delivering the curriculum on schedule. With respect to student attendance, teachers reported that more than three consecutive absences by a student results in a follow up from the school support officer to reach out to the family and investigate the reason for the absences.

Student performance data was captured using a literacy assessment and a numeracy assessment, as well as a Catch Up assessment specific to that program component. Teachers described how these data were used to identify struggling students and provide additional assistance. Impact Network staff also mentioned that student performance data was used to tailor trainings for teachers to assist them on topics where there are common difficulties.

Teachers used the performance data to group students by ability level, with Impact Network providing trainings so teachers felt confident to manage these groupings themselves.

No challenges were reported with data usage or using the Comm Care system, suggesting that this aspect of program implementation was well managed and that the software was intuitive.

Child Protection

Respondents agreed that the most prevalent child protection concerns are related to dropout, cattle herding, early marriage, pregnancy, and assault. While some respondents reported an uptick in teenage pregnancy cases during and following COVID-19-related school closures, others said they did not notice much of a difference in the types of child protection issues they encountered. There is a clear referral process in place to respond to such concerns, including a fast-track for more severe cases such as assault. Additionally, community sensitizations were held about how to recognize and respond to different types of abuse. Impact Network has a strict child protection policy that all employees are required to sign and a zero-tolerance policy for corporal punishment. Parents reported that Impact Network takes child protection issues very seriously, and children consistently reported knowing to approach their teacher if they have a problem at home or at school.

While teachers are trained on child protection, both Impact Network staff and teachers themselves said they need more training on how to handle more complex child protection challenges. A Petauke teacher commented, *"I think on this one we might require additional training, cause being in rural we cannot know how to handle some of the problems, and sometimes if you come in you try to help, sometimes the feedback you will be getting is* 'you came with both your legs to this village you will go back without them if you continue interfering in other people's business.'*"* Other teachers echoed this sentiment, saying they sometimes struggle with how to respond to child protection issues and sometimes fear for their safety when following up on these issues..

Teacher Working Conditions and Pay

Teacher pay and teacher working conditions were significant concerns for respondents. While teachers are paid regularly and on time, the amount is so nominal that one Impact Network staff member referred to it as *"essentially a stipend."* Teachers, Teacher Supervisors, and Impact Network staff all agreed that low salaries were a demotivating factor for teachers. Inflation and the fact that some teachers also support other family members financially exacerbated teachers' frustrations with their low salary.

In addition to dissatisfaction with salary, some teachers also struggled to adjust to life in more rural settings and in some cases reported strained relations with the local community. As one Impact Network staff member put it, "*[We] brings in some extra teachers to the initial training to accommodate for some of them leaving after seeing the village."* Limited mobile network, bad roads, poor accommodations, and limited access to health care were among teachers' most frequent complaints. School infrastructure, too, was a source of frustration for some teachers who complained that toilets were insufficient, and they were forced to share toilets with learners in some cases.

Impact Network staff found teachers smart, ambitious, and often eager for the next step which contributes to turnover. Respondents noted an upcoming[20] large-scale government recruitment of teachers and voiced concerns about losing many of their Impact Network teachers: *"Teacher turnover is a bit of a challenge. Teachers we hired now, there is about to be a big government recruitment of 30,000 teachers. All of our teachers will apply for that, it's a good opportunity and salaries are better."* Others added that the recent requirement for a teacher diploma led to a significant number (25, by one account) of Impact Network teachers not being renewed (because of a government mandated change in the requirement).

Variation by Geography, Culture, and Season

Geography affected the Impact Network program's implementation in limited ways. Lusangazi appeared to have more problems with attendance due to longer distances from school. This might have impacted student attendance rates, lessening the program's impact on encouraging students to attend regularly. PTAs in Lusangazi also seemed to be less established and not have a lot of participation from parents. Low parental involvement could also be a side effect of a more distributed community, and the program's sustained impact on the community might suffer as a result.

Seasonality mainly impacted the program's implementation with respect to the solar panels' efficacy. During the rainy season or periods of bad weather, the solar panels would struggle to capture sufficient charge to power the tablets or projector. This meant more days spent relying on printed materials and textbooks for delivering the curriculum during the rainy season. Because so much of the program's content delivery rests on the technology to have power, seasonal lack of access presents a significant issue. This reinforces the need for alternative

[20] This recruitment had happened at the time of the endline survey, resulting in significant teacher turnover.

electricity sources to be provided (like a diesel generator) or for lessons plans for the rainy season to be designed around hard copies of the content.

COVID-19 also impacted the program's implementation. Beyond providing sanitizer and masks to classrooms, qualitative reports from Impact Network staff revealed that the pandemic affected classroom use of tablets – there was a period where teachers lessened tablet use to keep students socially distanced in the classroom, because otherwise they would crowd around the devices. The student tablets were mostly utilized for group work, and the Impact Network staff observed that sanitizing the tablets between users was very time consuming. Lessons were also shortened according to one operations coordinator, "*[W]e had a limited time for learners to be in class attending the full lessons. So, it was not being used from the start to the end, just because you find that the lesson that was supposed to run for an hour it was reduced to thirty minutes, and for the teacher to connect the system to project, it wasn't that helpful for the learners.*" To supplement the shortened lessons, teachers distributed homework packs and performed regular home visits to check in on students' progress. Parents responded positively to these home visits, with some getting engaged with their child's education to assist with the homework packs and prepare for the teacher's visit.

7. Conclusion

While the impact estimates are smaller than during the midline survey, the cluster-randomized controlled trial of the eSchool 360 program continues to show positive impacts on literacy and mixed evidence for positive impacts on mathematics more than four years after the start of the multi-faceted technology-aided instruction program. The results show positive impacts of 0.26 standard deviations on EGRA. We do not find statistically significant average effects on EGMA outcomes or on the Zambian achievement and oral vocabulary assessments, but evidence for some positive effects on sub-groups.

We find evidence for larger heterogeneous impacts for boys and students who scored higher on the learning assessments during the baseline. We find positive and statistically significant effects of 0.20 standard deviations on EGMA for boys, and 0.19 standard deviations on EGMA for children who scored higher than the median EGMA score during the baseline. The study also shows smaller impacts on EGRA outcomes for girls and children who scored lower on the EGRA assessment during the baseline.

The sustained program impacts on EGRA outcomes are likely associated with the higher quality of education in Impact Network schools. We find large and statistically significant effects on the

perceived quality of education in Impact Network schools. This is consistent with the process evaluation findings demonstrating overall positive perceptions of the quality of education in Impact Network schools. Process evaluation respondents also suggest that teacher attendance is notably higher in Impact Network schools. In addition, many respondents report the strong alignment of the eSchool 360 program with the Catch-Up program, which focuses on teaching at the right level and activity-based learning, just like the eSchool 360 program. This is consistent with impact evaluation results demonstrating a higher fidelity of implementation of the Catch-Up program in Impact Network schools.

We also find evidence for higher learning outcomes in Impact Network schools than in government schools. Impact network students in grades 5 and 6 score considerably higher in EGRA, EGMA, ZAT, and the oral vocabulary test than students in government schools. The higher quality of education in Impact Network likely contributed to these sizeable and statistically significant differences. The process evaluation further suggests that student attendance is higher in Impact Network than in government schools, which may also have contributed to the differences in learning outcomes.

The COVID-19 pandemic and the associated school closures and lower enrollment likely contributed to the smaller program impacts during the endline compared to the midline. In contrast to the midline survey, we did not find statistically significant effects of the program on school enrollment or attendance. In addition, drop-out from the program was larger among students who had lower EGRA and EGMA scores during the baseline assessment. It remains challenging, however, to attribute the smaller program impacts to COVID-19 and the associated school closures.

One additional reason for the smaller impacts is the relatively low dosage of the program among children in the treatment group. Because the program was only implemented in two cohorts, students who repeated class or enrolled in school later did not benefit fully from the eSchool 360 program. In fact, only 49 percent of the sampled children in the treatment schools ever enrolled in an Impact Network school, and only 20 percent participated for the full duration of the program.

The effects for students enrolled in school and students who attend school regularly indeed often are larger, but the point estimates are not always statistically significant. We find sizeable and statistically significant effects on EGRA outcomes for those who enroll and regularly attend school, but the effects on EGMA, ZAT and oral vocabulary assessment scores are often not statistically significant for these sub-groups despite the larger point estimates. While the results are vulnerable to weak instrument bias, we did find sizeable and statistically significant effects on EGRA and EGMA outcomes for students who enrolled in Impact Network classes for the full

duration of the program (20 percent of the students in the treatment group). The lower grade repetition among these students than in the control group may also have contributed to the larger positive impacts for this group.

The cost-effectiveness analysis at endline suggests that the program generates between 0.35 and 1.40 standard deviations in EGRA outcomes and 0.11 to 0.25 standard deviations in EGMA outcomes for every additional $100 spent. The cost-effectiveness estimates differ depending on 1) whether they include costs that the government would likely also have made (but for which we do not have data), and 2) whether they include teacher salaries paid by Impact Network (which the government would also pay after program scale-up).

Scaling back costs to midline where Treatment schools were 14 months into program implementation, we find substantially larger cost-effectiveness of the program, owing to both larger impacts and lower costs at midline. Estimates suggest that every $100 spent on the program resulted in an additional 1.38 to 3.30 SD gain in EGRA and 0.73 to 1.62 SD in EGMA.

8. Recommendations

We provide various recommendations to the AIR Opportunity fund, the Zambian government and Impact Network based on the midline and endline findings. These recommendations include suggestions on how to scale up the program, specific implementation features, and evidence-gaps that require more research to provide conclusive evidence. When providing recommendations based on the endline findings, it is important to consider the context of the COVID-19 pandemic and the associated school closures. The program impacts four years after the start of the program implementation were likely smaller because of increased drop-out after COVID-19 and the associated school closures. For this reason, we also consider the possibility that the midline impact estimates are more externally valid for the Zambian context after the COVID-19 pandemic than the endline findings. At the same time, it remains important to consider the endline findings in the recommendations. While the endline results may not be fully externally valid, they provide relevant information about the potential longer-term impacts of the eSchool 360 program. The midline results only provide information about the program impacts 14 months after the start of the program implementation.

Recommendations to scale up the eSchool 360 program

We recommend the AIR Opportunity Fund to invest in Impact Network to expand the program to grade 1 and grade 2 of the control schools: The midline and endline findings indicate that upgrading the infrastructure of control schools and implementing the eSchool 360

model in treatment and control schools will bring substantial educational benefits to the control schools and enable Impact Network to continue expanding its primary school program in a cost-effective manner. The endline survey shows that the eSchool 360 model improves literacy outcomes and mathematics outcomes for boys and children with higher EGMA scores during the baseline. Most of these impacts were generated during the first two years of the program implementation during which the program showed larger impacts on literacy and mathematics outcomes than during the endline survey. These findings suggest that expanding the program to Grade 1 and 2 of the control schools will generate substantial benefits.

While it is possible that the program would have generated larger impacts on students in grade 5 and 6 in the absence of the COVID-19 pandemic, current evidence is insufficient to recommend expanding the program to all grades in the control schools. The program effects and cost-effectiveness reduced considerably between the midline and the endline survey. It is likely that the COVID-19 pandemic contributed to reducing the impact of the program. However, it is uncertain whether expanding the program to all grades in control schools would result in sufficient benefits to justify a large investment. In addition, the costs of the program would increase considerably when expanding the program to all grades in the control schools.

We recommend the AIR Opportunity Fund to examine opportunities for integrating literacy elements of the eSchool 360 model as part of Zambian government schools: The midline and endline results show substantial positive impacts of the eSchool 360 model on literacy outcomes, especially in grades 1 and 2. In addition, Impact Network has invested in a new curriculum based on phonetic charts (the "Read Smart" program) that could generate additional impacts on literacy outcomes. However, at this moment only community schools implement the Read Smart program. Expanding the eSchool 360 program will require integrating the literacy components of the eSchool 360 model as part of government schools, while conducting additional research on the impact and cost-effectiveness of the "Read Smart" program in community and government schools.

We recommend the AIR Opportunity Fund, the Zambian government and international donors to look for additional opportunities to improve the quality of secondary education for community school students who enroll in secondary schools in the Impact Network catchment areas: While the eSchool 360 model has positive impacts on learning outcomes in primary school, it remains uncertain how much children would benefit from the program if the quality of secondary education remains relatively low. Low-quality secondary education may result in the fade-out of improvements in learning outcomes in primary schools after students enroll in secondary school. Strengthening and empowering school leaders to create a secondary school environment that enables high-quality learning could also incentivize secondary school

participation and bring sustainable learning benefits to students who enroll in secondary school. This suggests that additional investments are needed to pilot and assess the effectiveness of innovations to improve the quality of secondary education in rural Zambia.

Recommendations to improve implementation of the eSchool 360 program

Based on the process evaluation findings, we suggest that Impact Network consider the following recommendations to improve its programming:

- **Device maintenance and charging** – Despite regular maintenance checks, Impact Network schools faced challenges related to malfunction of projectors and solar panels. To minimize these challenges and improve the reliability of technological tools at Impact Network schools, Impact Network should consider having a stockpile of commonly replaced parts for projectors and solar panels to avoid wait-times on shipping the parts from the United States. Additionally, providing more training and guidance to teachers and school staff on proper use of the devices can help to address the use of solar power to charge personal devices.
- **Analog teaching materials** - Given the potential for failure and/or malfunction of technological devices, Impact Network should consider if more textbooks could be provided to supplement the existing printed materials provided to teachers (e.g.: backup lesson plans) and ensure that teachers can continue to deliver content in the case of technological issues. Additionally, providing some printed learning materials to students can help meet the demands from some parents and caregivers for more resources for at-home learning, such as storybooks.
- **Seating** - When budgeting, Impact Network should consider allocating funds to provide additional classroom seating (benches or desks) to ensure all children have their own seat. Many respondents noted inadequate classroom seating, with some saying that the lack of seats is especially problematic for the type of group work incorporated in both Catch-Up classes and the e-school 360 model.
- **Coaching** – Although coaching was generally considered effective in supporting teachers, some teachers noted that the frequency of coaching/observation sessions was sometimes excessive and that less frequent sessions could give teachers more "breathing room." In addition, Teacher Supervisors reported difficulties in the logistics of traveling to schools. Impact Network should consider developing a progressive observation schedule where the coaching frequency can decrease to biweekly sessions for teachers with a certain length of service who are performing well. This would decrease transportation costs/time for Teacher Supervisors while still supporting teachers and keeping them accountable. Also, Impact

Network should consider making the coaching rubric available to teachers so they can track progress and areas of improvement.

- **Additional support to the Lusangazi district**[21] - Given the challenges with student attendance and parental engagement in parent-teacher associations in Lusangazi, we recommend Impact Network provides additional support to the Lusangazi parent-teacher association to encourage parental participation and highlight the importance of regular student attendance.
- **Support for more complex child protection challenges** - Interviews with Impact Network teachers showed that they need more support to safely resolve more complex child protection challenges. We suggest Impact Network provides additional training and written guidance on how to resolve more complex child protection issues and consider additional ways to ensure that teachers feel both empowered and safe as the work to resolve these issues.
- **Teacher salaries** - We suggest that Impact Network regularly reassesses and potentially recalibrates teacher salaries based on required credentials and the government pay scale. At the time of data collection, teachers expressed considerable frustration with their compensation and reported it was much less than what government teachers were paid.[22]

The eSchool 360 model has considerable potential to generate positive impacts on learning outcomes at scale when the government would scale-up the program. Incorporating the recommendations to scale up and improve implementation of the program may generate even larger impacts at a lower cost.

[21] Lusangazi district was made independent from Petauke district in 2018
[22] It is important to recognize, however, that community school teachers in control schools are volunteers

References

Angrist, N., de Barros, A., Bhula, R., Chakera, S., Cummiskey, C., DeStefano, J., Floretta, J., Kaffenberger, M., Piper, B., & Stern, J. (2021). Building back better to avert a learning catastrophe: Estimating learning loss from COVID-19 school shutdowns in Africa and facilitating short-term and long-term learning recovery. *International Journal of Educational Development*, 84.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., … Wasserman, M. (2016, October). *Mainstreaming an effective intervention: Evidence from randomized evaluations of "Teaching at the Right Level" in India* [NBER Working Paper No. 22746]. Cambridge, MA: National Bureau of Economic Research. Doi: 10.3386/w22746

Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Pariente, W., … Udry, C. (2015). A multifaceted program causing lasting progress for the very poor: Evidence from six countries. *Science, 348*(6236). Doi: 10.1126/science.1260799

Barrera-Osorio, F., & Linden, L. L. (2009). The use and misuse of computers in education: Evidence from a randomized controlled trial of a language arts program. Retrieved from http://www.leighlinden.com/Barrera-Linden%20Computadores_2009-03-25.pdf

Beg, S.A., Lucas, A.M., Halim, W., & Saif, U. (2019). *Beyond the basics: Improving post-primary content delivery through classroom technology*. New York, NY: National Bureau of Economic Research. doi: 10.3386/w25704

Crawfurd, L., Evans, D., Hares, S., & Sandefur, J. (2023). Live tutoring calls did not improve learning during the COVID-19 pandemic in Sierra Leone, *Journal of Development Economics*, 164.

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.* Washington, DC: American Institutes for Research.

De Hoop, T., Brudevold-Newman, A., & Davis, D. (2018). *Baseline report for the mixed-methods cluster-randomized controlled trial of Impact Network's eSchool 360 model in rural Zambia.* Washington, DC: American Institutes for Research.

De Hoop, T., Ring, H., Siwach, G., Dias, P., Tembo, G., Rothbard, V., & Tounguis, A. (2020). Midline report for the mixed-methods cluster-randomized controlled trial of Impact Network's eSchool 360 model in rural Zambia. Retrieved from https://www.air.org/sites/default/files/2021-06/Midline-Report-Mixed-Methods-Cluster-RCT-of-eSchool-360-Model-in-Rural-Zambia-February-2020-rev2.pdf

De Hoop, T., Ring, H., Siwach, G., Dias, P., Tembo, G., Rothbard, V., & Tounguis, A. (2023). 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.*

Dhaliwal, I., Duflo, E., Glennerster, R., & Tulloch, C. (2011, August). *Comparative cost-effectiveness analysis to inform policy in developing countries: A general framework with applications for education*. Retrieved from http://economics.mit.edu/files/6959

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.

Eble, A., & Escueta, M. (2022). *How demand and supply interact to generate learning in settings of extreme poverty.* Retrieved from https://www.edworkingpapers.com/sites/default/files/Eble%20Escueta%20-%20Demand%20supply%20learning%202022-10-20.pdf

Evans, D., & Mendez Acosta, A. (2020). Education in Africa. What are we learning? *Journal of African Economies, 30*(1), 13-54.

Evans, D., Hares, S., Mendez Acosta, A., & Saintis, C. (2021). It's been a year since schools started to close due to COVID-19. Retrieved from https://www.cgdev.org/blog/its-been-year-schools-started-close-due-covid-19

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.

Ganimian, A., & Murnane, R. (2016). Improving education in developing countries: Lessons from rigorous impact evaluations. *Review of Educational Research, 86*(3), 719-755.

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. Retrieved from https://bfi.uchicago.edu/wp-content/uploads/2022/06/BFI_WP_2022-68-1.pdf

Haelermans, C., Jacobs, M., Van der Velden, R., Van Vugt, L., & Van Wetten, S. (2022). Inequality in the effects of primary school closures due to the COVID-19 pandemic: Evidence from the Netherlands. American Economic Review: Papers and Proceedings, 112, pp. 303-307 https://doi.org/10.1257/pandp.20221107

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, 17(3), 630-651.

Imbens, G. (2015). Matching methods in practice. *Journal of Human Resources, 50*(2), 373–419.

Jack, R., & Oster, E. (2023). COVID-19, school closures, and outcomes. *Journal of Economic Perspectives*, 37(4), 51-70.

Ministry of General Education, Republic of Zambia. (2016, February). *Educational statistical bulletin*. Retrieved from https://www.moge.gov.zm/download/statistics/annual_statistics/Educational-Statistical-Bulletin/2014-Educational-Statistical-Bulletin.pdf

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.

Nakamura, P., Molotsky, A., Castro Zarzur, R., Ranjit, V., Haddad, Y., & de Hoop, T. (2023). Language of instruction in schools in low – and middle-income countries: A systematic review. *Campbell Systematic Reviews*.

Organisation for Economic Co-Operation and Development (2023). Decline in educational performance only partly attributable to the COVID-19 pandemic. Retrieved from https://www.oecd.org/en/about/news/press-releases/2023/12/decline-in-educational-performance-only-partly-attributable-to-the-covid-19-pandemic.html

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.

Piper, B., Zuilkowski, S. S., Kwayumba, D., & Strigel, C. (2016). Does technology improve reading outcomes? Comparing the effectiveness and cost-effectiveness of ICT interventions for early grade reading in Kenya. *International Journal of Educational Development, 49,* 204–214.

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 Daniel, Schueler Beth E. Can learning be measured by phone? Evidence from Kenya Annenberg Institute at Brown University (2022) Technical Report.

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

Snilstveit, B., Stevenson, J., Menon, R., Phillips, D., Gallagher, E., Geleen, M., … Jimenez, E. (2016, September). *The impact of education programmes on learning and school participation in low- and middle-income countries: A systematic review summary report.* London, UK: International Initiative for Impact Evaluation. Retrieved from http://pubdocs.worldbank.org/en/429341481774674870/The-Impact-of-Education-Programmes-on-Learning-and-School-Participation.pdf

Tauson, M. and Stannard, L. (2018). EdTech for learning in emergencies and displaced settings. New York, NY: Save the Children. Retrieved from https://resourcecentre.savethechildren.net/node/13238/pdf/edtech-learning.pdf

Valiente, O. (2010). "1-1 in Education: Current Practice, International Comparative Research Evidence and Policy Implications", *OECD Education Working Papers*, No.44, OECD Publishing, Paris. doi: 10.1787/5kmjzwfl9vr2-en.

Valk, J.-H., Rashid, A. T., & Elder, L. (2010, March). Using mobile phones to improve educational outcomes: An analysis of evidence from Asia. *The International Review of Research in Open and Distributed Learning, 11*(1), 117–140. Retrieved from https://files.eric.ed.gov/fulltext/EJ881581.pdf

World Bank. (2018). *World development report 2018: Learning to realize education's promise.* Washington, DC: Author. Retrieved from https://www.worldbank.org/en/publication/wdr2018

Young, A. (2019). Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. The Quarterly Journal of Economics, 134(2), 557–598.

Appendix A. Adjusted P-Values for Multiple Corrections in EGRA & 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 Heterogenous 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 Heterogenous 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 Heterogeneus 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

Appendix C. Outcome Measures Related to Learning

We fielded a consistent assessment instrument across the baseline and midline to measure the impact on student outcomes 1 year (midline) after the introduction of the Impact Network model in schools. The primary cognitive skills outcomes are aggregate scores on the EGRA, EGMA, and ZAT; the secondary outcomes are the EGRA, EGMA, and ZAT subtasks as well as measures of oral reading fluency. As students learn different concepts at different ages, we expect to see different impacts on subtasks at the midline relative to the endline (yet to be executed).

At the midline, the secondary literacy outcome measures were the four-emergent literacy EGRA subtasks: concept of print, oral vocabulary, phonological awareness, and decoding. For EGMA, the secondary math outcomes at midline were oral counting fluency, one-to-one correspondence, number identification, quantity discrimination, and the time it takes to complete each of these sections. In the following we define each of these outcome measures. Much of the discussion on the early grade math assessment constructs is based on RTI International (2009).

Concept of print: This task measures whether a child understands how print "functions"—how to hold a book, where the beginning and end of a book is, and so forth.

Phonological awareness: This task measures phonemic awareness by using a sound identification task in which the data collector sounds out three words (with corresponding pictures on the student sheet), and the child identifies the one word that has a different first sound (syllable and phoneme).

Decoding: This task measures the ability to sound out words in print. A child needs to be able to do this automatically (without time or effort) to free up the cognitive resources required for reading larger amounts of text (Perfetti, 1985). We will test decoding skills using real words and pseudo-words separately. (Pseudo-words are combinations of letters that do not form meaningful words but are not precluded by the grammatical rules of the language.)

Reading comprehension: This task measures reading comprehension by using a short passage with literal and basic inferential comprehension questions.

Oral reading fluency: This task measures oral reading fluency by counting the correct words read per minute.

Oral counting fluency: This task measures oral counting fluency by assessing a child's ability to produce numbers fluently. The task asks a child to count as high as possible, usually beginning with the number 1, until they make an error (Floyd, Hojnoski, & Key, 2006).

One-to-one correspondence: This task measures one-to-one correspondence by assessing the extent to which a child recognizes the items they need to count and the extent to which the child recognizes, and mentally tags, those items that have already been counted.

Number identification: This task measures number identification by assessing the extent to which a child orally identifies printed number symbols that are randomly selected and placed in a grid.

Quantity discrimination: This task measures quantity discrimination by assessing a child's ability to make judgments about differences by comparing quantities in object groups. We will measure this ability by presenting two groups of objects and ask which group has more objects.

Missing number: This task measures the ability to find a missing number. In this task, we will present children with a string of three numbers with the first, middle, or last number in the string missing. The child reports which number is missing.

Addition and subtraction: This task measures addition and subtraction skills by presenting a child with oral or written problems with a focus on addition and subtraction. We will show a visual representation of the mathematics problem and read the problem out loud.

Geometry: This task measures whether a child recognizes shapes. The child is asked to point to all representations of one shape on a sheet of paper. The score is based on the number of correctly and incorrectly marked shapes.

Oral vocabulary: This task measures receptive oral language skills separately from any decoding/script-processing ability. In this test, a child sees four pictures, listens to the data collector say the name of one picture, and is asked to point to the correct picture in their test booklet.

Previous experience in Zambia suggests that a majority of students will score quite low on their EGRA and EGMA tests. Furthermore, it is likely that any improvements by endline will be small because of the difficulty of these tests. These so-called floor effects raise concerns about the ability of our impact evaluation to detect statistically significant effects of the program on EGRA and EGMA outcomes. For this reason, we included the prereading recognition subtest of the ZAT as a complementary test of preliteracy skills.

The ZAT assessment, developed for use in multiple Zambian languages, was specifically constructed for the context of Zambia (and of Zambia's Eastern Province) to measure academic achievement in mathematics, reading (letter and word) recognition, pseudo-word decoding, and reading comprehension. The prereading skills subtest consists of 34 items and is constructed so that a child simply needs to show that they can recognize the shapes and sounds of certain letters. A previous nonexperimental evaluation of the Impact Network program shows that the ZAT is less vulnerable to floor effects when implemented in the Eastern Province.

Examples of ITT Effects on Learning Outcomes

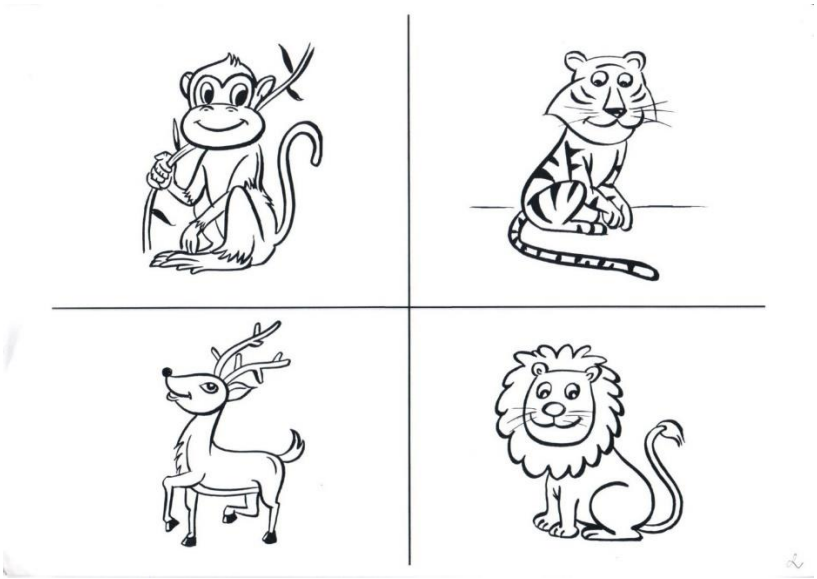
Below we provide several examples from the four tests to show how ITT effects translated into differences in learning outcomes.

1. ZAT: On average, children living in the Impact Network catchment areas score 2.5 percentage points higher than other children on tasks such as:
   - Recognizing the letters FI among a list of letter groups that includes IP, TT, TI and FI.
   - Recognizing and pointing to the letter R when shown a group of pictures that includes a lion, a giraffe, a rhinoceros and the letter R.
   - Pointing to a picture of an object that begins with the same sound as another object among a group of objects.
2. EGRA: On average, children living in Impact Network catchment areas scored 2.9 percentage points higher than other children on tasks such as:
   - Answering questions based on reading a story about a friendship between a child and a snake.
   - Determining where a sentence begins, in which direction they should read, and where to go when they reach the end of a line.
   - Answering questions based on listening to a story about a child going to school for the first time during the rainy season.
3. EGMA: On average, children living in Impact Network catchment areas scored 3.3 percentage points higher than other children on tasks such as:

- Recognizing numbers such as 9, 97, or 468.
- Identifying which number is missing from a set of patterns (for example, from a pattern of 10, 20, …, and 40, the child was asked to identify the missing number).
- Adding two numbers, such as 67 + 25.

4. Oral Vocabulary: Children living in Impact Network catchment areas were 5.3 percentage points more likely to listen to a word given to them and point to a picture in a booklet that best represented that word. Exhibit 1 is an example given to children participating in this study: An enumerator may, for instance, have said the word “lion” and then asked a child to point to the image of a lion.

Table C1. . Example of a Question from the Oral Vocabulary Test



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**

XXXXX_MO/23