Inception Report for the Mixed-Methods Cluster-Randomized Controlled Trial of Impact Network's eSchool 360 Model in rural Zambia

Thomas de Hoop, Andrew Brudevold-Newman, Hannah Ring, Arianna Zanolini, Claire Nowlin, and Dustin Davis
American Institutes for Research

September 2017

AMERICAN INSTITUTES FOR RESEARCH®

Contents

Tables

Figures

Introduction

There is a global learning crisis (Pritchett, 2013; Berry, Barnett, & Hinton, 2015). Improvements in students' learning achievement have lagged behind in low-and middle-income countries despite significant progress in school enrollment numbers. For several years now, large-scale early grade reading and math assessment data (e.g., Annual Status of Education Report [ASER], 2013; EdData II, n.d.) have highlighted worryingly high "zero" scores in reading assessments across the world. Approximately 250 million children across the world are not acquiring basic reading and math skills, even though about half of them have spent at least 4 years in school (UNESCO, 2014).

Although the use of technology in education shows a lot of promise for improving learning outcomes in low-and middle-income countries, the evidence on its impacts is underwhelming. Its effectiveness could, however, be improved if technology were combined with a strong focus on pedagogical practices. A review of high-quality studies with a focus on the impact of the use of technology in education in low-and middle-income countries finds "mixed evidence with a pattern of null results" (Bulman & Fairly, 2016). Recent evidence shows more promising results, however. For example, Muralidharan, Singh, and Ganimian (2016) show that a technology-based after-school instruction program with a strong emphasis on learning at the right level can produce large and statistically significant effects on reading outcomes. A cost-effectiveness analysis suggests that this program was also highly cost-effective (Muralidharan et al., 2016). Assessing the cost-effectiveness of successful technology-based education programs is not straightforward, however. A recent review demonstrates that technology-based education programs may not be cost-effective even if they produce large impacts on learning outcomes (Piper, Simmons Zuilkowski, Kwayumba, and Strigel, 2016).

This study focuses on the effects and cost-effectiveness of Impact Network's eSchool 360 program in producing improvements in learning outcomes among students in community schools in rural Zambia. The core of the eSchool 360 model is e-learning technology; tablets and projectors provided by Impact Network's partner, iSchool, loaded with a curriculum in the local language approved by the Zambian government. Impact Network supplements the technology by providing teacher training and professional development to community school teachers, providing electricity via solar power, and creating community ownership. Locally hired teachers receive weekly training focused on using the technology and enhancing their pedagogical skills.

In Zambia's transition away from a socialist economy in the 1990s, Zambian communities began forming their own community schools because of concerns about the large number of uneducated children (De Stefano, 2006). The number of community schools has grown rapidly in Zambia, from an estimated 100 in 1996 to about 3,000 schools with 600,000 children today (Chimese, 2014). Community schools are often staffed by untrained and underpaid teachers; they lack management and school supplies, and they teach a substandard curriculum despite the official recognition of community schools by the Zambian government in 1998 (De Stefano, 2006). Many community schools have fallen into disrepair and do not offer the full range of primary grades. The schools are operating autonomously from the government and are funded with less than $91 per year (1,000 Zambian Kwacha) per school (De Stefano, 2006). They are in need of a cost-effective solution for delivering quality education to improve learning outcomes.

A previous nonexperimental evaluation of the eSchool 360 program suggests that the program may be cost-effective in improving learning outcomes among primary school students in poor isolated areas in rural Zambia (Schling & Winters, 2015). The study compared the academic performance of first and second graders at Impact Network schools, government schools, and community schools using a longitudinal design with two rounds of data collection. The analysis indicated that improving math outcomes by one percentage point cost 88% less in Impact Network Schools than it did in government schools (Schling & Winters, 2015).

It is with this history that American Institutes for Research (AIR) decided to fund the expansion of the eSchool 360 program to 30 additional community schools in rural Zambia (from a total of 5 schools to 35 schools). Impact Network will expand the eSchool 360 program to 30 schools across three Zambian districts (Katete, Sinda, and Petauke). The schools are located in the Eastern Province of Zambia, in a rural area with no running water and limited electricity. Households with children in Impact Network schools are significantly poorer than students in government schools (Schling & Winters, 2015).

The expansion will be accompanied by a rigorous mixed-methods randomized controlled trial (RCT) to determine the impact of the program on students' learning outcomes. To achieve this goal, AIR and Impact Network closely consulted with Zambian government officials to obtain letters of approval for random assignment of the eSchool 360 program to 30 treatment and 34 control schools. We successfully completed this random assignment by the end of May. The randomization was implemented by Ministry of Education officials and the integrity of the process was ensured by AIR staff. We chose an unbalanced design with a smaller number of treatment schools because of limited resources to implement the eSchool 360 model. We will supplement the RCT with a process evaluation to assess the fidelity of implementation of the eSchool 360 program, and a cost-effectiveness analysis.

The study focuses on the impact of the program on learning outcomes and on various intermediate outcomes along the causal chain of the theory of change, as well as on the implementation of the program to examine why and how the program achieves its impacts. In addition, we will examine the cost-effectiveness of the program in achieving learning outcomes. The overarching research questions for the study fall into three categories and are as follows:

Impact Evaluation

a. What is the effect of eSchool 360 on students' numeracy, pre-literacy and literacy skills?
b. Do students enrolled in the eSchool 360 program improve in numeracy and literacy skills?
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 parents' and childrens' aspirations?

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 math outcomes?

Process Evaluation

a. Was the 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 does the program implementation vary by geography, culture, and time of year?

c. Do perceptions of the quality of teachers differ among students, parents, teacher supervisors, and teachers? If yes, how?

The rest of this inception report is structured as follows. We start with a description of the background of the program, including an overview of the existing literature on the impact of technology-based education programs. Next we present a description of the program, followed by a description of the theory of change. We end the report with a detailed overview of the mixed-methods research design that will enable us to address our research questions.

Background

The current Zambian educational system is not widely accessible to the country's potential students, and the quality of education it provides is low. Public spending on education constitutes only 1.3% of gross domestic product, compared to an average of 5.6% in Southern and Eastern Africa. Literacy rates among young adults ages 15–24 are only 58.5% for females and 70.3% for males (UNICEF, 2015). Zambian government policies focus too much on educational inputs at the expense of investing in the quality of education. For example, a rigorous study found no statistically significant association between learning outcomes and a fixed block grant provided by the Zambian government (Das et al., 2013).

In addition, the Zambian government does not provide adequate resources for autonomously operating community schools (DeStefano, 2006). The government funds community schools with less than $91 per year (1,000 Zambian Kwacha) per school. However, the number of community schools has grown rapidly in Zambia, from an estimated 100 in 1996 to about 3,000 schools with 600,000 children today (Chimese, 2014). Community schools are often staffed by untrained and underpaid teachers, lack management and school supplies, and teach a substandard curriculum. Many community schools have fallen into disrepair and do not offer the full range of primary grades. They are in need of a cost-effective solution for delivering quality education in order to improve early grade reading and early grade math outcomes.

Previous research demonstrates that a combination of teacher training and curriculum improvement such as that provided by eSchool 360 can be effective in improving learning outcomes. Kremer, Brannen, and Glennerster (2013) highlight the importance of adapting the curriculum to the child's level to improve the effectiveness of pedagogical interventions. A review of Conn (2014) suggests that among interventions that included teacher training as a component, those with adaptive instruction had larger effect sizes than those without adaptive instruction. Muralidharan et al. (2017) find that in an RCT of a personalized computer-aided after-school instruction program in India, students in the treatment group made significant gains in math and Hindi test scores. They conclude that the impact was due primarily to the computer-aided learning system's ability to target and adapt to the wide variation in student learning levels.

However, Schling and Winters (2015) note that some research in developing countries on the effect of changes to educational access and quality has yielded mixed results. An increased provision of traditional school resources such as textbooks or flip charts had no impact on student attainment (Glewwe, 2002). Banerjee, Cole, Duflo, & Linden (2007) note that increasing inputs to schooling fails to have an impact on student attainment if what is being taught remains too difficult for students to follow. Similarly, a number of studies focusing on computer-assisted learning programs do not find significant impacts throughout. For example, Cristia, Ibarrarán, Cueto, Santiago, & Severín (2012) analyzed the effect of the One-Laptop-Per-Child program for students in rural Peru and found little impact on the attendance and educational attainment of students. They argue that this lack of impact is due to the computers not containing software that was directly linked to class material, such as mathematics or reading, as well as a lack of clear instruction on how teachers should use the computers in class. However, Banerjee et al.

(2007) found a significant positive effect from introducing a computer-assisted learning program to elementary schools in India, arguing that the program directly improved learning and indirectly increased attendance by making school more attractive.

As demonstrated in Muralidharan et al. (2017) and Cristia et al. (2012), the overall literature on technology in education suggests that the provision of technology must be focused on pedagogical improvements to be effective in improving learning outcomes. A review of 45 studies examining the effects of technology interventions in developing countries finds that interventions solely focused on technology hardware do little to improve students' active learning and learning outcomes (Power, Gater, Grant, & Winters, 2014). An evaluation of the Rwandan government's efforts to introduce and expand the use of computers reveals similar results; evidence from classroom observations showed that the integration of computers in regular teaching had not been properly implemented in the large majority of the targeted schools. Furthermore, the authors found that teaching and learning was more teacher-centric in schools where the use of information and communication technology (ICT) was actively encouraged, leading to less time being allocated to students in ICT-enabled classrooms (Rubagiza, Were, & Sutherland, 2009).

Evidence suggests that when implemented with a strong pedagogical focus, technology interventions are more likely to produce positive changes (Power, Gater, Grant, & Winters, 2014). A case study of the Intel 'Teach Essentials' course in India, Turkey and Chile found that a proper pedagogical context was key to effective e-learning integration. Intel's program focused on training teachers to integrate e-learning technology across the curricula as a tool for learning and to design and implement project-based learning activities. Students interviewed during the course of the study spoke positively about new learning activities such as project-based work, which gave them a chance to collaborate, utilize multiple resources, and direct their own learning, as well as schoolwork that was more relevant to their lives outside school, making learning more meaningful. Teachers also demonstrated better understanding of student-centered teaching practices and of ICT knowledge and skills (Light, 2009).

The evaluation of Impact Network's eSchool 360 program stands to add a significant contribution to this body of literature given the proposed rigorous evaluation design and supplementation with context-rich qualitative research. Because the eSchool 360 program incorporates the provision of e-learning technologies with a strong focus on teacher training and pedagogical improvements, the evaluation will provide an excellent comparison point for the growing literature that suggests ICT interventions in education must include deep pedagogical focus in order to be effective in improve learning outcomes. Additionally, the setting of the eSchool 360 program in rural Zambian community schools provides new cultural, geographic, and administrative context; this, especially in conjunction with the qualitative research, will be invaluable in understanding the generalizability of the evidence surrounding a popular and growing type of intervention throughout the developing world.

Program Description

eSchool 360, implemented by Impact Network, delivers low-cost education to children in rural communities through a holistic solution. eSchool 360 costs $3 per month per student, which is 70% less than the Zambian government spends (Winters, Schling, and Winters, 2013). The core of the model is e-learning technology whereby tablets and projectors are provided by Impact Network's partner, iSchool, and are loaded with a curriculum in the local language approved by the Zambian government. Impact Network supplements the technology by providing teacher training and professional development, providing electricity via solar power, and creating community ownership. Locally hired teachers receive weekly training focused on using the technology and enhancing their pedagogical skills. The approach represents a significant innovation not only because technology is used but also because it incorporates the practice of training local high school graduates to be teachers, and providing them with systematic, ongoing support.

Impact Network's schools are located in the Eastern Province of Zambia, in a rural area with no running water and limited electricity. Households with children in Impact Network schools are significantly poorer than students in government schools (Schling & Winters, 2015). The current eSchool 360 program affects 2,200 primary school–aged children and nearly 10,700 community members of students' families. In the next 3 years, Impact Network aims to serve an additional 7,000 students and benefit 34,650 related community members across 35 additional schools in the districts of Katete, Petauke, and Sinda. In the next 5 years, Impact Network hopes to reach 1 million students, and to benefit 5 million citizens across Zambia in partnership with the government.

Theory of Change

The theory of change underlying the program suggests that the eSchool 360 program may lead to improvements in learning outcomes through various mechanisms. First, the teacher professional development component of the program may lead to improvements in the knowledge and practices of untrained teachers, which may in turn 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 may in turn lead to improvements in pre-literacy, early grade reading and early grade math outcomes. Second, the infrastructure improvements in the community schools may lead to increases in the demand for education, which may in turn result in increases in education enrollment and attendance. The infrastructure improvements may also result in decreases in the age-at-enrollment of Zambian students. These improvements in school attendance and enrollment may then result in increases in the time spent on education, which may lead to improvements in learning outcomes.

In addition to the improvements in learning outcomes, the program may also result in improvements in students' and parents' aspirations. Improvements in the quality of education may increase students' and parents' expectations about their future. These increased expectations may lead to higher aspirations in the domains of education, the labor market, and family outcomes. For example, parents may increase their expectations about the likelihood that their children will be able to finish 12th grade. In addition, the improved quality of education may result in increased expectations about the returns to education, which may in turn lead to higher expectations about labor market outcomes. Finally, increased aspirations in the education and labor market domains may result in increases in expectations about the marriage prospects of students as well as their age-at-marriage.

The validity of the theory of change depends on several assumptions. Perhaps most importantly, teachers need to comply with the e-school programming. In addition, the community schools need to have sufficient capacity to implement the program. Furthermore, locally selected teachers need to have the right incentives to provide quality education. The language of instruction also needs to be consistent with the student population's needs.

The effects of the program may also vary with several individual-level, household-level, and community-level moderators. For example, the effects of the program may vary by gender, language, age, and socioeconomic household-level characteristics. In addition, the program may be less effective in improving school attendance and enrollment for students who live further away from the Impact Network schools. Furthermore, the program impacts may be moderated by baseline pre-literacy, reading, and math outcomes, as well the mother and father's education levels. We will test each of these potential heterogeneities in the impact evaluation.

Ultimately, the goal of our evaluation is to inform how the eSchool 360 program can be effectively scaled up in Zambia. Achieving this goal requires a combination of different research and evaluation approaches to guide an iterative program design, where the implementing partner uses each evaluation finding to reflect on and, if needed, refine the program design. Such an approach incentivizes the use of evaluation findings, which in turn increases the value for money of the evaluation. We highlight this adaptive programming approach to guiding the scale-up of the program in the conceptual framework in Figure 2. We build on the work of McClure and Gray (2015), who created a framework to explain which factors contribute to the effective scale-up of development programs. They emphasize the importance of defining the "big-picture goal" of a program before determining the strategy underpinning the scale-up model. They also highlight the importance of scaling up using an iterative process, continuously learning and adapting the program as needed as the scope of the innovation expands. This iterative, evidence-driven approach requires effective evaluation.

Figure 1. Theory of Change

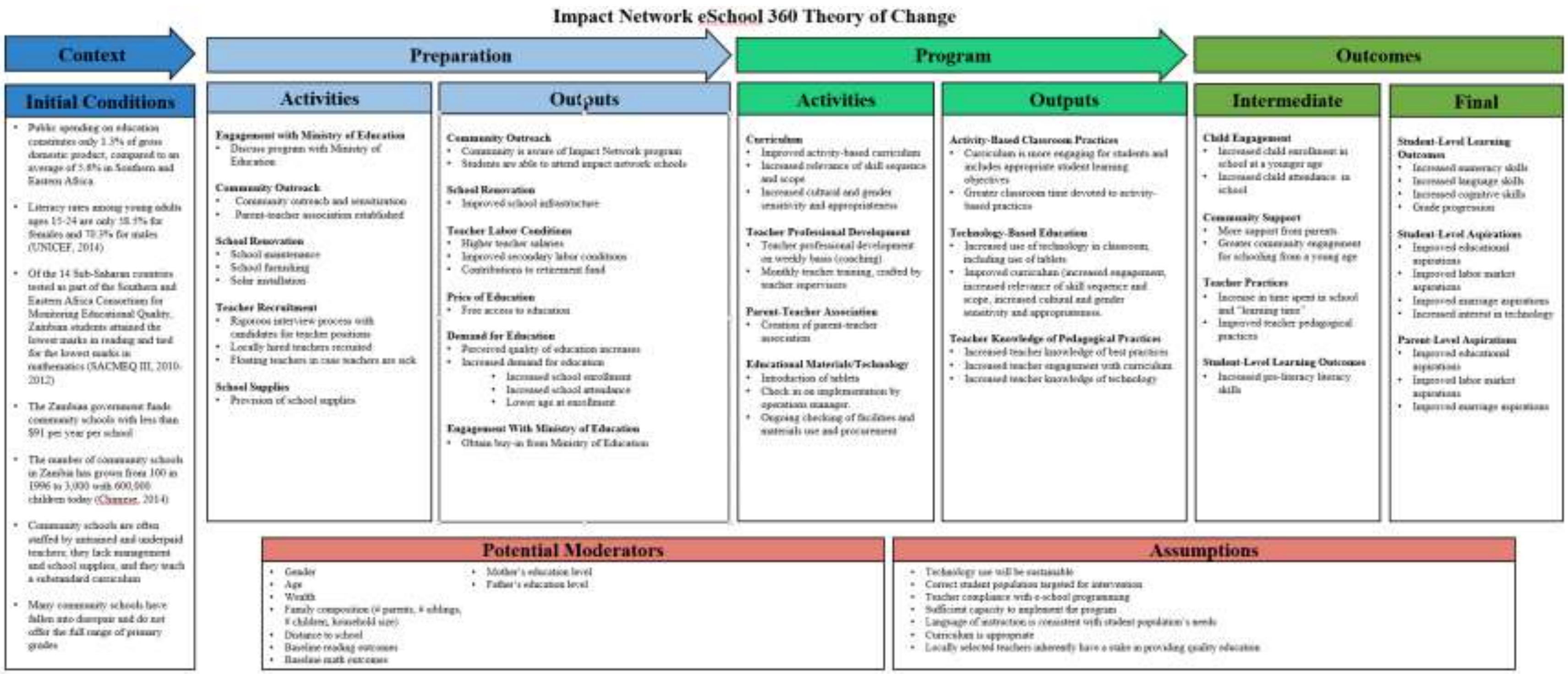
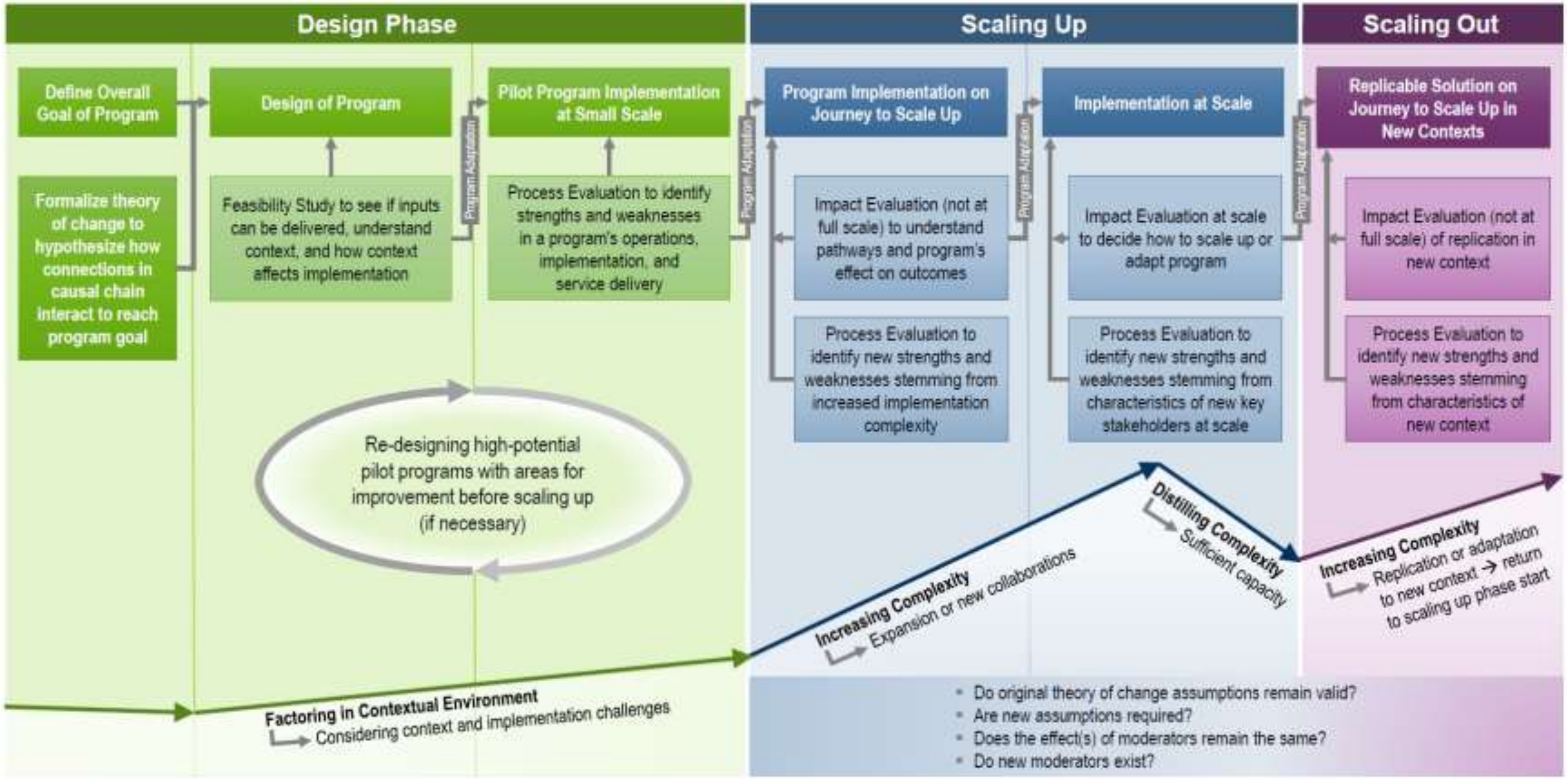


Figure 2. Scaling Frameworks



Research Questions

The proposed study 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. Each component is designed to answer different, but complementary, questions:

Impact Evaluation

a. What is the effect of eSchool 360 on students' numeracy, pre-literacy and literacy skills?
b. Do students enrolled in the eSchool 360 program improve in numeracy and literacy skills?
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 parents' and childrens' aspirations?

Cost-Effectiveness

f. How cost-effective is the eSchool 360 program in improving literacy outcomes?
g. How cost-effective is the eSchool 360 program in improving math outcomes?

Process Evaluation

h. Was the 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?
i. How does the program implementation vary by geography, culture, and time of year?
j. Do perceptions of the quality of teachers differ among students, parents, teacher supervisors, and teachers? If yes, how?

Study Design

We will design and implement a two-stage mixed-methods evaluation design that starts with quantitative baseline data collection and then follows with quantitative and qualitative data collection to inform the scale-up of the program 1 year and 3 years after the start of the baseline data collection. Impact evaluations of programs with an emphasis on innovations in education often rely on quantitative designs without triangulating the results with qualitative methods. As a result, it remains unclear how and why programs influence education outcomes even if programs are effective, which limits the learning potential of impact evaluations. In their assessment of the explanatory power of two RCTs of education programs, Burde (2012) argues,

> "When properly designed and executed, randomized trials can produce robust and significant findings even in the most difficult circumstances. Had they relied exclusively on quantitative methods, however, the studies discussed here would not have fared as well in explaining why these programs had the impact they had. Mixed methods enhance explanatory power for studies that explore impact and cause-and-effect questions."

In addition to the quantitative and qualitative impact analyses, we will conduct a cost-effectiveness analysis. The use of cost-effectiveness analyses will enable us to estimate the cost of achieving certain benefits, such as improvements in learning outcomes. These estimates will in turn guide policy makers in assessing the value for money of investing in the eSchool 360 program.

Quantitative Study Design

To determine the impact of the program we will use a cluster-randomized, controlled trial (cluster-RCT) in which 64 eligible schools will be randomly assigned to either receive Impact Network's eSchool 360 program ("treatment schools") or not receive the program ("control schools"). A well-designed and well-implemented cluster-RCT permits researchers to make causal statements about the impact of a program; if the randomization is valid and other conditions are met, any differences observed between the treatment and control students or households are directly attributable to the program (Duflo, Glennerster, & Kremer, 2007).

Randomization Across Eligible Schools

The cluster-RCT evaluation of the eSchool 360 program involves randomly assigning the program among schools that satisfy Impact Network's geographic, infrastructure, and organizational structure eligibility criteria for the eSchool 360 expansion. The geographic criteria arise from Impact Network's goal of introducing 34 community schools across three districts in Zambia's Eastern Province: Petauke, Sinda, and Katete. Among schools in these areas, Impact Network sought schools that had sufficient infrastructure to accommodate the eSchool 360 program; in other words the school had to have a dedicated physical structure. Impact Network also selected schools that were largely informal, and decided to implement the program only in schools that had more community teachers than government teachers. In addition to these criteria, the evaluation imposed one additional geographic eligibility criteria: pairs of eligible schools that operated in close proximity to each other (within 3 kilometers) were excluded to reduce bias from spillovers or contamination.

To determine which schools met all of the above eligibility criteria, Impact Network collected data from all community schools in the region; this process enabled Impact Network to identify 64 eligible schools.[1] Impact Network and AIR started this process by closely consulting with Zambian government officials to obtain a list of all community schools. Impact Network staff then visited each of these schools to obtain information about the structure of the school, the number of government and volunteer teachers, the state of the infrastructure, the grades served, and the distance to other schools. In addition, the Impact Network went well beyond this process. They were able to identify several community schools that were not on the official list provided by the government.

Following the selection of eligible schools, AIR oversaw the process during which representatives from local ministry of education departments in each of the three districts (Petauke, Sinda, and Katete) randomly assigned 30 of the eligible schools to the eSchool 360 program and 34 schools to the control group. To achieve this goal, it was important to obtain letters of approval from ministry officials. The participation of the local ministry of education officials in the randomization will encourage buy-in to the evaluation by the Zambian government, which in turn will enable us to help maintain the fidelity of the randomization and the credibility of the study. Finally, this process also encourages government officials to be invested in and knowledgeable about the evaluation. These processes will all contribute to an increased likelihood of a successful scale-up of the eSchool 360 program if the evaluation of the eSchool 360 program shows evidence for a cost-effective, quality education for students who attend community schools.

Sampling

To identify the sample of children with the potential to be affected by the program and determine intention-to-treat effects, we will conduct a census in the areas surrounding the sample schools to identify all households with children who are eligible for enrolling in first grade from January 2018. Children eligible for first grade are those who are 6 years old or older in January 2018 and who did not attend first grade in the prior school year.[2] We will identify all households with children eligible to enroll in first grade within a diameter of 1.5 kilometers of the schools, because conversations with local experts suggest that children eligible for first grade generally do not walk more than 1.5 kilometers to school.

It is important to identify intention-to-treat effects, because the announcement of the treatment schools and the construction of the additional infrastructure was visible for the beneficiaries. The randomization of schools to treatment occurred almost 7 months before the first evaluation cohort will be admitted to the sample schools (in January 2018). This visibility may result in a different composition of students in the treatment schools relative to the control schools—for example, by

[1] Impact Network staff had the impression that they had originally identified 65 Impact Network eligible schools. However, further analysis suggested that the 65th school was too far away from the Impact Network offices to consider for the program. This school had originally been assigned to the treatment group. However, because it was too far away, we decided to replace the school with a randomly selected school from the control group in the same district.

[2] Because only a small sub-sample of 6-year-olds enrolls in first grade, we will, however, exclude 6-year-old children from our sample. This will increase the evaluation's statistical power to detect small but meaningful program effects.

influencing school enrollment and attendance. In fact, a previous nonexperimental evaluation of the Impact Network program demonstrated that the introduction of Impact Network schools may have resulted in a reduction in the age-at-enrollment in community schools (Schling & Winters, 2015). Such changes in the composition could result in a bias in impact estimates that compare students enrolled in the Impact Network schools with students who are enrolled in other community schools because of differences in either observable (e.g., age, gender, parents' education level) or unobservable characteristics (e.g., motivation, noncognitive skills).

To address concerns regarding these composition effects and enable the estimation of intention-to-treat effects, we will sample 30 households from the census-generated sample frame for each of the sample schools. For households with more than one eligible child, we will select the oldest child for inclusion in the sample. Thus, we expect to have a sample of 30 children from the area surrounding each of the 30 treatment and 34 control schools, for a total of 1,920 children. We will sample more than 30 children if feasible in anticipation of the possibility that some villages may not have 30 households with eligible children. This will ensure that we maintain a sufficient sample size throughout the study.

To increase statistical power we will use stratified random sampling by age. We will not include 6-year-olds in our sample because descriptive statistics suggest that 6-year-olds are very unlikely to enroll in first grade; we will also over-sample 8- and 9-year-olds, because descriptive statistics indicate that these children are more likely to enroll in first grade than 7-year-olds (Ministry of General Education, 2014). Of the children enrolled in first grade in the Eastern Province in 2014, 4,484 were under 7 years old, 23,206 were 7 years old, and 35,220 were 8 years old or older. To increase statistical power, we will use the same age distribution in our sample (after excluding 6-year-olds), while assuming that 8-year-old and 9-year-old children will enroll in first grade at an equal rate. We present this sampling strategy in Table 1 below. Oversampling groups that are more likely to enroll in school will increase the expected effect size on learning outcomes, which will in turn increase statistical power.

Table 1. Oversampling of 8-and 9-Year-Olds

Age Category	Percentage of Children in First Grade	Sample Size in Impact Evaluation of eSchool 360 Model
7-Year-Olds	39.72%	763
8-Year-Olds	30.14%	579
9-Year-Olds	30.14%	578

To maintain the ability to estimate intention-to-treat effects for 7, 8, and 9-year-olds we will conduct analysis with and without sampling weights. In our analysis with sampling weights, we will weigh the sample by using post-stratification weights. This will bring the sample back to being representative of the population by gender and by likelihood of enrolling in community schools. Importantly, however, the use of post-stratification weights may reduce statistical power if 7-year-olds are less likely to enroll in first grade than 8- and 9 year-olds. For this reason, we will conduct analyses with weights and sub-sample analyses by age category. The analyses with weights will enable us to estimate intention-to-treat effects for the full sample, and the sub-sample analyses will enable us to estimate intention-to-treat effects by age category.

The study will use a longitudinal panel design that follows each of the sampled children for 3 years. We will follow the children regardless of where and when they enroll in school to estimate intention-to-treat effects of the program on school attendance and enrollment as well as pre-literacy, literacy, and numeracy outcomes.

We will sample the households from the census in the field through SurveyCTO software. This software will be used to collect data on tablets. The software will store basic data from each of the eligible households through the census we will conduct in the areas surrounding the sample schools. In addition, SurveyCTO has built-in features to randomly sample households by category.

Data Collection

Quantitative Data Collection

AIR will partner with Palm Associates, a Zambian research organization that specializes in data collection and social science research, to collect data for the evaluation. AIR has worked successfully with Palm Associates in Zambia on two longitudinal RCTs of cash transfer programs, a longitudinal quasi-experimental evaluation of a nutrition program, an RCT of a condom distribution program, and a performance evaluation of a water and sanitation program. Palm collected data for the nutrition and condom distribution studies in the same districts as this Impact Network evaluation, so the organization's staff are familiar with the geography and culture. AIR will work closely with Palm to train enumerators before each round of data collection and will then follow them into the field to observe data collection. Palm uses Zambian enumerators who speak the local language of the areas included in the study and who are familiar with the assessments that we will implement. AIR will also help build the capacity of Palm's staff to conduct high-quality research, thus empowering Zambian researchers to help grow the country's ability to generate its own evidence for decision-making—an effort closely aligned with AIR's mission.

Palm will use tablets to collect data, thus improving the quality of data collection, minimizing the need to clean data, and eliminating data entry. As discussed above, we will collect the data on tablets running SurveyCTO. The use of SurveyCTO will minimize errors in the field because skipping patterns can be automated and the software allows for built-in checks to ensure the quality of data. For example, we can ensure that children's age and date of birth will be consistently entered in the database. The Survey CTO software runs on the Open Data Kit (ODK) platform. ODK enables users to collect data on a tablet and send it to a server, then aggregate the collected data and extract it in Stata format. Data coming in from the field will be examined weekly by the AIR team. We will develop Stata "dofiles" with high-frequency checks to examine the most common errors and discrepancies. These Stata dofiles will identify enumerators with data that are systematically different from their colleagues' (for example, different in the amount of time it takes to administer the survey and assessment but also in the survey and assessment data collected). In addition, we will closely examine GPS coordinates to ensure that enumerators work in their assigned enumeration areas.

Outcome Measures

Quantitative Indicators

The primary achievement indicators for numeracy, pre-literacy, and literacy will come from the EGRA, EGMA, and Zambian Achievement Test (ZAT) that have already been adapted and used in the same region as this study. These instruments have all been translated and were validated in the context of Zambia. We will collect assessment data in Bemba because this is the language of instruction in Grade 1-3 in Eastern Zambia. We will also collect enrollment and attendance data through the household survey. These data will also be validated through the use of attendance-data from the treatment schools. Last, we will collect outcome indicators on parental and community perceptions of school, teachers, and their children's education through a household survey. Specifically, we will collect data on parents' and children's educational, labor market, and marriage aspirations.

In addition to these outcome indicators, the study will also collect control and moderating variables at the student, household, and school level including:

- Student-level: gender, age, and orphan status
- Household-level: distance from school, poverty level, parents' education level, and household size
- School-level: size, number of teachers, experience in years of teachers, age of teachers, and average class size.

Outcome Measures Related to Learning

We will field a consistent assessment instrument across the midline and endline to measure the impact of the Impact Network program on student outcomes 1 year and 3 years after the introduction of the Impact Network schools. The primary cognitive skills outcomes are aggregate scores on the early grade reading assessment (EGRA), early grade math assessment (EGMA), and Zambian achievement tests, and the secondary outcomes are the EGRA, EGMA, and Zambian achievement test subtasks as well as measures of oral reading fluency. As students will be learning different concepts at different ages, we expect to see different impacts on subtasks at the midline relative to the endline.

We will use different secondary outcome measures for the midline and endline surveys. At the midline, the secondary literacy outcome measures will be the four emergent literacy EGRA subtasks: concept of print, oral vocabulary, phonological awareness, and decoding. At the endline, the secondary literacy outcome measures will be reading comprehension and oral reading fluency. Similarly, for EGMA, the secondary math outcomes at midline will be oral counting fluency, one-to-one correspondence, number identification, quantity discrimination, and the time to complete each of these sections, while at the endline, the secondary math outcomes will be the missing number, addition and subtraction, and geometry sections, and the time to complete each of these sections. Below we define each of these outcome measures. Much of the discussion on the early grade math assessment constructs is based on RTI (2009).

Concept of print: This will measure whether children understand how print "functions"—how to hold a book, where the beginning and end of a book is, and so forth.

Oral vocabulary: This task will measure receptive oral language skills separately from any decoding/script-processing ability. In this test, children will see four pictures, will listen to the data collector say the name of one picture, and will be asked to point to the correct picture in their test booklet.

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

Decoding: Decoding measures the ability to sound out words when they are seen in print. Children need 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: We will measure reading comprehension by using a short passage with literal and basic inferential comprehension questions.

Oral reading fluency: We will measure oral reading fluency by counting the correct words read per minute.

Oral counting fluency: We will measure oral counting fluency by assessing children's ability to produce numbers fluently. The task usually begins with the number 1, and asks children to continue counting until they reach the highest number they can before making a counting error (Floyd et al., 2006).

One-to-one correspondence: We will measure one-to-one correspondence by assessing the extent to which children are able to recognize the items they need to count and by assessing the extent to which children are able to recognize, and mentally tag, those items that have already been counted.

Number identification: We will measure number identification by assessing the extent to which children are able to orally identify printed number symbols that are randomly selected and placed in a grid.

Quantity discrimination: We will measure quantity discrimination by assessing children's ability to make judgements 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: In this task children will be asked to name a missing number in a set or sequence of numbers. We will present children with a string of three numbers with the first, middle, or last number in the string missing. We will then ask the children to report what number is missing.

Addition and subtraction: We will measure addition and subtraction skills by presenting children with oral or written problems with a focus on addition and subtraction. Children will be shown a visual representation of the mathematics problem, and also have the problem read aloud to them.

Geometry: This task measures whether children are able to recognize shapes. An interviewer will ask the child to identify and point to all representations of one shape on an 8 ½ x 11-inch sheet of paper. At the end of the assessment, the interviewer will base the score on the number of correct shapes and incorrect shapes that were marked.

Previous experience in Zambia suggests that the large majority of the students will score very low on their EGRA and EGMA tests in 2018. Furthermore, it is likely that any improvements by endline will only be small because of the difficulty of the EGRA and EGMA 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 will include the pre-reading recognition subtest of the Zambian achievement test as a complementary test of student pre-literacy skills. The ZAT assessment, developed by the PACE Center (New Center for the Psychology of Abilities, Competencies, and Expertise) and EgLab (which is part of the Child Study Center at Yale School of Medicine) 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 Pre-Reading Skills subtest consists of 34 items and was constructed so that children simply need 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 Zambian achievement test is less vulnerable to floor effects when implemented in the Eastern Province of Zambia.

To mitigate concerns about floor effects it will still be important to add additional outcome measures. We will therefore also estimate impacts on oral reading fluency. We have created a short oral vocabulary subtask to enable these impact estimates. This test will measure the spoken language skills of children. Estimating impacts on oral reading fluency is less vulnerable to floor effects.

Factor Analysis

In addition to estimating impacts on EGRA and EGMA aggregate scores, we will also conduct factor analysis to understand what latent constructs the assessments are tapping into. In educational assessment, certain abilities, such as math skill, logical reasoning, and reading ability, are posited to be latent constructs. The existence of these constructs must be demonstrated through the accumulation of behavioral or performance evidence that supports that claim. Data collected from the EGRA and EGMA administrations will provide AIR with the opportunity to conduct empirical analyses (factor analyses) of the underlying internal data structure of the subtasks. The primary purpose of factor analysis is to determine the number of distinct dimensions or constructs (also referred to as *factors*) that theoretically underlie a domain of knowledge, trait, or ability measured by an assessment or survey instrument (Kim & Mueller, 1978). Although EGRA and EGMA have been validated for the context of Zambia, it will be useful to examine whether the factor structure will be different for the population of children who are eligible for Impact Network schools.

For this purpose we will employ *principal axis* factor analysis. The factor analyses will be conducted with subtask results by assessment test to assess the dimensionality of the entirety of each assessment battery. Factor analysis interpretation will be guided by examining factor loadings in a rotated factor matrix. Based on previous research on reading and math ability, it is plausible that our underlying factors of interest will be correlated; thus, oblique (Oblimin) rotation will be selected. The resulting pattern matrices will allow interpretation of the overall structure of the data by examining how factors are clustered on the matrix. High factor loadings (above .4) can indicate which subtasks are tapping into which common dimensions.

Comparison With Outcomes in Government Schools

In addition to the cluster-RCT, we will also compare 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 will randomly select these government schools from a sample of government schools in the three districts. The comparison with government schools will not enable us to estimate the impact of enrolling in Impact Network schools versus enrolling in government schools. Nonetheless, the comparison will serve 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.

Other Outcome Measures

In addition to the EGRA, EGMA, Zambian achievement test, and oral reading fluency measures, we will also estimate the program impacts on school attendance and enrollment, parents' perceptions of school and education quality, student-level aspirations, and parent-level aspirations. Using these outcome measures will enable us to determine impacts along the causal chain of the theory of change and let us examine the mechanisms underlying the program impacts.

School enrollment and attendance: We will measure school enrollment and attendance by relying on self-reported student data. Specifically, we will ask parents whether their children are enrolled in school and ask children how many days they attended school in the week before the survey. We will measure impacts on school enrollment and attendance using a single-difference model (without controlling for the baseline value of the outcome of interest) because our baseline survey will not include measures of school enrollment and attendance. During the baseline we will ask parents whether they expect to enroll their children in 2018. However, we will not use this variable as a control variable because it could be affected by the announcement of the treatment schools and the construction of the additional infrastructure.

Parents' and children's perceptions of school and education quality: We will measure parents' perceptions of school and education quality by asking parents four-point Likert-scale questions related to the quality of education. These questions will relate to their general perceptions of 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 parents' perceptions about teachers' pedagogical practices. We will also ask children for their level of engagement in the classroom, the time devoted to activity-based learning activities, the use of technology in the classroom, and their interaction with teachers.

Student- and parent-level aspirations: We will also measure students' and parents' aspirations with respect to education, marriage, and labor market outcomes. To measure student aspirations, we will ask students about the level of education they would like to achieve, and the age at which they would like to get married. To measure parent aspirations, we will ask them for the bride price they expect to receive for their daughter, the probability they assign to their child graduating from grade 12, and their child's preferred age-at-marriage.

Power to Detect Effects

Power calculations suggest that interviewing 30 children in each of the 64 school catchment areas will be sufficient to detect small but meaningful effects on school enrollment and attendance, on various intermediate outcomes, and on the learning outcomes of children enrolled in Impact Network schools. On the basis of our sample of 1,920 children in 64 school catchment areas, we would have an 80% chance of detecting an intention-to-treat effect of 0.247 SMD on school enrollment and attendance and learning outcomes when we assume an intraclass correlation of 0.15 and an R squared of 0.4. Das, Dercon, Habyarimana & Krishnan (2007) show that average learning during the year in both English and mathematics was approximately 0.4 standard deviations. Hence, we are powered to detect impact estimates that are equivalent to what students learn in approximately 0.6 school years. The values of the intraclass correlation and the R squared are based on AIR's experience with impact evaluations of similar interventions and an analysis of the data from the previous evaluation of the Impact Network program in Zambia (Schling & Winters, 2014). These power calculations also assume a 10% attrition rate, though we believe this to be a conservative assumption based on our plan to do intensive tracking of sample members for data collection, including visits to their home, and low overall mobility in the study areas.

We will also have sufficient statistical power to detect intention-to-treat effects by age category. Power calculations indicate that we would have an 80% chance of detecting an intention-to-treat effect of 0.289 SMD on 7-year-olds' outcomes when we make the same assumptions as discussed above and assume that 39.72% of the sample consists of 7-year-olds. We would also have an 80% chance of detecting an intention-to-treat effect of 0.308 SMD on 8-year-olds' or 9-year-olds' outcome measures assuming that 30.14% of the sample consists of 8-year-olds and 30.14% of the sample consists of 9-year-olds. The blocked randomization by district and the use of analysis of covariance (ANCOVA) should further increase our statistical power beyond these levels.

Cluster-RCT Analyses

Research Question 2 focuses on the impact of the Impact Network program on a variety of student and parent outcomes. The main impact analyses will be conducted as intention-to-treat analyses, which will measure the impact of living within 1.5 kilometers of an Impact Network school on student academic outcomes, school attendance, and perceptions of quality, regardless of whether the student chooses to attend the Impact Network school. In this chapter, we describe the analytic samples and describe the statistical models that will be employed for the impact analyses. We also describe how we will handle attrition, and our approach to testing of multiple hypotheses in the analysis.

Baseline Balance

Following the collection of baseline data, we will compare mean values of demographic and baseline cognitive skill measures across the treatment and control school catchment areas to confirm their comparability. As we will be testing a variety of measures, we expect that some will be significant as a result of randomization with finite numbers. In addition to ordinary least squares (OLS) regression analysis, we will also calculate the normalized difference to examine balance. In keeping with the work of Imbens (2015), we will control for any variables that have a normalized difference of more than 0.25.

Impact Analysis Sample

Our impact evaluation research design includes a baseline survey and two rounds of post-intervention data collection: a midline survey and an endline survey. Our main evaluation sample will include all respondents who are surveyed at each round; the midline analysis will include all individuals that are surveyed at the midline, and the endline analysis will include all individuals that are surveyed at the endline. Following the endline, we will also produce analyses that restrict attention to the sample of respondents who are surveyed at all three rounds: the baseline, midline, and endline. This supplemental analysis set will allow us to track program impacts across time using a consistent sample.

Statistical Models

The proposed 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. The probability of assignment to treatment is orthogonal to individual characteristics after controlling for stratum fixed effects. Thus, the direct effect of treatment (residing in an Impact Network school catchment area) on outcome Y_i can be estimated using the regression specification below:

$$Y_{it} = \alpha + \beta * IN_i + \delta * S_i + \sigma * Y_{it-1} + \mu * C_i + \varepsilon_i$$

Here IN_i is an indicator variable equal to one if individual i resides in the catchment area of an Impact Network school and equal to zero otherwise, S_i is a vector of dummies for the anticipated district strata, Y_{it-1} 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. Since treatment is randomized within strata, the inclusion of the Y_{it-1}, S_i, and C_i variables should increase efficiency but not impact the estimated value of β. We will use cluster-robust standard errors clustered at the school level to account for potential correlation in outcomes within a school catchment area.

We will use an ANCOVA model to estimate the intention-to-treat effect of the program. The ANCOVA approach uses a regression specification that includes the baseline measures of outcome variables as an additional explanatory variable. This empirical approach can improve statistical power by exploiting information and variation contained in the baseline data (McKenzie, 2012). In other words, the use of ANCOVA increases the likelihood that statistically significant effects will be detected if the program indeed causes statistically significant effects. The intention-to-treat effect is the impact of the intervention on those households that are targeted by the intervention regardless of whether these households participate in the intervention.

Correction for Multiple Comparisons

To address the potential inflation of Type I error and statistical significance owing to multiple comparisons, we will apply corrections for multiple comparisons to multiple outcome measures within the same outcome domain using the Benjamini-Hochberg method, as recommended by the What Works Clearinghouse and employed by Banerjee et al. (2015). The outcome measures for the impact analyses will be organized into four domains: primary EGRA outcomes, primary EGMA outcomes, EGRA subtask outcomes, EGMA subtask outcomes. The primary EGRA and EGMA outcomes are single measures and will not be corrected for multiple comparisons. The EGRA and EGMA subtask outcomes are related within the domains, and the results will be corrected for multiple comparisons using the Benjamini-Hochberg method. When reporting study findings, we will note both the statistical significance after correction for multiple comparisons (q-values) and provide the uncorrected p-values so that readers can apply their own corrections as they see appropriate.

Treatment Heterogeneity

The theory of change identifies a number of potential moderators to the impact of the Impact Network program. We will test for whether the program has a differential impact by gender, region, socioeconomic status, mother's education level, and baseline EGRA/EGMA attainment. For non-categorical moderators, we will construct a binary variable equal to one for individuals with a value greater than the median and equal to zero for those with a value less than the median. We will then interact the binary variable with treatment to test for whether the treatment is statistically different for groups with high and low levels. For socioeconomic status, we will construct an asset index using the calculated values from the first principal component of a list of assets as recommended by Filmer and Pritchett (2001).

Attrition

For the midline and endline analysis, we will present information about both overall and differential attrition rates, by treatment status, for students. We do not expect high overall rates of attrition given AIR's previous experience on another project in rural Zambia that tracked 98% of respondents for a 3-year follow-up (AIR, 2014). If the rate of attrition is statistically different across students in the treatment and control groups, we will apply the Lee (2009) bounds correction and report both the original and corrected impact estimates.

Treatment Effects on the Treated

In addition to intention-to-treat effects we will also estimate treatment effects on the treated by using the program assignment as an instrumental variable for self-reported school attendance in Impact Network schools. The treatment effect on the treated is the impact of the intervention on those children who participated in the program. In our case we will define these children as children who self-report attending Impact Network schools at least three times in the week before the survey. In an additional analysis, we will define these children as children who have ever been enrolled in Impact Network schools. For the instrumental variable approach we will use two successive regressions, in which the first regression explains the treatment variable using the treatment assignment, and the second regression explains the outcome variables with the predicted treatment variable.

Cost-Effectiveness Analysis

To determine the cost-effectiveness of the program we will combine the benefits estimated in our impact analyses with the costs obtained from our implementation research to carry out a **cost-effectiveness analysis.** For this purpose, we will need to specify all of the ingredients that are necessary to replicate the program and then collect data on the unit costs of all of these ingredients (Dhaliwal, Duflo, Glennerster, & Tulloch, 2011). AIR will work with Impact Network to gather information on resources used for the intervention to create an exhaustive list of resources with costs. Using this information, AIR will create a cost database that contains basic descriptive information and, if the data are available, information to permit analysis of the patterns of variation of resources—for example, by geography or scale. We will then estimate the costs of the intervention for the average beneficiary and divide these costs by the expected gain in outcome derived from the impact analysis to serve as the cost-effectiveness measure of the intervention. We will also consider including opportunity costs for the beneficiaries in this cost analysis.

Performance Evaluation of SimPrints

In addition to the cluster-RCT and the cost-effectiveness analysis, we will examine the performance of attendance tracking for teachers and students through fingerprinting technology. Impact Network is partnering with an organization called SimPrints in setting up a fingerprinting system to track attendance for teachers and students in the 10 schools where Impact Network set up its pilot program. These 10 schools will not be part of the cluster-RCT. However, we will track teacher and student-attendance in these five schools in order to track the performance of the fingerprinting system. Previous studies suggest that monitoring attendance can be effective in improving teacher attendance. Duflo, Hanna, and Ryan (2012) show that combining the monitoring of teacher attendance with financial incentives reduced teacher absenteeism with 21 percentage points relative to the control group. Our analysis of teacher and student attendance will not allow for assessing the causal effect of the fingerprinting system. However, it should enable AIR and its partner to obtain useful information about the performance of SimPrints when the program would be implemented at a larger scale.

Qualitative Study Design

AIR is well versed in supplementing impact evaluations with qualitative research. Qualitative research is well suited to complement impact evaluations because it enables researchers to explore not only formal activities and anticipated outcomes but also informal patterns and unanticipated interactions, and it gives the researcher flexibility to explore unforeseen areas of interest. Incorporating a qualitative process evaluation into a larger impact evaluation allows researchers to explore how and why a given link in a program's theory of change is not working optimally. Qualitative research also gives researchers the opportunity to investigate why we do or do not see impacts and to explore the nuances and variation in our sample, particularly with regard to sensitive or complex topics. Qualitative data provide rich contextual information to complement the numerical findings of an impact evaluation while also allowing researchers to triangulate and explain quantitative findings. This section provides information about the qualitative component of the evaluation, including the instruments, methods, and timing for data collection.

District Selection and Sampling

Qualitative data collection will take place in two schools in each of the three treatment districts of Katete, Petauke, and Sinda. In each district the research team will work with Impact Network staff to select two schools to visit. The selection will be made based on a set list of criteria established by the research team and Impact Network. Illustrative sampling criteria include school size, distance from district center, perceived success in delivering eSchool 360 curriculum, perceived difficulty in delivering eSchool 360 curriculum, and high/low performing schools based on student learning outcomes from prior years. This selection will maximize our ability to capture both perceived strengths and weaknesses of the different components of the eSchool 360 program.

Data Collection Methods

We will use three primary approaches to qualitative data collection for this evaluation: key informant interviews (KIIs) with community leaders, eSchool 360 program staff, teachers, and students; focus group discussions (FGDs) with students and parents; and classroom observations. Table 2 provides a summary of the qualitative data collection approaches and respondents.

Key Informant Interviews

A key informant is a person who possesses expert knowledge about the program or a topic related to the program. KIIs with carefully selected eSchool 360 teachers, students, and school community leaders are an important piece of our evaluation design because they will shed light on perceptions of eSchool 360 quality both in terms of how the technology and teacher trainings have affected classroom instruction and how the program engages with school communities. In KIIs students and teachers will discuss the perceived impact of the program, and we will explore respondent beliefs about whether the program is being implemented as planned and students are being instructed as intended.

Focus Group Discussions

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 will create a social dynamic that encourages participants to reflect upon their opinions and experiences and express them verbally. From past experience, we expect to benefit from synergies in which contributions from participants inspire other participants to think about and discuss their own experiences. Focus group discussions are helpful for facilitating a deeper understanding of the factors enabling or inhibiting implementation of the eSchool 360 curriculum in the classroom. From FGDs we hope to learn about experiences that parents and PTA members have had in communicating with eSchool 360 staff, engaging in general with the program, and using the school in the evenings and on the weekends. We also hope to engage students in a participatory exercise to reflect on the techniques that teachers are using in the classroom as well as the friendliness of the school environment.

Classroom Observations

We will conduct formal classroom observations at two schools in each of the three treatment districts. Classroom observations will allow us to examine the extent to which teachers at eSchool 360 schools are effectively using e-learning technology and applying pedagogical techniques gained during eSchool 360 teacher trainings. Observations will also take note of the eSchool 360 environment, particularly in regard to its child-friendliness. Classroom observations will be conducted in three to four randomly selected academic classes at each of the six selected schools using a piloted and validated instrument.

Logistics of Data Collection

Two-person teams at each school will undertake data collection. Wherever possible, one field researcher will be responsible for interviewing or facilitating, while the second researcher will have primary responsibility for recording responses. The researcher will write down responses (in local languages where necessary) on the response sheets provided with each data collection instrument. The researchers will digitally record all KIIs and FGDs on portable digital recorders, using an external microphone where possible. The researchers will download recordings to field laptops each day, renamed according to an anonymized code system held in an encrypted Excel sheet and copied to external media for backup. At the end of each day, the field researchers will transcribe the handwritten field recording sheets to Microsoft Word documents, translating the material where necessary.

Researchers will use audio recordings to supplement and validate the written transcriptions and translations. They will assign all transcriptions new names according to the code system in order to ensure data and informant confidentiality. The data collection team leader will make supervisory visits to all field sites to ensure that the work is carried out in accordance with research plans, on schedule, ethically, safely, and completely. They will also read transcribed interviews to ensure high standards of data quality. Table 2 presents our proposed data collection plan for the six schools we plan to visit in each of the three treatment districts.

Table 2. Data Collection Plan

Method	Respondent(s)	Key Content Areas	Number per district	Total Number
FGD	Students	Classroom instruction, school environment, community engagement, eSchool 360 program perceptions	2	6
FGD	Parents/PTA members	Community engagement, perception of eSchool 360 quality	2	6
KII	Community leaders	Community engagement, perception of eSchool 360 quality	4	12
KII	Students	Classroom instruction, school environment, community engagement, eSchool 360 program perceptions	4	12
KII	Teachers	Teacher trainings, classroom instruction, professional development, eSchool 360 programmatic support	4	12
KII	Teacher supervisors, Operations managers, and director	Teacher training, classroom instruction, professional development, eSchool 360 programmatic support, perceptions of program impact	2–3	6–9
KII	Impact Network staff	Teacher training, classroom instruction, professional development, eSchool 360 programmatic support, perceptions of program impact	2–4	
	Classroom observations	School environment, classroom instruction	6–8	18–24

Analysis

AIR adheres to strict data analysis principles, regardless of the type of data being collected. All data from KIIs and FGDs will be coded and analyzed using the NVivo qualitative software program. Our team will create a preliminary coding outline and structure on the basis of the research questions, interview protocols, and memos of ideas that emerged during data collection. This coding outline will serve as the tool to organize and subsequently analyze the information gathered in the interviews and focus groups. The outline will be a living document that may be modified as new themes and findings emerge during data analysis. A list of definitions for the codes will accompany the outline to ensure that coders categorize data using the same standards. After inputting the raw data into NVivo, coders will select a sample of interviews to double-code, to ensure inter-rater reliability. The team will subsequently code the data into the structure. Using this coded data, the qualitative team will use grounded theory to identify themes, categories, and theories that emerge from the data and that confirm or refute the researchers' initial impressions. In other words, rather than basing the analysis on a hypothesis, the researchers will create concepts and categories based on the data, refining the concepts as they go along to eventually inform the overall findings. During this process of data reduction, researchers will characterize the prevalence of responses, examine differences among groups, and identify key findings and themes related to the research questions.

Ethical Considerations

AIR conducts rigorous ethical reviews though our Institutional Review Board (IRB) for all of our own internal research activities and provides this service for a variety of subcontractors and collaborators. AIR's IRB has conducted expedited and full board reviews of research involving human subjects for more than 25 years. AIR is registered with the Office of Human Research Protection as a research institution and conducts research under its own Federalwide Assurance. We will obtain full approval from the AIR IRB and from the Zambian ethical review board at the University of Zambia. The following outlines how we will obtain consent and maintain confidentiality.

Consent

We will inform participants that the information they share is confidential. We will also inform them that their participation is voluntary and that they can end their participation at any time or skip any questions they do not wish to answer. During the qualitative research we will obtain informed verbal consent from each participant after reading the consent form aloud. During the quantitative research we will obtain written informed consent after reading the consent form aloud. We will obtain informed consent through thumbprints if the respondents are illiterate.

Assurances of Confidentiality

AIR handles all data in accordance with the procedures and protocols approved by our IRB. Standard practices include digital recording, transcription and translation where necessary, complete anonymization of data, and protection of confidentiality.

The study will protect confidentiality by a number of methods. First, we will not identify any individual household or member by name in any report or publication about this study. We will not share specific information about a household with anyone outside the research team. We have developed data handling procedures to safeguard completed forms. Each participant will be assigned a unique identification code that we will use to link participant records across modules and survey rounds. After we enter the data, we will encrypt and password-protect the complete data file.

We will develop an anonymized data set, stripping away any identifying information, and we will use this anonymized data set for all analyses. We will keep these identification numbers and associated names on a master file which will only be accessible to the Project Investigators (PIs) at Palm Associates and AIR. The PIs will save the electronic file on their computers and will protect the file with a password so that it is accessible by only the PIs. The team will analyze data collectively so that information from any one participant remains anonymous. We will also ensure that study staff members are trained to understand ethical research.

Communication and Dissemination Plan

High-quality research with concrete policy recommendations is a necessary, but not sufficient, condition for policy impact. To achieve maximum policy impact, it is important to engage with the Zambian ministries from the beginning of the evaluation. To achieve this goal, we will work closely with Impact Network while maintaining our independence.

AIR will create a communication plan that will describe the agenda for disseminating the research and informing evidence-based policy. This communication plan will include planning for events related to the research, blogs, social media activities, podcasts, policy briefs, presentations at academic and policy conferences, presentations in Zambia, and academic papers to disseminate the research findings to policy makers and researchers. Importantly, AIR will ensure that the dissemination of the research findings will be customized to the audience. For example, we do not expect Ministry of Education officials to read long impact evaluation reports or peer-reviewed papers. Thus, we will emphasize the use of policy briefs and presentations in the dissemination of our research findings to these policy makers.

Perhaps most importantly, however, we will support Impact Network in its activities to closely engage with the Zambian Ministry of Education. Impact Network is working to join all relevant working groups related to education and technology in Zambia in the coming year. Currently, we plan to present at the Projects Coordinating Committee (PCC), a coordinating committee for all non-governmental and civil society organizations that work with the Ministry of General Education. Impact Network also aims to join or chair the information and communication technologies (ICT) subcommittee of the PCC in the coming years. Further, Impact Network is in the process of joining the Zambia National Education Coalition, where Impact Network can further disseminate results as they become available. Last, Impact Network will be reaching out to Ministry contacts in order to join any other relevant working groups, and participate in other annual meetings, such as the Joint Annual Review.

Through Impact Network's contacts at the Ministry of General Education and funding community in Zambia, AIR aims to disseminate the research findings across the Ministry of General Education and multilateral funders through a variety of small group meetings. Impact Network staff perceive this type of briefing as more beneficial than large group meetings.

Capacity Building

In addition to the dissemination of the findings, the evaluation will also serve to build the capacity of local researchers from Palm Associates. AIR believes that it is important to work with colleagues in developing countries to conduct high-quality research and build the capacity of local institutions and researchers. AIR's collaborative process serves four purposes: it (1) gives local researchers opportunities to present research findings; (2) expands exposure to policy makers to increase local researchers' reputation as experts; (3) helps provide legitimacy to the findings through local context and contribution; and (4) helps policy makers engage with researchers from the start of the research. The local researchers will provide legitimacy and context and contribute to the data collection, analysis, and reporting.

Work Plan

Work for this evaluation commenced in April 2017, and will continue through the third quarter of 2021 with the dissemination of the final results and conclusion of the sustainability activities. Impact Network will need the first year to retrofit schools, hire staff, and train teachers in preparation for the first full year of implementation in 2018. AIR will use this time to finalize the design of the study and instruments, select the sample, and conduct baseline data collection from October through December 2017. The primary cohort for the study will start first grade in January of 2018. We will collect our first follow-up data at the end of 2018 after this cohort has completed a year of the program. We will follow these same children for 2 more years, and collect the final data at the end of 2020, after they have completed third grade (for those who stay in school). We will track children throughout the study regardless of whether they remain in school or not and will collect data at their house if they are no longer in school. After finalizing each report, we will disseminate the results through a series of meetings, presentations, and conferences, the most important of which will be our engagement with the Zambian government and large donors in the country as part of the sustainability activities. Each report will serve as an opportunity to educate these important stakeholders about the program and the evidence generated from the study. The performance evaluation of the SimPrints program will occur in 2018 and 2019 as the program scales up to the 10 original impact schools. Figure 3 shows the Gantt chart with the full work plan.

Figure 3. Work Plan

Task	2017										2018				2019				2020				2021	
	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2
Random assignment																								
Seek AIR Institutional Review Board (IRB) approval																								
Develop theory of change																								
Develop qualitative research instruments for process evaluation																								
Develop survey materials																								
Develop inception report																								
Obtain local IRB approval																								
Translate survey materials																								
Pilot survey tools baseline survey																								
Collect quantitative data, baseline																								

Task	2017										2018				2019				2020				2021	
	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2
Quantitative data analysis, baseline																								
Deliverable 1: Baseline report writing																								
Develop policy brief for baseline																								
Develop blogs for baseline																								
Revise quantitative survey instrument for midline																								
Translate qualitative research instruments for process evaluation																								
Train qualitative team for process evaluation																								
Midline quantitative data collection																								
Collect cost data																								
Collect qualitative data for process evaluation																								

Task	2017										2018				2019				2020				2021	
	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2
Transcribe qualitative data for process evaluation																								
Deliverable 2: Midline report writing																								
Develop policy brief for midline																								
Develop blogs for midline																								
Develop articles for peer-reviewed journals																								
Revise survey materials for endline data collection																								
Endline data collection																								
Deliverable 3: Endline report writing																								
Conduct cost-effectiveness analysis																								
Develop impact evaluation reports																								

Task	2017										2018				2019				2020				2021	
	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2
Develop policy brief for endline																								
Develop articles for peer-reviewed journals																								
Develop blogs for endline																								
Presentation to implementers																								
Presentations at academic conferences																								
Presentations at policy conferences																								

Timeline of Deliverables

Deliverable	Dates
Baseline report	March 31, 2018
Midline report	March 31, 2019
Endline report	March 31, 2021

References

ASER. (2013). *Annual status of education report 2013—rural.* Retrieved from http://www.prathamusa.org/sites/default/files/aser_2013.pdf

Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *The Quarterly Journal of Economics*, *122*(3), 1235–1264.

Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Pariente, W., Shapiro, J., Thuysbaert, B., & Udry, C. (2015). A multifaceted program causing lasting progress for the very poor: Evidence from six countries. *Science, 348* (6236).

Berry, C., Barnett, E., & Hinton, R. (2015). What does learning for all mean for DFID's global education work? *International Journal of Educational Development, 40,* 323-329.

Bulman., G. & Fairlie, R.W. (2016). *Technology and education: Computers, software, and the internet.* (No. 9432). IZA Discussion Papers

Burde (2012). Assessing impact and bridging methodological divides: Randomized trials in countries affected by conflict. *Comparative Education Review, 56*(3): 448-473.

Chimese, K. (2014). A case of vulnerables teaching OVCs. Retrieved from http://zambiaopencommunityschools.org/news/a_case_of_vulnerables_teaching_ovcs_100011.php.

Conn, K.M. (2014). Identifying effective education interventions in sub-Saharan Africa: A metaanalysis. Retrieved from http://academiccommons.columbia.edu/catalog/ac%3A175014

Cristia, J., Ibarrarán, P., Cueto, S., Santiago, A., & Severín, E. (2012). *Technology and child development: Evidence from the one laptop per child program* (No. IDB-WP-304). IDB Working Paper Series.

Das, J., Dercon, S., Habyarimana, J., & Krishnan, P. (2007). Teacher shocks and student learning: Evidence from Zambia. *Journal of Human Resources, 42*(4): 820-862.

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., & Sundararaman, V. (2013). School inputs, household substitution, and test scores. *American Economic Journal: Applied Economics, 5*(2): 29-57.

DeStefano. (2006). Meeting EFA: Zambia Community Schools. EQUIP2. Retrieved from: http://www.equop123.net/docs/e2-ZambiaCaseStudy.pdf

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

EdData II. (n.d.). Early grade reading. Retrieved from https://www.eddataglobal.org/reading/index.cfm

Duflo, E., Kremer, M., & Glennerster, R. (2007). *Using randomization in development economics research: A toolkit.* (No. 6059). CEPR Discussion Paper Series.

Duflo, E., Hannah, R., & Ryan, S.P. (2012). Incentives work: Getting teachers to come to school. *American Economic Review, 102*(4): 1241-1278.

Filmer, D., & Pritchett, L. (2009). Estimating wealth effects without expenditure data – Or Tears: An application to educational enrollments in states in India. *Demography, 38(1)*: 115-132

Floyd, R.G., Hojnoski, R., & Key, J. (2006). Preliminary evidence of the technical adequacy of the preschool numeracy indicators. *School Psychological Review, 35*(4), 627-644.

Glewwe, P. (2002). Schools and skills in developing countries: Education policies and socioeconomic outcomes. *Journal of Economic Literature, 40*(2), 436–482.

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

Kim, J., & Mueller, C. (1978). Introduction to factor analysis: What it is and how to do it. *Quantitative Applications in Social Sciences*. Sage University Press.

Kremer, M., Brannen, C., & Glennerster, R. (2013). The challenge of education and learning in the developing world. *Science*, *340*(6130), 297–300.

Lee, D.S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies, 76*(3): 1071-1102.

Light, D. (2009). The role of ICT in enhancing education in developing countries: Findings from an evaluation of the Intel® teach essentials course in India, Turkey, and Chile. *Journal of Education for International Development*, *4*(2), 52–66.

McKenzie, D. (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics*, *99*(2), 210–221.

McClure, D., & Gray, I. (2015a). *Engineering scale up in humanitarian innovations' missing middle*. doi: 10.1109/GHTC.2015.7343963

Muralidharan, K., Singh, A., & Ganimian, A. J. (2016). *Disrupting education? Experimental evidence on technology-aided instruction in India* (No. w22923). National Bureau of Economic Research.

Perfetti, C. A. (1985*). Reading ability.* New York: Oxford University Press.

Power, T., Gater, R., Grant, C., & Winters, N. (2014). Educational technology topic guide.

Pritchett, L. (2013). *The rebirth of education: Schooling ain't learning.* Washington, DC: Center for Global Development.

RTI International (2009). Early Grade Mathematics Assessment (EGMA): A conceptual framework based on mathematics skills development in children. Retrieved from http://pdf.usaid.gov/pdf_docs/Pnads439.pdf

Rubagiza, J., Were, E., & Sutherland, R. (2011). Introducing ICT into schools in Rwanda: Educational challenges and opportunities. *International Journal of Educational Development*, *31*(1), 37–43.

UNICEF. (2015). The State of the World's Children 2015: Executive Summary. Retrieved from: http://www.unicef.org/publications/files/SOWC_2015_Summary_and_Tables.pdf.

Winters, C., Schling, M., and Winters, P. (2013) Cost Effectiveness Report for the Impact Network *eLearning Education Program in Rural Zambia*: Final Report Prepared for the Impact Network.