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

APRIL 2018

Thomas de Hoop | Andrew Brudevold-Newman | Dustin Davis

MAKING RESEARCH **RELEVANT**

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

APRIL 2018

Thomas de Hoop | Andrew Brudevold-Newman | Dustin Davis

AMERICAN INSTITUTES FOR RESEARCH®

1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
202.403.5000

www.air.org

4117_04/18

Contents

Tables

Figures

Executive Summary

Low- and middle-income countries have made significant progress getting children into school, but student learning and achievement are often dreadfully low (Berry, Barnett, & Hinton, 2015; Pritchett, 2013). Approximately 250 million children across the world are not acquiring basic reading and math skills, even though about half have spent at least 4 years in school (United Nations Educational, Scientific, and Cultural Organisation, 2014).

Zambia faces many common educational challenges. Literacy rates among young Zambian adults ages 15–24 are 58.5% for females and 70.3% for males, despite an average of 7.7 years and 7.9 years of education, respectively (Zambia Demographic and Health Surveys, 2013–14; UNICEF, 2015). Furthermore, public spending on education is low relative to other regional countries: Zambia spends about 1.3% of its gross domestic product as compared with an average of 5.6% in both Southern and Eastern Africa an (UNICEF, 2015). Zambia also has a large, autonomous community schooling system that formed during Zambia's transition from a socialist economy. The community schooling system has expanded over the past 20 years to increase education access in remote areas: the number of community schools is estimated to have increased from 100 schools in 1996 to about 3,000 schools with 600,000 children in 2014 (Chimese, 2014; DeStefano, 2006). However, community schools are often staffed by untrained and underpaid teachers who teach a substandard curriculum and who lack management skills and school supplies. Community schools in Zambia are in need of a cost-effective solution for delivering quality education in order to improve learning outcomes.

The Impact Network program represents a promising approach to improving educational outcomes by incorporating three potentially high-impact components that could create important synergies: e-learning, ongoing teacher training and professional development, and community ownership. Each component could, on their own, have positive impacts on student outcomes by engaging the three main actors in the education system: students, teachers, and parents. Combining these components into a single program may be particularly effective by aligning all three actors towards improving the educational outcomes of the 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, 2016).

The American Institutes for Research (AIR) has designed and is implementing a mixed-methods cluster-randomized controlled trial (cluster-RCT) to determine the impact of Impact Network's eSchool 360 model. The 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. To determine the impact of the program, we are using a cluster-RCT in which 64 eligible schools have been randomly assigned either to receive Impact Network's eSchool 360 program (30 treatment schools) or not to receive the program (34 control schools).

The primary cognitive skills outcomes are aggregate scores on the early grade reading assessment (EGRA), early grade math assessment (EGMA), and the Zambian Achievement Test (ZAT). The population consists of children between 6 years old and 9 years old who live close to the 64 selected community schools across three districts in Zambia's Eastern Province: Petauke, Sinda, and Katete.

This report presents the baseline results of the cluster-RCT used to determine the impact of Impact Network's eSchool 360 model. It examines the differences between the treatment and control households along the causal chain of the theory of change. In addition, it analyzes the potential for floor effects on the assessment instruments that will be used to measure the impact of the Impact Network program on student outcomes 1 year and 3 years after the introduction of the Impact Network eSchool 360 model.

The eSchool 360 model successfully started operating in the 30 study schools and five quasi-governmental schools in January 2018, despite challenges associated with a cholera epidemic in Zambia. Although the cholera outbreak primarily affected the capital Lusaka (with more than 2,800 cases since September 2017), the government decided to implement a nationwide postponement of the start of the school year (originally scheduled for January 16, 2018). Since then, schools have been allowed to start only after receiving an inspection certificate that proves that the latrines and handwashing materials are adequate. The vast majority of Impact Network schools started the school year on January 29, after teacher training was conducted during the week of January 22. All Impact Network schools started the school year by February 5th. As of the writing of this baseline report, we do not have information when the control schools started. We will collect this information during the follow-up survey that will occur in November–December 2018.

The baseline findings in this report suggest that the cluster-RCT was successful in creating equivalence in observable characteristics between treatment and control households. We did not find evidence for systematic statistically significant differences. Furthermore, almost none of the statistically significant differences at baseline are larger than 0.3 standard deviations. This finding indicates that the randomization will enable AIR to make causal claims about the impacts of Impact Network's eSchool 360 program after the midline data collection and analysis (one year after the introduction of the program) and the endline data collection and analysis (three years after the introduction of the program).

In addition, the analyses suggest the estimation of program impacts on EGRA and EGMA outcomes might be subject to floor effects, but there is encouraging evidence that the ZAT and the oral vocabulary test follow approximately a normal distribution. Unsurprisingly, the children in our sample scored very low on the EGRA and EGMA assessments at baseline. The midline and endline analyses will have to determine whether students improve sufficiently to mitigate concerns about floor effects in the estimation of program impacts on EGRA and EGMA outcomes. In the presence of floor effects, we will have to rely primarily on the estimation of the impacts of Impact Network's eSchool 360 program on the ZAT and the oral vocabulary test, in addition to intermediate outcomes (e.g., school attendance and enrollment) of the theory of change.

We plan to collect midline data in November–December 2018. These midline data will again include the collection of EGRA, EGMA, and ZAT data and the collection of household-level survey data on school enrollment and attendance, student-level aspirations, and parental-level aspirations. In addition, we will collect qualitative data in two schools in each of the three evaluation districts. We will use three primary approaches to qualitative data collection for the midline evaluation: key informant interviews with community leaders, Impact Network's eSchool 360 program staff, teachers, and students; focus group discussions with students and parents; and classroom observations. We also plan to collect cost data to inform the cost-effectiveness analysis in November–December 2018.

Introduction

There have been dramatic increases in educational attainment over the past few decades but the quality of the education and overall student learning continue to lag far behind (Berry et al., 2015; Pritchett, 2013; World Bank, 2018). Approximately 250 million children across the world are not acquiring basic reading and math skills, even though about half have spent at least 4 years in school (UNESCO, 2014). The current set of global development goals—the Sustainable Development Goals—shifts the focus from educational attainment to education quality, with the goal to "ensure inclusive and equitable **quality** education and promote lifelong learning opportunities for all" (United Nations, 2017, emphasis added).

Zambia is emblematic of many low- and middle-income countries that face several educational challenges. First, overall education quality is low: literacy rates among young Zambian adults aged 15–24 are 58.5% for females and 70.3% for males, despite an average of 7.7 years and 7.9 years of education, respectively (DHS 2013–14; UNICEF, 2015). Second, public spending on education is low relative to other regional countries: Zambia spends about 1.3% of its gross domestic product as compared with an average of 5.6% in Southern Africa and Eastern Africa (UNICEF, 2015). There is evidence, however, that additional funding to schools may not be a solution to low education quality. A rigorous study found no evidence of a positive impact of a fixed block grant provided by the Zambian government on student learning outcomes (Das et al., 2013). Third, Zambia has a large, autonomous community schooling system that formed during Zambia's transition from a socialist economy. The system has expanded over the past 20 years to increase education access in remote areas: the number of community schools is estimated to have increased from 100 in 1996 to about 3,000 schools with 600,000 children in 2014 (Chimese, 2014; DeStefano, 2006). However, community schools are often staffed by untrained and underpaid teachers who teach a substandard curriculum and who lack management skills and school supplies. Improving education quality in community schools may be an effective entry point to improve educational outcomes for vulnerable children in remote areas.

This baseline study focuses on the effects and cost-effectiveness of Impact Network's eSchool 360 model, which represents a promising approach to delivering quality education and improving educational outcomes for students in community schools in rural Zambia. It incorporates three potentially high-impact interventions that could offer important complementarities: e-learning technology, ongoing teacher training and professional development, and community ownership. The e-learning component includes electricity via solar power (provided by Impact Network), and projectors and tablets (provided by Impact Network's partner, iSchool) for the community schools loaded with materials in the local language that are structured around a curriculum approved by the Zambian government. Impact Network supplements the technology by providing teacher training and professional development to community schoolteachers and creating community ownership. Locally hired teachers receive weekly training focused on using the technology and enhancing their pedagogical skills.

Combining e-learning, ongoing teacher training and professional development, and community ownership components into a single program may be particularly effective by aligning the incentives of students, teachers, and parents towards improving student educational outcomes. The components could each, on their own, have positive impacts on student outcomes. Earlier research has highlighted that engaging all three actors in the education system (students, teachers, and parents) may be particularly effective because it creates important and sizable synergies. For example, an educational technology program in urban India that included a strong focus on pedagogical practices showed impacts larger than the sum of those obtained from separate educational technology or pedagogical interventions (Muralidharan et al., 2017). A cost-effectiveness analysis of the same program found that the program was also highly cost-effective. This is an important point considering a recent review found that technology-based education programs may not be cost-effective, even if they produce large impacts on learning outcomes (Muralidharan et al., 2016; Piper, Simmons, Zuilkowski, Kwayumba, and Strigel, 2016).

The study comprises three main evaluation parts: an impact evaluation of the eSchool 360 model, an analysis of the cost-effectiveness of the model, and a process evaluation of the expansion of the model. This report presents findings only related to the quantitative data of the evaluation because the baseline study included only quantitative data collection. We have also produced an inception report which details the design of all parts (De Hoop et al., 2017). We will collect our first round of qualitative data in summer 2018 and will present an analysis of these data in the midline report (one year after the introduction of the program). However, in the following, we present research questions related to each of the parts. Each part of the evaluation is designed to answer different, but complementary, questions:

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?

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?

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 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?

This report presents the baseline results of the cluster-randomized controlled trial (cluster-RCT) to determine the impact of Impact Network's eSchool 360 model. It examines the differences between the treatment and control households along the causal chain of the theory of change. In addition, it analyzes the potential for floor effects on the assessment instruments that we will use to measure the impact of Impact Network's eSchool 360 program on student outcomes 1 year (midline) and 3 years (endline) after the program's introduction. The primary cognitive skills outcomes are aggregate scores on the early grade reading assessment (EGRA), early grade math assessment (EGMA), Zambian Achievement Test (ZAT), and an oral vocabulary test. The report also presents multivariate regression analyses to analyze the predictive power of student-level and household-level observable characteristics in determining EGRA, EGMA, ZAT, and oral vocabulary test outcomes.

The rest of this baseline report is structured as follows. It starts with a description of the background of the Impact Network's eSchool 360 model, and includes an overview of the existing literature on the impact of technology-based education programs. Next, it presents a description of the model, followed by a description of the theory of change and the quantitative research design. It then offers a detailed overview of the quantitative baseline results, which is followed by a conclusion.

Background

The current Zambian educational system provides low-quality education yet remains inaccessible for many of the 52.5% of the Zambian population under the age of 18 (Central Statistics Office Zambia, 2013). An estimated 600,000 students attend nongovernmental, autonomous, community schools that do not offer a full range of grades, are in poor condition, and are funded through minimal government funding of less than $91 per year (1,000 Zambian Kwacha) per school (DeStefano, 2006). Students in Zambia routinely score below their regional neighbors: students across both government and community schools in Zambia attained the lowest marks in reading and tied for the lowest marks in mathematics of the 14 sub-Saharan countries tested as part of the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ, 2010–12). The Zambian government has introduced evidence-backed interventions in government schools to improve school quality, such as teaching at the right level.[1] This intervention groups students according to learning level rather than by age or grade (Banerjee et al., 2016). The autonomous nature of the community schools raises concerns related to whether they may be neglected from this quality push as well as whether a different set of interventions may be necessary to improve their quality.

Previous research demonstrates that multifaceted education programs such as Impact Network's eSchool 360 model can be effective in improving learning outcomes. A comprehensive systematic review on the impact of education programs in low- and middle-income countries concludes that successful education programs address constraints at multiple levels (Snilstveit et al., 2016), which can only be achieved by multifaceted education programs, such as Impact Network's eSchool 360 model. Kremer, Brannen, and Glennerster (2013) also

[1] The Teaching at the Right Level program groups children based on their learning level, which is based on an assessment test. Banerjee et al. (2016) show that the program led to significant gains in language skills for children who participated in the program as it was implemented by state governments in Haryana and Uttar Pradesh in India. However, these learning gains were only achieved after the introduction of careful, top-down support and the monitoring to ensure that classrooms were reorganized around initial learning levels by government teachers.

highlight the importance of adapting the curriculum to the child's level to improve the effectiveness of pedagogical interventions. Furthermore, Conn (2014) suggests that among interventions that include 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 afterschool instruction program in India, students in the treatment group made significant gains in math and Hindi test scores. They concluded 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. Finally, Banerjee Cole, Duflo, and Linden (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.

By contrast, there is evidence that programs in low- and middle-income countries that focus on increasing educational inputs without addressing other constraints to learning are not sufficient to improve learning outcomes (Schling & Winters, 2015; Snilstveit et al., 2016). An increased provision of traditional school resources such as textbooks or flipcharts had no impact on student attainment (Glewwe, 2002). Banerjee et al. (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 learn. Similarly, a number of studies that have focused on computer-assisted learning programs did not find significant impacts. For example, Cristia, Ibarrarán, Cueto, Santiago, and Severín (2012) analyzed the effect of the One-Laptop-Per-Child program for students in rural Peru; they 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 directly linked to class material, such as mathematics or reading, as well not having clear instruction on how teachers should use the computers in class.

The overall literature on technology in education, as demonstrated in Muralidharan et al. (2017) and Cristia et al. (2012), suggests that the provision of technology must be focused on pedagogical improvements to be effective in improving learning outcomes. A review of 45 studies that examined 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 majority of the targeted schools. Furthermore, the authors found that teaching and learning was more teacher-centric in schools in which 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, 2011).

Evidence suggests that when implemented with a strong pedagogical focus, technology-in-education programs are more likely to produce positive changes (Power et al., 2014). A case study of Intel's 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 over the course of the study spoke positively about new learning activities such as project-based work that gave them a chance to collaborate, use multiple resources, and direct their own learning, as well as about schoolwork that was more relevant to their lives outside of school, making learning more meaningful. Teachers also demonstrated better understanding of student-centered teaching practices and of ICT knowledge and skills (Light, 2009).

The eSchool 360 Program. Impact Network developed the eSchool 360 model to deliver low-cost education to children in rural communities through a holistic solution. The cost of eSchool 360 is $3 per month per student, which is 70% less than the Zambian government spends per student (Winters, Schling, and Winters, 2013). The core of the model is e-learning technology whereby tablets and projectors, provided by Impact Network's partner iSchool, are loaded with curricula approved by the Zambian government and in the local language. Impact Network provides electricity via solar power, and supplements the technology by providing teacher training and professional development and creating community ownership. Locally hired teachers receive weekly training that is 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 provides them with systematic, ongoing support.

A previous nonexperimental evaluation of the eSchool 360 model suggests that it 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-grade students 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 1 percentage point cost Impact Network Schools 88% less than it cost government schools (Schling & Winters, 2015).

These encouraging results led the American Institutes for Research (AIR) to fund the expansion of the eSchool 360 model to 30 additional community schools in rural Zambia (from a total of 9 community schools to 39 community schools).[2] Impact Network conducted the expansion in 2017 by implementing the eSchool 360 model in 30 community schools across three rural Zambian districts (Katete, Sinda, and Petauke) in areas with no running water and limited electricity. The first cohort of Impact Network students was admitted in January 2018. Impact Network now enrolls more than 5,000 primary-school children in 39 community schools and 5 quasi-governmental schools across these three districts. Over the next five years, Impact Network hopes to reach 1 million students and to benefit 5 million citizens across Zambia through a partnership with the government.

The eSchool 360 model successfully started operating in the 30 study schools and 5 quasi-governmental schools in January 2018, despite challenges associated with a cholera epidemic in Zambia. Although the epidemic primarily affected the capital Lusaka (with more than 2,800 cases since September 2017), the government decided to implement a nationwide postponement of the start of the school year (originally scheduled for January 16, 2018). Since then, schools have been allowed to start only after receiving an inspection certificate to prove that the latrines and handwashing materials are adequate. The vast majority of Impact Network schools started the school year on January 29, after teacher training was conducted during the week of January 22. All Impact Network schools started the school year by February 5. As of the writing of this baseline report, we do not have information when the control schools started. We will collect this information during the follow-up survey that will occur in November–December 2018.

Rigorous Evaluation. AIR and Impact Network designed the expansion to include a rigorous mixed-methods cluster-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. The random assignment of schools was conducted in May 2017. Ministry of Education officials implemented the randomization, and AIR staff ensured the integrity of the process. AIR chose an unbalanced design with a smaller number of treatment schools because of limited resources to implement the eSchool 360

[2] AIR decided to fund the expansion of the program to five quasi-governmental schools that were not included in the cluster-RCT. We will compare learning outcomes of the students in these five schools to the learning outcomes of students in five comparable governmental schools during the final data collection, which will be three years after the introduction of the model in the quasi-governmental schools.

model. We will supplement the cluster-RCT with a process evaluation to assess the fidelity of implementation of the eSchool 360 model and will conduct a cost-effectiveness analysis.

The evaluation of Impact Network's eSchool 360 model 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 model uses a multifaceted approach to incorporate the provision of e-learning technologies and has 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 a pedagogical focus and a focus on addressing constraints at multiple levels to be effective at improving learning outcomes. Additionally, the setting of the eSchool 360 model in rural Zambian community schools will provide new cultural, geographic, and administrative contexts; especially in conjunction with the qualitative research, this will be invaluable in understanding the generalizability of the evidence of a popular and growing type of intervention throughout the developing world.

Theory of Change

The theory of change that underlies the program suggests that the eSchool 360 program may lead to improvements in learning outcomes through various mechanisms (Figure 1). First, the teacher professional development component of the model may lead to improvements in the knowledge and practices of untrained 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 may lead to improvements in preliteracy, 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 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 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 be able to finish 12th grade. In addition, the improved quality of education may result in increased expectations of the returns of education, which may lead to higher expectations for labor market outcomes. Finally, increased aspirations in the education and labor market domains may result in increases in expectations related to the marriage prospects of students as well as increase their age-at-marriage.

The validity of the theory of change depends on several assumptions. Perhaps most important, teachers need to comply with the e-school programming. In addition, the community schools need to have sufficient capacity to implement the model. 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 needs of the student population.

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

The moderators we identified are closely related to a conceptual model developed by the World Bank for its World Development Report on education (World Bank, 2018). The conceptual framework highlights how learning outcomes are directly affected by the quality of school inputs, school management, and teachers, as well as the education preparedness of learners. In theory, improvements in the quality of one of these factors could lead to improvements in learning outcomes (Figure 2). However, the World Development Report presents evidence showing that learning is unlikely to be positively affected unless the quality of each factor improves (World Bank, 2018). This result is consistent with the idea that education models are unlikely to improve learning outcomes unless they address multiple constraints, as highlighted in Snilstveit et al. (2016). It also shows the importance of the multifaceted approach Impact Network uses to improve the quality of education. The eSchool 360 model aims to improve learning outcomes through interactions with a wide range of stakeholders, which could lead to improvements in the quality of all factors.

Ultimately, the goal of our evaluation is to inform how the eSchool 360 model can be scaled up effectively in Zambia. Achieving this goal requires a combination of different research and evaluation approaches to guide an iterative model design in which the implementing partner uses each evaluation finding to reflect on and, if needed, refine the model design. Such an approach incentivizes the use of evaluation findings, which then increases the value of the evaluation. We highlight this adaptive approach to guiding scale-up of the program in the scaling framework in Figure 3. 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 an intervention 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 model as needed when the scope of innovation expands. This iterative, evidence-driven approach requires effective evaluation.

The World Development Report also highlights the importance of working with a range of key stakeholders to successfully scale-up education programs that produce improvements in learning outcomes (World Bank, 2018). Specifically, the report presents how different stakeholders (teachers, principals, bureaucrats, politicians, parents and students, the judiciary, employers, NGOs, suppliers of educational inputs, and international donors) each have learning-aligned interests and competing interests (World Bank, 2018). These interests (which are depicted in Table 1) need to be taken into consideration to successfully scale the Impact Network's eSchool 360 model.

Figure 1. Theory of Change

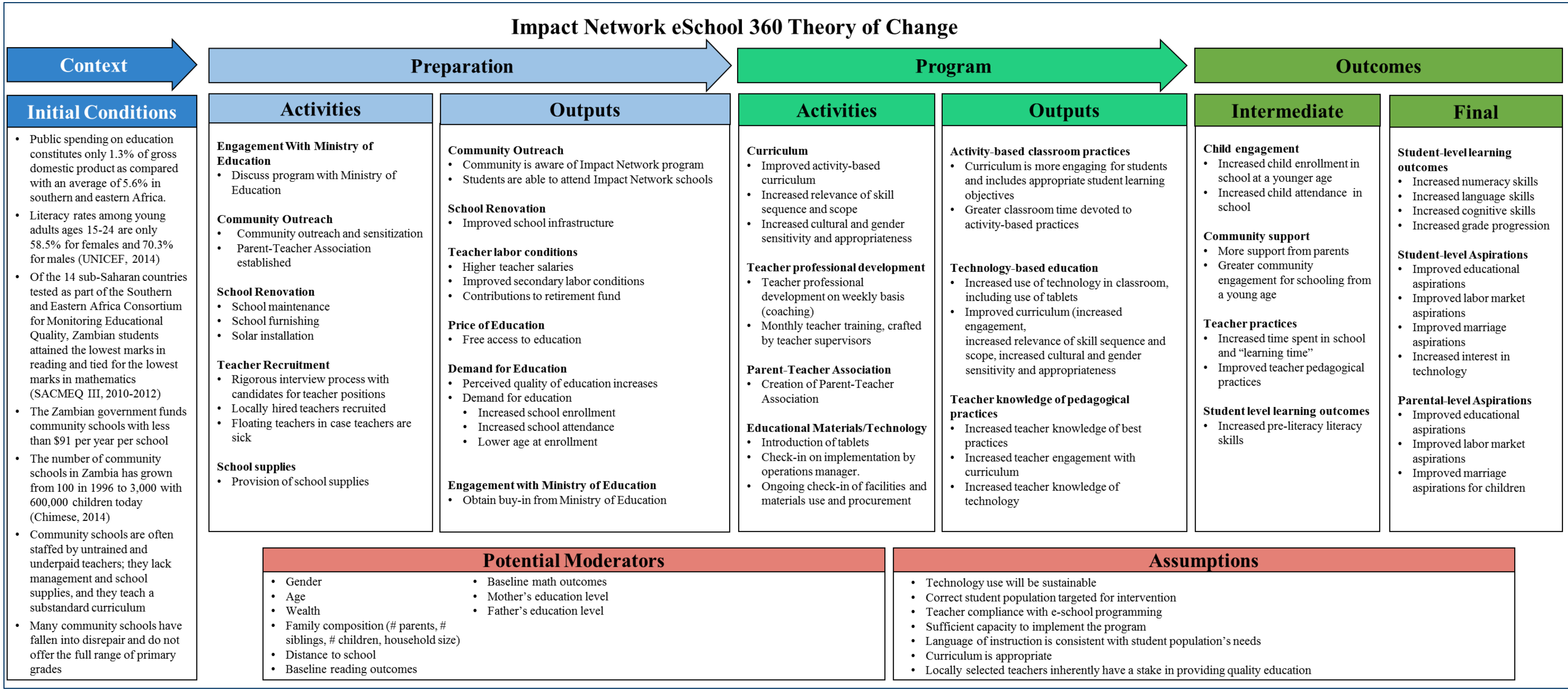
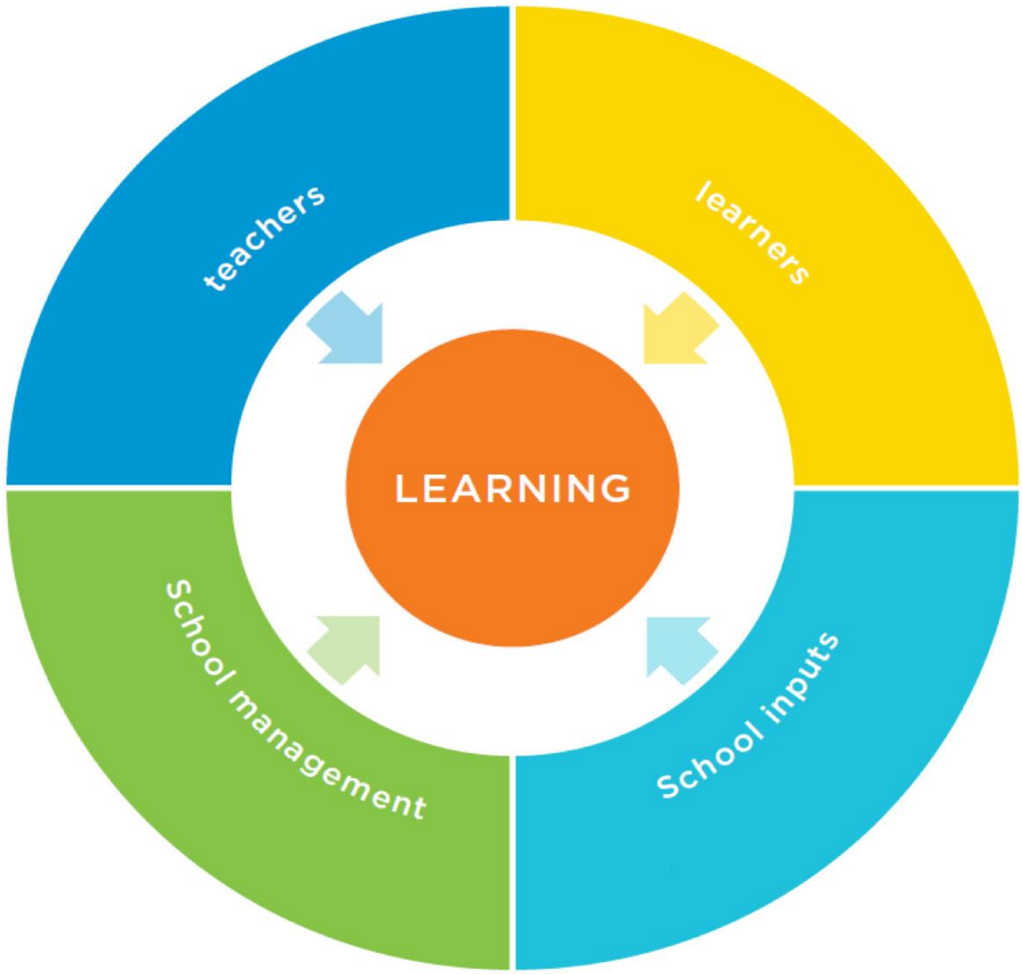


Figure 2. Conceptual Model for Improving Learning Outcomes



Note: Reprinted from World Bank (2018)

Figure 3. Scaling Framework

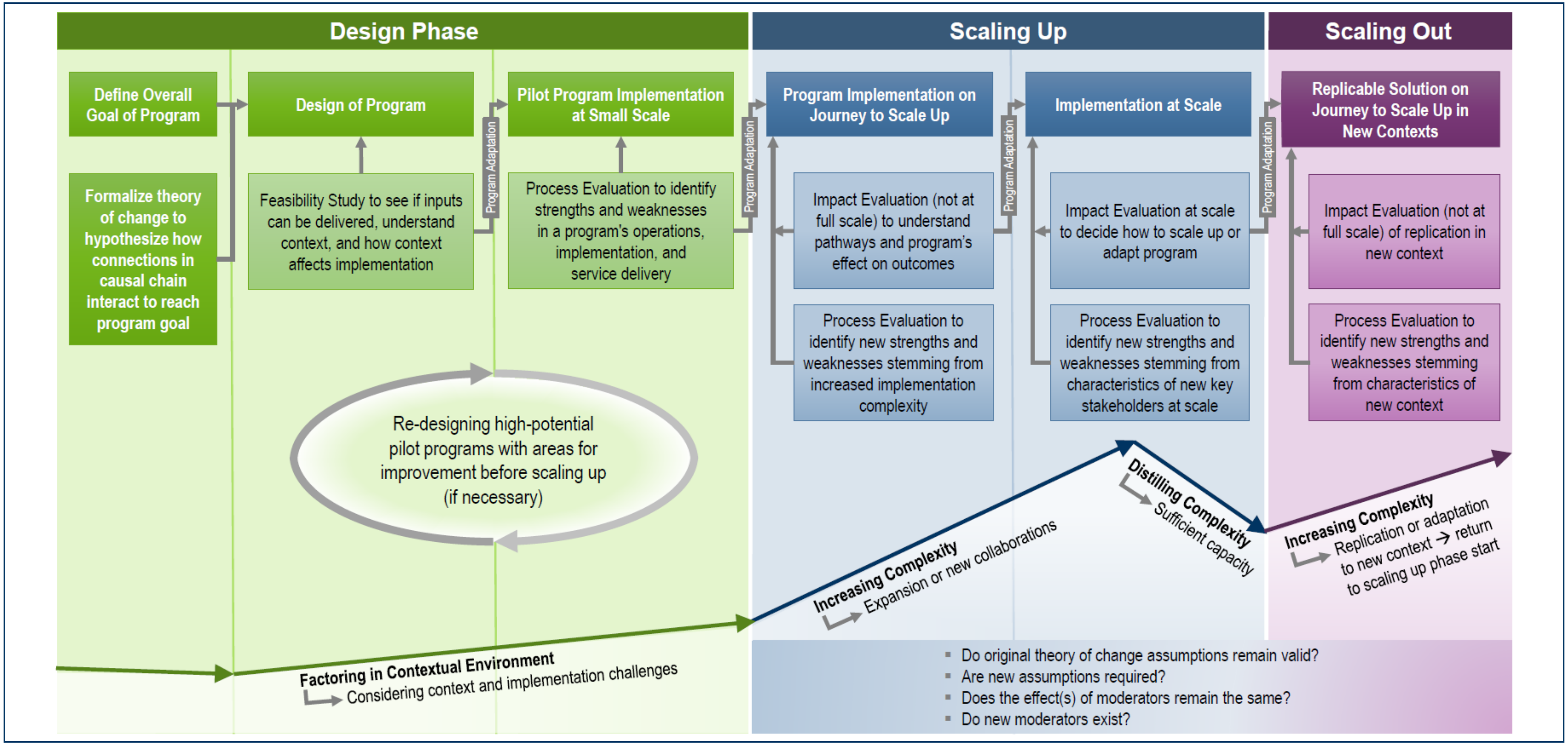


Table 1. Multiple Interests of Key Stakeholders

Stakeholders	Examples of learning-aligned interests	Examples of competing interests
Teachers	Student learning, professional ethics	Employment, job security, salary, private tuitions
Principals	Student learning, teacher performance	Employment, salary, good relations with staff, favoritism
Bureaucrats	Well-functioning schools	Employment, salary, rent-seeking
Politicians	Well-functioning schools	Electoral gains, rent-seeking, patronage
Parents and students	Student learning, employment of graduates	Family employment, family income, outdoing others
Judiciary	Meaningful right to education	Favoritism, rent-seeking
Employers	Skilled graduates	Low taxes, narrowly defined self-interests
Nongovernmental schools (religious, nongovernmental, for-profit)	Innovative, responsive schooling	Profit, religious mission, funding
Suppliers of educational inputs (e.g. textbooks, information technology, buildings)	High-quality relevant inputs	Profit, influence
International donors	Student learning	Domestic strategic interests, taxpayer support, employment

Note. Reprinted from World Bank (2018)

Study Design

AIR has designed and is implementing a mixed-methods cluster-RCT. The data collection started with quantitative baseline data, which will be followed by quantitative and qualitative data collections (1 year and 3 years after the start of the baseline data collection) to inform possible scale-up of Impact Network's eSchool 360 model. Impact evaluations of interventions 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 models influence education outcomes, even if they are effective, which limits the learning potential of impact evaluations. In an 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. Specifically, we will assess the costs of the eSchool 360 model using the ingredients method. For this purpose, we will need to specify all the ingredients that are necessary to replicate the model and then collect data on the unit costs of all 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 consider including opportunity costs for the beneficiaries in this cost analysis. The cost-effectiveness estimates will guide policymakers in assessing the value for money of investing in the eSchool 360 model.

This report presents findings from the baseline study, which comprised only quantitative data collection. The inception report includes a preliminary qualitative research design and a description of the methods we propose to use for the cost-effectiveness analyses (De Hoop et al., 2017). We will collect a first round of qualitative data in summer 2018 and will present an analysis of these data in the midline report.

Quantitative Study Design

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

Randomization Across Eligible Schools

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

Impact Network and AIR first consulted with local Zambian government officials to obtain a list of all community schools in the region. Impact Network was able to identify several community schools that were not on the government list; and it then collected data on each school. Of the 149 community schools that were identified, 64 met all the eligibility criteria.[3] Impact Network staff then visited each of the 64 schools to obtain information on 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.

AIR oversaw the randomization that determined which schools received the eSchool 360 model. Of the 64 eligible schools, 10 are in Katete, 18 in Sinda, and 36 in Petauke. Representatives from the local Ministry of Education implemented the randomization, assigning to the treatment group (receiving the eSchool 360 model): 5 schools in Katete, 9 in Sinda, and 16 in Petauke. The remaining 34 schools were assigned to the control group (not receiving the model). It was important to obtain letters of approval from local ministry officials as their participation in the randomization encouraged buy-in from the Zambian government. This

[3] Impact Network originally identified 65 eligible schools. However, further analysis suggested that the 65th 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.

allows us to maintain the fidelity of the randomization and the credibility of the study. Finally, this process encouraged government officials to be invested in and knowledgeable about the evaluation. As highlighted in the World Development Report (World Bank, 2018) and in our scaling framework, these processes contribute to an increased likelihood of a successful scale-up of the eSchool 360 model if the evaluation shows evidence of a cost-effective, quality education for students who attend community schools.

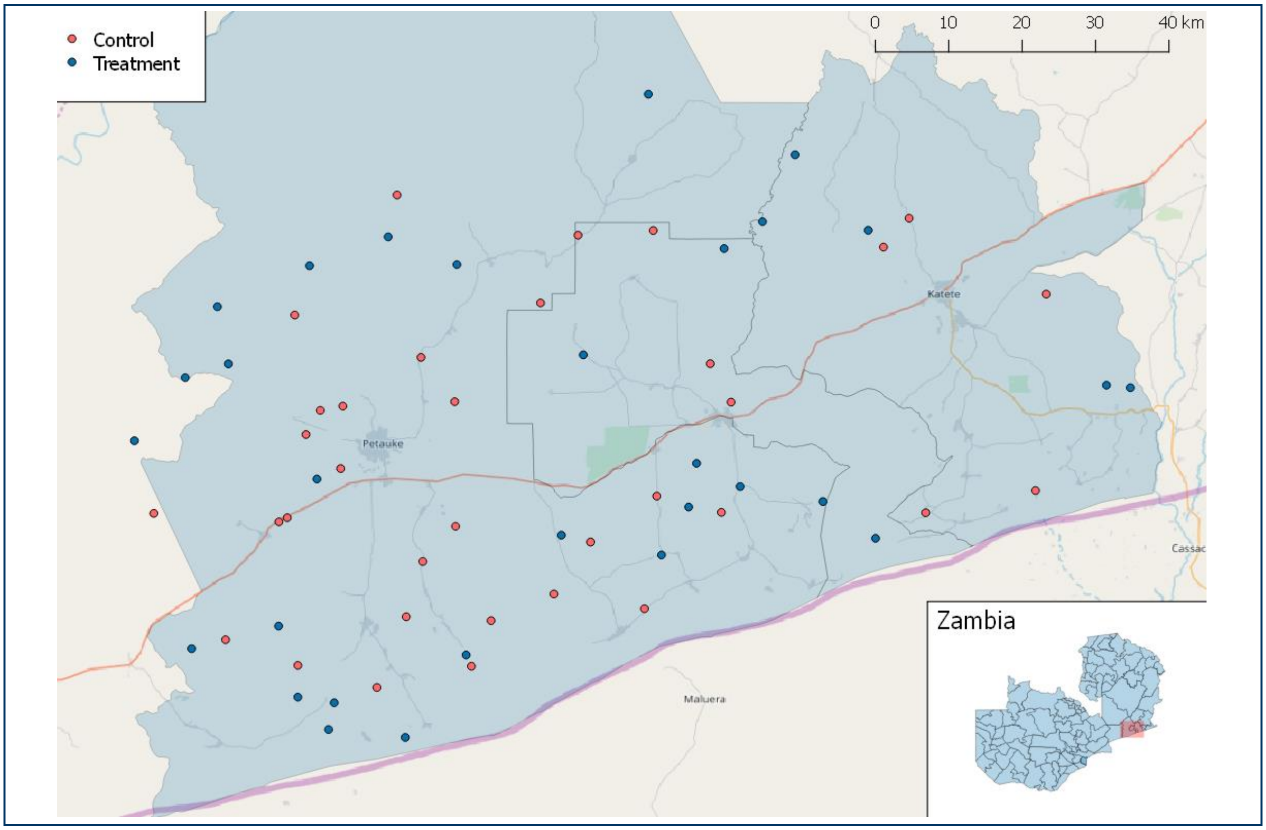
Sampling

This study will estimate the impact of the eSchool 360 model on children *eligible* to enroll in first grade and who live near the 64 schools. We focus on these children because the introduction of the eSchool 360 model was not designed to benefit all students in a school at the same time; rather, it is designed to expand to an additional grade each year that the model is operating. In the first year, only the first-grade cohort will receive the full package. This will expand to Grades 1 and 2 in the second year, and so on. The study will use a longitudinal panel design that follows each of the sampled children for 3 years, regardless of where and when they enroll in school, to estimate intention-to-treat effects of the model on school attendance and enrollment as well as preliteracy, literacy, and numeracy outcomes.

To identify the sample of children with the potential to be affected by the model and determine intention-to-treat effects, 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 and who did not attend first grade in the prior school year. We identified all households with children eligible to enroll in first grade within a diameter of 1.5 kilometers of the schools. We iteratively expanded the distance by 0.5 kilometers in communities with insufficient numbers of eligible children within the initial or subsequent sampling areas until we found sufficient eligible households: this procedure was implemented consistently across treatment and control school-catchment areas.[4] A map of the schools is shown in Figure 4.

[4] 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.

Figure 4. Map of Sample Schools



It is important to identify intention-to-treat 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 7 months before the first evaluation cohort was 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 influencing school enrollment and attendance. In fact, a previous nonexperimental evaluation of the Impact Network model 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 enrolled in other community schools because of differences in either observable (e.g., age, gender, parents' education level) or unobservable (e.g., motivation, noncognitive skills) characteristics.[5]

[5] We are currently considering the feasibility of collecting data on school enrollment and attendance in April 2018. If feasible, we will collect these data through SMS messages to both treatment and control households. We will provide phone credit to respondents who respond to the SMS messages to incentivize respondents to participate in the survey.

To address concerns regarding these composition effects and to enable the estimation of intention-to-treat 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. Thus, we expected 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.

To increase statistical power, we initially planned to use stratified random sampling by age. We planned to oversample 8-year-olds and 9-year-olds because descriptive statistics indicated that they are more likely than 7-year-olds to be enrolled in first grade (Ministry of General Education, 2014). Of the children enrolled in first grade in the Eastern Province in 2014, 4,484 were younger than 7 years old; 23,206 were 7 years old, and 35,220 were 8 years old or older. We also planned to exclude 6-year-olds in our sample because descriptive statistics suggested that they are very unlikely to be enrolled in the first grade.

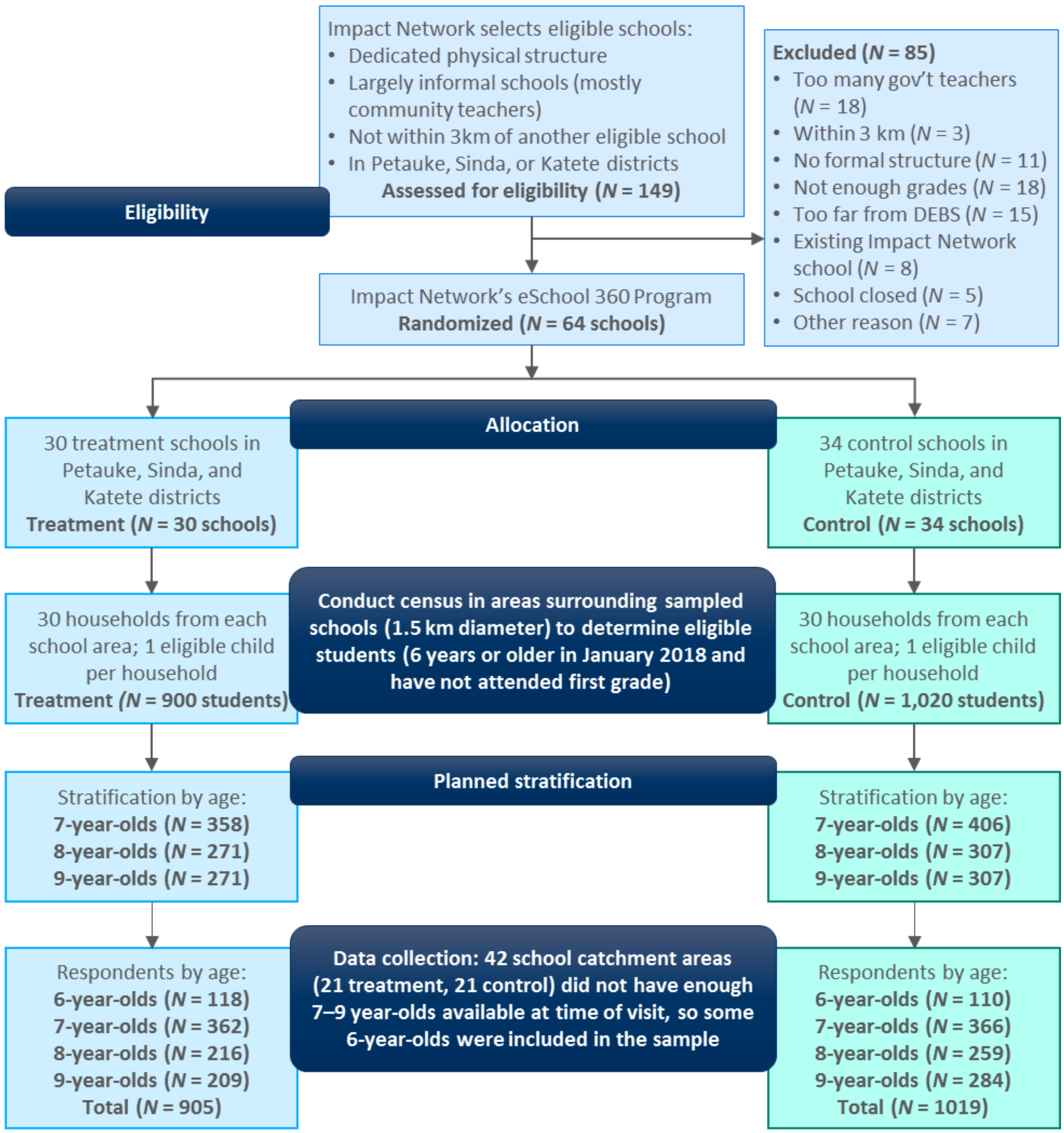
To increase statistical power, we planned to use the same age distribution in our sample (after excluding 6-year-olds), while assuming that 8-year-olds and 9-year-olds will be enrolled in first grade at an equal rate. Oversampling groups more likely to be enrolled in school will increase the take-up of the model and our ability to estimate its effects on learning outcomes with sufficient precision thereby increasing statistical power. This preliminary sampling strategy is presented in Table 2.

Table 2. Oversampling of 8-Year-Olds and 9-Year-Olds

Age Category	Percent of Children in First Grade	Sample Size (*N*) 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

In practice, we had to slightly adjust the sampling strategy because, as noted earlier, we were not able to find 30 households meeting these criteria within a distance of 1.5 kilometers from each of the schools in the sample. To adjust the sample, we first randomly sampled 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 and then sampled within the expanded range. Figure 5 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 the data collection.

Figure 5. Consort Flow Chart



Data Collection

Quantitative Data Collection

AIR is partnering with Palm Associates, a Zambian research organization that specializes in data collection and social science research, to collect data for the evaluation of Impact Network's eSchool 360 model. AIR previously worked with Palm Associates 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 the Impact Network evaluation, so the organization is familiar with the geography and culture. AIR worked closely with Palm to train enumerators before the baseline data collection and then followed 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 is also helping to 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 used tablets to conduct the baseline data collection, thus improving the quality of data collection, minimizing the need to clean data, and eliminating data entry. We collected the data on tablets running SurveyCTO software, which minimizes errors in the field because skipping patterns can be automated and built-in checks ensure the quality of data. The SurveyCTO software runs on the Open Data Kit (ODK) platform. ODK enables users to collect data on a tablet, send data to a server, aggregate the collected data, and extract data in Stata format.

Outcome Measures

Quantitative Indicators

The primary achievement indicators for numeracy, preliteracy, and literacy come from the EGRA, EGMA, and ZAT, which have already been adapted and used in the same region as this study. All of these instruments have been translated and validated in the context of Zambia. We collected baseline assessment data in Nyanja because this is the language of instruction in Grades 1–3 in Eastern Zambia.

In future data collection rounds, we will collect enrollment and attendance data during the household survey. These data will be validated using administrative attendance-data from the treatment schools. Outcome indicators on parental and community perceptions of school, teachers, and their children's education will also be collected through this household survey. We will collect specific data on parents' and children's educational, labor market, and marriage aspirations.

In addition to these outcome indicators, the survey will also collect control and moderating variables at different levels, 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 teaching, 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 on student outcomes 1 year (midline) and 3 years (endline) 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.

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 it takes to complete each of these sections. At the endline, the secondary math outcomes will be filling in the missing number, addition and subtraction, geometry, 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.

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.

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.

Previous experience in Zambia suggests that a majority of students will score quite low on their EGRA and EGMA tests in 2018. 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,[6] 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.

To mitigate concerns about floor effects, it will 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 task will measure the

[6] The ZAT assessment was 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.

spoken language skills of a child. Estimating impacts on oral reading fluency is also less vulnerable to floor effects.

Factor Analysis

In addition to estimating impacts on aggregate percentage scores on the EGRA and EGMA, 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 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 (higher than 0.4) can indicate which subtasks are tapping into which common dimensions.

Comparison With Outcomes in Government Schools

In addition to the cluster-RCT, we will compare the EGRA, EGMA, and ZAT outcomes of children enrolled in five quasi-governmental 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 quasi-governmental 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, ZAT, and oral reading fluency measures, we will estimate the program impacts on school attendance and enrollment, parents' perceptions of school and education quality, student-level aspirations, and parental-level aspirations for their children. Using these outcome measures will enable us to determine impacts along the causal chain of the theory of change and examine the mechanisms underlying the program impacts.

School enrollment and attendance: During the next data collection rounds, we will measure school enrollment and attendance by relying on self-reported student data because the intervention intends to change how schools measure enrollment and attendance data. This change could potentially make data from the Impact Network schools systematically different from the control schools. To obtain the self-reported data, 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 did not include measures of school enrollment and attendance. During the baseline, we asked parents only whether they expected to enroll their children in 2018. However, we will not use this as a control variable because it could be affected by the announcement of the treatment schools and the construction of the infrastructure associated with the eSchool 360 program.[7]

Parents' and children's perceptions of school and education quality: We measured parents' perceptions of school and education quality by asking 4-point Likert-scale questions related to their 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. In future data collection rounds, 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 parental-level aspirations: We also measured students' and parents' aspirations with respect to education, marriage, and labor market outcomes. To measure student aspirations, we asked students about the level of education they would like to achieve and the age at which they would like to get married. To measure parental aspirations, we asked parents

[7] In addition to this outcome measure, we are considering collecting SMS data on school enrollment and attendance in April 2018.

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.

Cluster-RCT Analyses

Research Question 1 focuses on the impact of the Impact Network program on a variety of student and parental outcomes. The main impact analyses will be conducted as intention-to-treat analyses, which will measure the impact of living within 1.5 kilometers (or within 3 kilometers for a small minority of the children) 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 section, 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

In this report of the baseline results, we 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 regression analysis, we will also calculate the normalized difference to examine balance. In keeping with the work of Imbens (2015), in our impact models we will control for any variables that have a normalized difference of more than 0.25.

Impact Analysis Sample

The impact evaluation research design includes a baseline survey (conducted) and two rounds of postintervention data collection: a midline survey (1 year after the introduction of the Impact Network model) and an endline survey (3 years after the introduction of the model). The midline analysis will include all individuals surveyed at the midline, and the endline analysis will include all individuals surveyed at the endline. Our main evaluation sample will include all respondents surveyed at each round (the baseline, midline, and endline). Following the endline, we will produce analyses that restrict attention to the sample of respondents surveyed at all three rounds. This supplemental analysis set will allow us to track program impacts across time using a consistent sample.

Statistical Models

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 of detecting a statistically significant effect if the program indeed causes statistically significant effects.

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:

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

Here IN_i is an indicator variable equal to 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 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.

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 procedure, 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 two domains: primary EGRA outcomes and primary EGMA outcomes. These 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 procedure. 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 model. We will test for whether it has a differential impact by age; gender; region; socioeconomic status; mother's education level; and baseline EGRA, EGMA, and ZAT attainment. For noncategorical moderators, we will construct 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 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 analyses, we will present information on 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 (American Institutes for Research, 2016). 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 model. In this 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: the first regression will explain the treatment variable using the treatment assignment, and the second regression will explain the outcome variables with the predicted treatment variable.

Baseline Results

Baseline Balance

To ensure the comparability of the treatment and control groups in terms of observable characteristics, we tested for balance between the two groups for a number of explanatory measures. We grouped explanatory measures into several categories based on the theory of change: (1) household background characteristics, (2) housing and sanitation, (3) asset ownership, (4) food security, and (5) early childhood development, (6) ZAT, EGRA, EGMA, and (7) aspirations. We present balance tables for each of these categories below.

We found almost no statistically significant differences in demographic characteristics between treatment and control households. Both treatment and control households have a household size of slightly more than 6 people. Only a small minority of the sample (12% of the treatment households and 6% of the control households) lives within 1.5 kilometers of a government school, while close to 95% of the households live close to a community school. On average, both treatment and control households own more than 4.5 acres of land, and a little less than 50% of the households considers itself very poor or worse off than a year earlier. We did not find any statistically significant differences in these demographic characteristics except that treatment households live further away from school than control households. Only one of the statistically significant differences is larger than 0.29 standardized mean differences (SMDs). On average, treatment households live 0.21 kilometers further from school than control households. In addition, treatment households are 3 percentage points less likely than control households to benefit from the Social Cash Transfer program. This difference is statistically significantly at the 5% level but not more than 0.15 SMDs. Table 3 shows these baseline results

Table 3. Household-Level Background Characteristics

Variables	Control		Treatment		T-C Diff	Diff SE	*p*-value	Std. Mean Difference
	Mean	*N*	Mean	*N*				
Household size	6.18	1,007	6.28	888	0.12	0.14	0.37	0.06
Distance from school (km)	0.69	997	0.90	881	0.22	0.10	0.03*	0.29
Government schools within 1.5 km	0.12	1,007	0.06	888	-0.06	0.05	0.21	-0.20
Community schools within 1.5 km	0.93	1,007	0.94	888	0.01	0.05	0.90	0.02
Private schools within 1.5 km	0.00	1,007	0.00	888	0.00	0.00	0.31	0.10
Other schools within 1.5 km	0.00	1,007	0.01	888	0.01	0.00	0.09	0.14

Variables	Control		Treatment		T-C Diff	Diff SE	*p*-value	Std. Mean Difference
	Mean	*N*	Mean	*N*				
Agricultural land owned (acres)	4.51	894	4.85	759	0.35	0.30	0.25	0.09
Received benefit from Social Cash Transfer	0.06	1,007	0.03	888	-0.03	0.01	0.02*	-0.15
Household considers itself very poor	0.49	1,007	0.48	888	-0.01	0.03	0.75	-0.02
Household considers itself worse off than one year ago	0.39	1,007	0.40	888	0.02	0.03	0.54	0.04

Note. Standard errors (SE) clustered at school level. Includes district-level fixed effects. *, **, and *** indicate significance at the 90, 95, and 99 percent confidence levels, respectively. "T-C diff" refers to Treatment mean minus Control mean.

In addition, we found no statistically significant differences between treatment and control households across housing, water, and sanitation characteristics. On average, approximately 75% of the households rely on a borehole as their main source of drinking water, and 85% of the households use a torch as their main source of light. In addition, almost all (99%) of the households use firewood as their main source of energy for cooking. These data demonstrate the lack of access to electricity of the target group of the intervention. Furthermore, almost 45% of the households live under an iron-sheet roof, while roughly 85% of the households have access to a pit latrine. These findings are presented in Table 4.

Table 4. Housing, Water, and Sanitation

Variables	Control		Treatment		T-C Diff	Diff SE	*p*-value	Std. Mean Difference
	Mean	N1	Mean	N2				
Distance of main water source from household	0.63	1,000	0.81	873	0.19	0.10	0.07	0.14
Treats drinking water	0.06	1,007	0.06	888	0.01	0.02	0.76	0.02
Connected to electricity	0.00	1,007	0.00	888	0.00	0.00	0.79	0.01
Iron-sheet roof	0.44	1,007	0.43	888	-0.02	0.05	0.69	-0.04
Finished walls	0.34	1,007	0.35	888	0.00	0.05	0.93	-0.01
Main source of water: borehole	0.80	1,007	0.73	888	-0.07	0.07	0.32	-0.17
Main source of light: torch	0.85	1,007	0.85	888	0.01	0.02	0.75	0.02

Variables	Control		Treatment		T-C Diff	Diff SE	*p*-value	Std. Mean Difference
	Mean	N1	Mean	N2				
Main source of cooking energy: firewood	0.99	1,007	0.99	888	0.00	0.01	0.90	-0.01
Main cooking device: brick/stone stand on open fire	0.98	1,007	0.99	888	0.01	0.01	0.53	0.06
Has access to latrine	0.86	1,007	0.83	888	-0.03	0.03	0.24	-0.10

Note. Standard errors (SE) clustered at school level. Includes district-level fixed effects. *, **, and *** indicate significance at the 90, 95, and 99 percent confidence levels, respectively. "T-C diff" refers to Treatment mean minus Control mean.

The data on asset ownership again show little evidence for statistically significant differences between treatment and control households and show the high levels of poverty of the target group. We found a statistically significant difference between treatment and control households only in the proportion of children who have a pair of shoes. Children in treatment households are 8 percentage points (or 0.17 SMDs) more likely to own a pair of shoes. However, we found no other statistically significant differences in the ownership of assets. Of the target group, less than 20% own beds, less than 50% own shoes for their children, and less than 5% own a television. Approximately 65% of the households own a mosquito net and 35% of the households own a mobile phone. These findings are depicted in Table 5.

Table 5. Asset Ownership

Variables	Control		Treatment		T-C Diff	Diff SE	p-value	Std. Mean Difference
	Mean	N1	Mean	N2				
Child has a blanket (shared or owned)	0.82	1,006	0.82	888	0.01	0.04	0.89	0.01
Child has a pair of shoes	0.46	1,007	0.38	887	-0.08	0.03	0.01**	-0.17
Child has at least 2 sets of clothes	0.89	1,007	0.85	888	-0.04	0.02	0.11	-0.11
Asset quantity: bed	0.19	1,007	0.19	888	0.01	0.05	0.83	0.01
Asset quantity: mattress	0.30	1,007	0.29	888	-0.01	0.04	0.89	-0.01
Asset quantity: mosquito net	0.65	1,007	0.63	888	-0.03	0.08	0.68	-0.03

Variables	Control		Treatment		T-C Diff	Diff SE	p-value	Std. Mean Difference
	Mean	N1	Mean	N2				
Asset quantity: table (dining)	0.21	1,007	0.22	888	0.01	0.04	0.87	0.01
Asset quantity: lounge suite/sofa	0.10	1,007	0.08	888	-0.02	0.03	0.35	-0.04
Asset quantity: radio/stereo	0.22	1,007	0.21	888	-0.01	0.03	0.69	-0.02
Asset quantity: television	0.05	1,007	0.04	888	-0.01	0.01	0.45	-0.04
Asset quantity: DVD/VCR player	0.02	1,007	0.02	888	0.00	0.01	0.79	0.01
Asset quantity: cellular phone	0.35	1,007	0.35	888	0.01	0.05	0.88	0.01
Asset quantity: electric iron	0.00	1,007	0.00	888	0.00	0.00	0.15	0.07
Asset quantity: watch	0.01	1,007	0.01	888	0.00	0.00	0.38	0.04
Asset quantity: clock	0.01	1,007	0.01	888	0.00	0.00	0.98	-0.00
Asset quantity: refrigerator	0.00	1,007	0.00	888	0.00	0.00	0.32	-0.04
Asset quantity: hand saw	0.03	1,007	0.03	888	0.00	0.01	0.92	-0.01
Asset quantity: axe	1.01	1,007	1.10	888	0.09	0.06	0.12	0.11
Asset quantity: pick	0.14	1,007	0.17	888	0.03	0.03	0.27	0.07
Asset quantity: hoe	2.98	1,007	3.10	888	0.09	0.11	0.44	0.05
Asset quantity: hammer	0.24	1,007	0.27	888	0.03	0.03	0.40	0.05
Asset quantity: shovel/spade	0.13	1,007	0.13	888	0.00	0.02	0.81	0.01
Asset quantity: fishing net	0.00	1,007	0.00	888	0.00	0.00	0.31	-0.07
Asset quantity: plough	0.46	1,007	0.45	888	0.00	0.06	0.97	0.00
Asset quantity: animal cart	0.14	1,007	0.13	888	-0.02	0.02	0.41	-0.04
Asset quantity: bicycle	0.50	1,007	0.49	888	-0.01	0.03	0.69	-0.02
Asset quantity: motorcycle	0.02	1,007	0.02	888	-0.01	0.01	0.49	-0.05
Asset quantity: canoe	0.00	1,007	0.00	888	0.00	0.00	0.31	-0.05
Asset quantity: oxen	0.83	1,007	0.82	888	0.00	0.13	0.99	0.00
Asset quantity: solar panel	0.33	1,007	0.33	888	0.01	0.04	0.88	0.01

Note. Standard errors (SE) clustered at school level. Includes district-level fixed effects. *, **, and *** indicate significance at the 90, 95, and 99 percent confidence levels, respectively. "T-C diff" refers to Treatment mean minus Control mean.

We also found no statistically significant differences between treatment and control households with respect to food security. For these measures, we primarily relied on the Household Food Insecurity Access Scale, which is an index comprised of responses based on individual questions about food security (based on the Food and Nutrition Technical Assistance III Project guidelines). There are questions, among others, about the number of meals per day; whether households ate meat in the last month; whether households go to bed hungry and, if so, how often; and the variation of the food consumed. We include average values for treatment and control households for each of these questions (as well as the scale) in Table 6 below. The scale ranges from 0 to 27, with higher values indicating greater levels of food insecurity.

Table 6. Food Security

Variables	Control		Treatment		T-C Diff	Diff SE	*p*-value	Std. Mean Difference
	Mean	N1	Mean	N2				
Household Food Insecurity Access Scale	13.18	1,007	13.49	888	0.29	0.51	0.58	0.04
Meals per day	2.15	1,007	2.13	888	-0.02	0.04	0.68	-0.03
No meat in last month	0.24	1,007	0.26	888	0.03	0.04	0.50	0.06
Vegetables more than 5 times in past week	0.85	1,007	0.87	888	0.01	0.03	0.71	0.03
Worried often about not having enough food	0.28	1,007	0.31	888	0.03	0.03	0.36	0.06
Unable to eat preferred types of food	0.32	1,007	0.35	888	0.03	0.04	0.48	0.05
Ate smaller meals often	0.21	1,007	0.23	888	0.01	0.03	0.64	0.04
Ate fewer meals per day often	0.21	1,007	0.20	888	-0.01	0.03	0.83	-0.02
Had no food to eat often	0.12	1,007	0.13	888	0.01	0.02	0.72	0.03
Went to bed hungry often	0.08	1,007	0.08	888	0.00	0.02	0.89	0.01
Went 24 hours with no food often	0.08	1,007	0.07	888	0.00	0.02	0.99	0.00
Ate limited variety of food often	0.35	1,007	0.37	888	0.02	0.04	0.67	0.04

Variables	Control		Treatment		T-C Diff	Diff SE	*p*-value	Std. Mean Difference
	Mean	N1	Mean	N2				
Children under 5 often ate unhealthy food	0.14	1,007	0.19	888	0.04	0.03	0.11	0.12
Children under 5 often did not have enough food	0.09	1,007	0.14	888	0.04	0.02	0.07	0.14

Note. Standard errors (SE) clustered at school level. Includes district-level fixed effects. *, **, and *** indicate significance at the 90, 95, and 99 percent confidence levels, respectively. "T-C diff" refers to Treatment mean minus Control mean.

Child-level characteristics, including age, the language of the child, and early childhood cognition, also appear to be similar across treatment and control households. We found few statistically significant differences in this area. For example, the results suggest that children in the control group are approximately one month older than children in the treatment group. This difference is statistically significant but smaller than 0.25 SMDs. We did not find other statistically significant differences. Treatment and control children show similar outcomes in terms of the percentage of parents reporting that their child has the ability to read, add numbers, sit still, identify shapes and colors, draw a circle, kick a ball, had a good attention span, and other early childhood development characteristics. We also did not find statistically significant differences in the gender of the child. Furthermore, almost all of the children speak either Nsenga or Chewa, which are both dialects of Nyanja. These results are depicted in

Table 7 below.

Table 7. Early Childhood Characteristics

Variables	Control		Treatment		T-C Diff	Diff SE	*p*-value	Std. Mean Difference
	Mean	N1	Mean	N2				
Age of child in years	7.70	1,007	7.57	888	-0.13	0.06	0.04*	-0.13
Child is female	0.48	1,007	0.45	888	-0.03	0.02	0.16	-0.06
Child is biological child of respondent	0.85	1,007	0.88	888	0.03	0.02	0.16	0.08
Main language of child: Nsenga	0.67	1,007	0.57	888	-0.06	0.07	0.38	-0.13
Main language of child: Chewa	0.32	1,007	0.42	888	0.07	0.07	0.32	0.14
Child can read	0.01	1,007	0.01	888	0.00	0.01	0.80	-0.02
Child can add numbers	0.25	1,007	0.25	888	0.00	0.04	0.90	-0.01
Child pays attention well	0.97	1,006	0.98	886	0.00	0.01	0.66	0.02
Child can sit still for at least 5 minutes	0.95	999	0.93	885	-0.02	0.01	0.11	-0.10
Child can identify shapes and colors	0.48	993	0.43	879	-0.05	0.03	0.13	-0.09
Child can count to 20 or higher	0.38	990	0.35	876	-0.03	0.03	0.31	-0.06
Child uses words to describe feelings	0.67	974	0.67	861	0.00	0.03	0.90	0.01
Child invites other children to play	0.96	994	0.95	883	-0.01	0.01	0.34	-0.05
Child has frequent conflicts with other children	0.38	987	0.38	871	0.00	0.04	0.89	-0.01
Child is angry frequently	0.33	990	0.31	882	-0.02	0.04	0.62	-0.04
Child can draw a circle	0.81	991	0.78	875	-0.04	0.03	0.25	-0.09
Child can stack objects	0.89	978	0.87	875	-0.02	0.02	0.44	-0.06
Child can kick a ball	0.97	992	0.97	877	-0.01	0.01	0.45	-0.04
Child can jump on one foot	0.97	977	0.96	863	-0.01	0.01	0.22	-0.08

Note. Standard errors (SE) clustered at school level. Includes district-level fixed effects. *, **, and *** indicate significance at the 90, 95, and 99 percent confidence levels, respectively. "T-C diff" refers to Treatment mean minus Control mean.

In terms of caregiver aspirations, we did not find statistically significant differences between treatment and control households. Both treatment and control caregivers reported a preferred age-at-marriage of around 25 years old. Furthermore, both treatment and control caregivers assessed the probability that their child will graduate from Grade 12 slightly above 75%, on average. In addition, approximately 97% of the treatment and control caregivers reported a preference for their child to achieve an education of Grade 12 or higher. Finally, we also found no statistically significant differences between treatment and control caregivers in their estimation of the bride price they would receive (for daughters) or should provide (for sons). These data are highlighted in Table 8 below.

Table 8. Parental Aspirations

Variables	Control		Treatment		T-C Diff	Diff SE	*p*-value	Std. Mean Difference
	Mean	N1	Mean	N2				
Age caregiver would like child to marry at age…	25.20	996	25.01	865	-0.20	0.30	0.51	-0.04
Caregiver's assessment of the probability the child will graduate from Grade 12	0.75	960	0.77	850	0.02	0.02	0.28	0.08
Caregiver wants child to achieve Grade 12 or higher	0.97	1,007	0.97	888	0.00	0.01	0.94	0.00
Bride price estimate (female) [Kwacha]	1,329.29	471	1,349.93	385	31.98	181.53	0.86	0.02
Bride price estimate (male) [Kwacha]	924.67	521	922.31	477	5.74	129.92	0.96	0.00

In terms of the ZAT, we also did not find statistically significant differences between treatment and control children. Treatment and control children scored between 65% and 70% on the easiest module of the ZAT. The test gets more difficult with each module, and children's performances decreased to a score of 37% on the second module, 21% on the third module, and 18% on the fourth module. None of modules had a statistically significant difference in performance across children in the treatment catchment areas and control catchment area. These results are presented in

Table 9.

Table 9. Zambian Achievement Test

Variables	Control		Treatment		T-C Diff	Diff SE	*p*-value	Std. Mean Difference
	Mean	N1	Mean	N2				
Zambian Achievement Test 1	0.70	1,007	0.66	888	-0.04	0.02	0.12	-0.12
Zambian Achievement Test 2	0.37	1,007	0.34	888	-0.03	0.03	0.31	-0.07
Zambian Achievement Test 3	0.21	1,007	0.21	888	0.01	0.03	0.80	0.02
Zambian Achievement Test 4	0.18	1,007	0.19	888	0.01	0.03	0.83	0.02

Note. Standard errors (SE) clustered at school level. Includes district-level fixed effects. *, **, and *** indicate significance at the 90, 95, and 99 percent confidence levels, respectively. "T-C diff" refers to Treatment mean minus Control mean.

The EGRA data show that treatment and control children both scored low on the assessment test. Only approximately 15% of the children were able to move beyond the orientation to print subtask. This is unsurprising considering that none of the children was enrolled in school at the time of the assessment. These results are in alignment with potential for floor effects in the estimation of Impact Network's eSchool 360 model on early grade reading outcomes. This potential for floor effects shows the importance of including the ZAT. These results are presented in Table 10 below.

Table 10. Early Grade Reading Assessment

Variables	Control		Treatment		T-C Diff	Diff SE	*p*-value	Std. Mean Difference
	Mean	N1	Mean	N2				
Oral vocabulary	0.62	1,007	0.60	888	-0.02	0.02	0.20	-0.09
Orientation to print	0.14	1,007	0.15	888	0.01	0.02	0.69	0.03
Letter sound knowledge	0.01	1,007	0.01	888	0.00	0.00	0.83	-0.02
Nonword decoding	0.00	1,007	0.00	888	0.00	0.00	0.17	-0.07
Oral passage reading 1	0.00	1,007	0.00	888	0.00	0.00	0.70	-0.02
Reading comprehension 1	0.00	1,007	0.00	888	0.00	0.00	0.86	0.01
Oral passage reading 2	0.00	1,007	0.00	888	0.00	0.00	0.31	0.06
Reading comprehension 2	0.00	1,007	0.00	888	0.00	0.00	0.31	0.07

Variables	Control		Treatment		T-C Diff	Diff SE	*p*-value	Std. Mean Difference
	Mean	N1	Mean	N2				
Listening comprehension	0.35	1,007	0.35	888	0.00	0.02	0.96	0.01

Note. Standard errors (SE) clustered at school level. Includes district-level fixed effects. *, **, and *** indicate significance at the 90, 95, and 99 percent confidence levels, respectively. "T-C diff" refers to Treatment mean minus Control mean.

Similar to the EGRA results, both treatment and control children scored relatively low on the EGMA assessment. We did find some differences between treatment and control children, however. Specifically, treatment children scored lower on the first addition questions subtask than did control children. This difference is statistically significant at the 5% level. Both differences are smaller than 0.20 SMDs, however. As with EGRA, the EGMA scores indicate potential for floor effects in the estimation of Impact Network's eSchool 360 model on early grade math outcomes. These results are presented in Table below.

Table 11. Early Grade Math Assessment

Variables	Control		Treatment		T-C Diff	Diff SE	*p*-value	Std. Mean Difference
	Mean	N1	Mean	N2				
Oral counting	0.78	1,007	0.76	888	-0.02	0.02	0.29	-0.07
Rational counting	0.12	1,007	0.11	888	-0.01	0.01	0.06	-0.11
Number recognition	0.07	1,007	0.05	888	-0.01	0.01	0.15	-0.11
Quantity discrimination	0.09	1,007	0.07	888	-0.02	0.01	0.07	-0.11
Pattern completion	0.03	1,007	0.03	888	0.00	0.01	0.93	0.01
Word problems	0.08	1,007	0.06	888	-0.02	0.01	0.10	-0.10
Addition questions 1	0.13	1,007	0.08	888	-0.05	0.02	0.04*	-0.18
Addition questions 2	0.01	1,007	0.01	888	0.00	0.00	0.11	-0.08
Subtraction questions 1	0.11	1,007	0.07	888	-0.04	0.02	0.05	-0.16
Subtraction questions 2	0.01	1,007	0.01	888	0.00	0.00	0.52	-0.03

Note. Standard errors (SE) clustered at school level. Includes district-level fixed effects. *, **, and *** indicate significance at the 90, 95, and 99 percent confidence levels, respectively. "T-C diff" refers to Treatment mean minus Control mean.

Based on the EGRA and EGMA outcomes, we decided not to include factor analysis in this baseline report. Factor analysis serves to understand what latent constructs the assessments are tapping into, but the low scores on more complex EGRA and EGMA subtasks limit our ability 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. We will, however, include factor analyses in the midline report because we anticipate higher scores on more complex EGRA and EGMA subtasks one year after the start of the school year.

Potential for Floor Effects

As noted earlier, we hypothesized that most students will score very low on their EGRA and EGMA tests in 2018. It is likely that short-term improvements will 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 Impact Network's model on EGRA and EGMA outcomes.

As discussed in the previous section, some of the first analyses on the full sample suggested that the children indeed scored very low on the EGRA and EGMA tests, but there is encouraging evidence that the ZAT and the oral vocabulary test follow approximately a normal distribution. These results suggest the potential for floor effects in the estimation of impacts on EGRA and EGMA outcomes. However, this concern is partially mitigated by the limited risk of floor effects in the estimation of model impacts on the ZAT and the oral vocabulary test outcomes. The distributions of the EGRA, EGMA, ZAT, and oral vocabulary test outcomes are highlighted in Figures 6–9 below.

Figure 6. Distribution of EGMA Scores

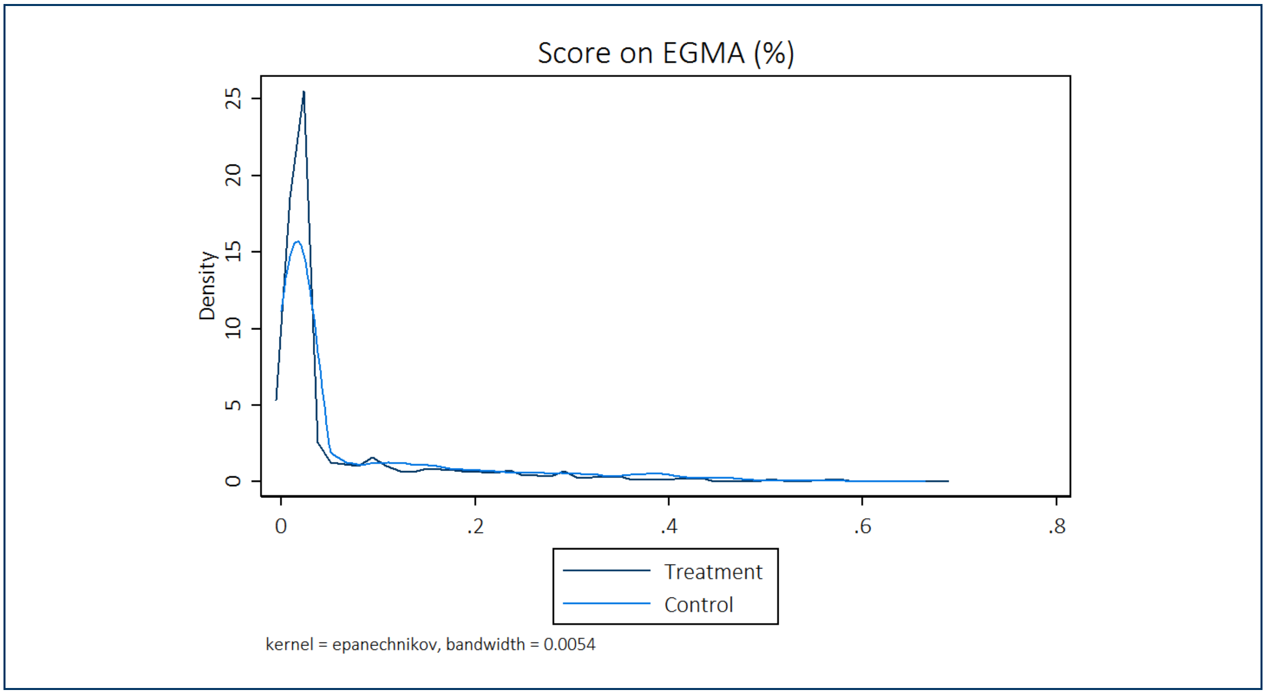


Figure 7. Distribution of EGRA Scores

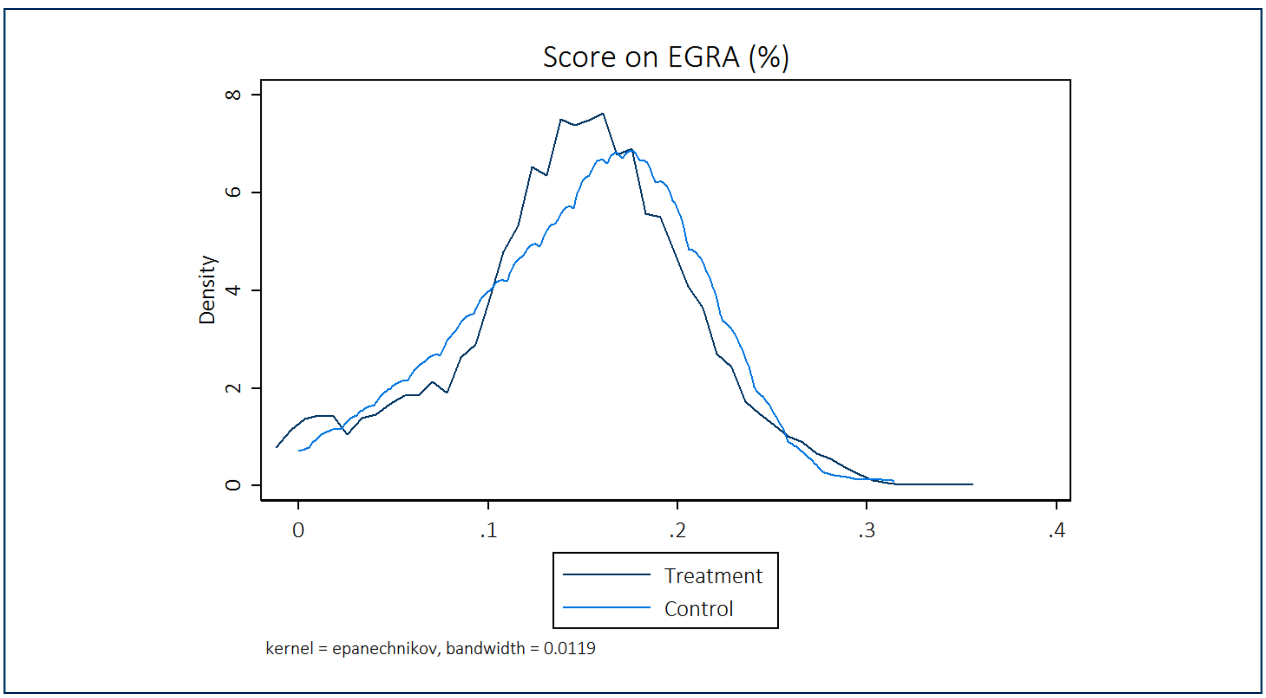


Figure 8. Distribution of ZAT Scores

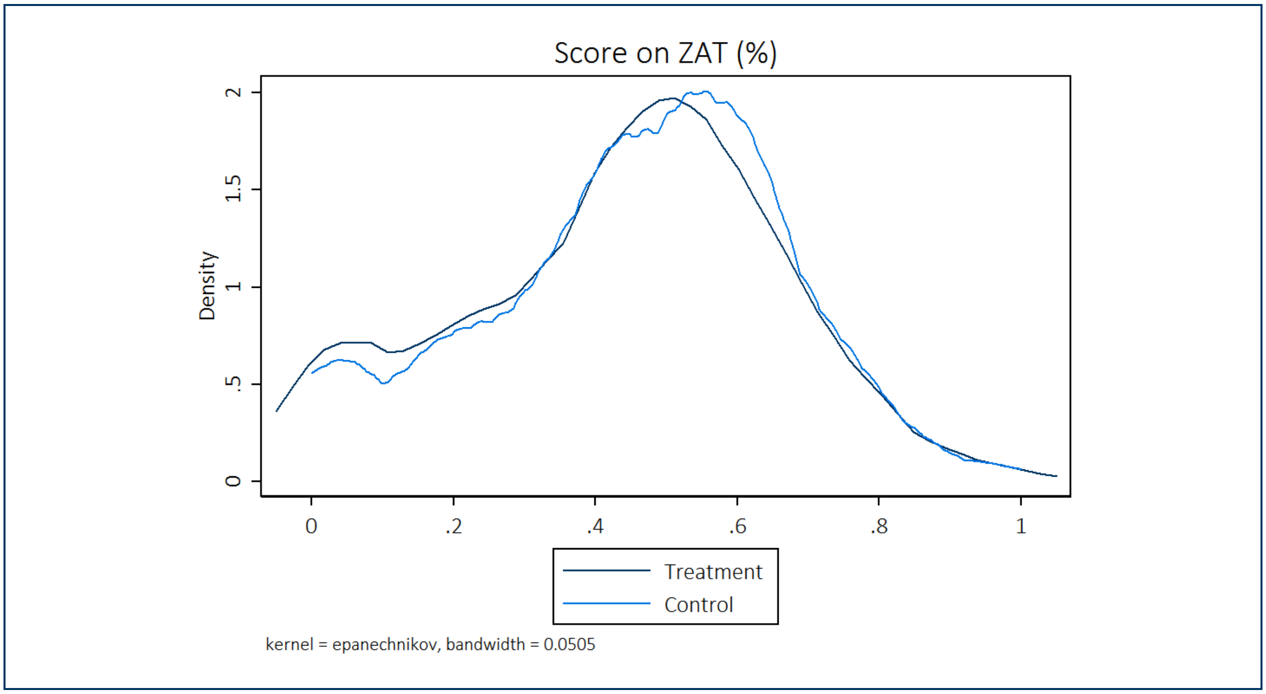
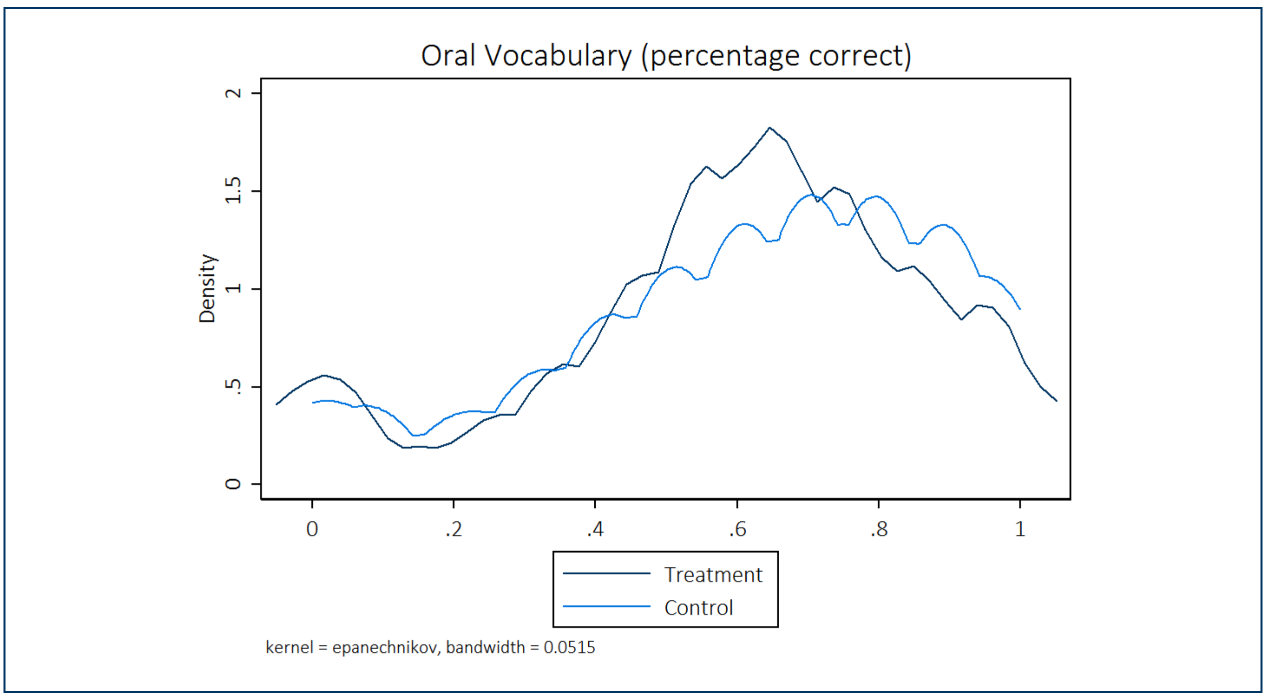


Figure 9. Distribution of Oral Vocabulary Test



Predictors of Baseline Assessment Tests

The baseline data also provide insights on the individual and household correlates of higher performance on assessments among young children in rural Zambia. We found few characteristics that are consistent correlates of high performance across the different assessments. The only relationship that is significant and persistent across all assessments and regression specifications is a positive correlation between child age and achievement. The results of these regressions are presented in Table 12: for each assessment, we examined the relationship excluding and including caregiver survey responses on child ability. The specification that excludes caregiver survey responses serves as a base estimate and focuses exclusively on static demographic and child characteristics that may be associated with assessment performance. The subsequent specifications add self-reported ability variables to assess the robustness of our estimates and to examine the extent to which these additional variables explain child assessment performance. It is worth noting that these regression results are not designed with causal inference in mind: rather, they examine correlations between assessment performance and child characteristics. Similarly, we hesitate to place too much weight on a single significant coefficient as many covariates are included and simple chance would dictate that some are significant; instead, we focus on general trends across the regression results.

For example, children score 3-5 percentage points better on the ZAT for each additional year of age (Table 12, columns 1a and 1b). None of the children in the sample had attended school at the time of the baseline survey, so these results indicate that children in rural Zambia develop the cognitive skills associated with shape recognition outside of school. The kernel density plot of ZAT scores by age (Figure 10) indicate that this difference is not only driven by particularly high-performing older students but also that older children score higher across the achievement spectrum. A similarly positive relationship between age and assessment performance is indicated in the various test scores (columns 2a through 5b), although the improvement associated with an additional year of age **is** smaller for other assessments: EGRA increases by about 1 percentage point, EGMA and listening comprehension both increase by 1-2 percentage points, and oral vocabulary increases by 2-4 percentage points.

We found limited evidence of other consistent correlates of high assessment performance. Child gender; child care assets such as having shoes, a blanket, or multiple sets of clothing; distance from school; and high self-reported poverty are generally not related to child performance. Perhaps, surprisingly, caregiver education is largely unrelated to child performance: the estimated coefficients for caregiver education in the ZAT, EGMA, and listening comprehension regressions, are precisely estimated to be zero. Caregiver education is statistically significantly associated with EGRA outcomes, but the magnitude of the relationship is small: 5 additional years of education is associated with only a 1 percentage point increase in EGRA performance. The one exception to the nominal relationship between caregiver education and child performance is oral vocabulary, in which each additional year of education is associated with an almost 1 percentage point increase in assessment performance. This increase is sizeable and equates to the effect of 0.25 additional years of age.

We found no relationship between household assets and children's performance on the ZAT, EGMA, or listening comprehension assessments, but children in wealthier households (as measured by an asset index) perform marginally better on the EGRA and 1 percentage point better on the oral vocabulary assessment. We found similar mixed evidence for household food security: a 1 standard deviation increase in the food security index (6.5 points) is associated with a 3 percentage point increase in ZAT performance, a gain equivalent to adding 0.6 years of age. It is possible that the ZAT measures a basic cognitive development that may be hampered by limited food intake.

Three caregiver-reported child abilities are significantly correlated with improved child performance on the various assessments: whether a child can draw a circle, identify shapes and colors, and count to 20. These results are presented in columns b and c for each of the assessments in Table and are presented as additions to the base regressions because they provide interesting information on the correlations but may be more likely to suffer from endogeneity as, in some cases, they represent self-reported measures of similar outcomes. Each of these variables, however, is strongly correlated with higher performance on the assessments. The assessments appear to be picking up different cognitive skills: the ZAT comprises multiple pattern recognition components; it is encouraging that a child's ability to identify shapes and colors as well as to draw a circle appear as positive and significant. Similarly, the ability to count to 20 has a positive and significant impact on the EGMA performance, which is designed to measure numeracy skills. Of note in the results is the fact that a small number of children (20) whose caregivers indicated that their child can read do not perform better on the EGRA examination, suggesting that self-reported reading may not give an accurate representation of a child's reading skills.

Overall, we found only a few consistent predictors of child assessment performance using characteristics measured during the baseline. Our regression estimates show that including a battery of child and household characteristics explains only 20% of the observed variation in assessment performance.

Table 12. Demographic and Household Correlates of Test Performance

	Zambian Achievement Test (percent correct)		Early Grade Reading Assessment (percent correct)			Early Grade Math Assessment (percent correct)			Listening Comprehension (percent correct)		Oral Vocabulary (percent correct)	
Variables	(1a)	(1b)	(2a)	(2b)	(2c)	(3a)	(3b)	(3c)	(4a)	(4b)	(5a)	(5b)
Age of child in years	0.049***	0.035***	0.012***	0.008***	0.008***	0.022***	0.023***	0.014***	0.023***	0.012**	0.038***	0.025***
	(0.005)	(0.005)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.005)	(0.006)	(0.007)	(0.008)
Child is female	0.012	0.007	0.003	0.002	0.002	0.007*	0.006	0.004	0.010	0.008	0.006	0.000
	(0.010)	(0.009)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.008)	(0.009)	(0.013)	(0.013)
Child's assets (blanket, shoes, clothes)	0.013*	0.009	-0.001	-0.002	-0.001	0.005	0.007*	0.005	-0.007	-0.009	0.001	0.000
	(0.007)	(0.007)	(0.002)	(0.002)	(0.002)	(0.004)	(0.004)	(0.004)	(0.007)	(0.007)	(0.010)	(0.010)
Distance from school (Km)	0.004	0.003	-0.005*	-0.005	-0.005*	0.004	0.004	0.004	0.004	0.005	-0.031**	-0.032**
	(0.009)	(0.010)	(0.003)	(0.003)	(0.003)	(0.005)	(0.005)	(0.005)	(0.008)	(0.008)	(0.015)	(0.015)
Female caregiver's years of education	0.002	0.001	0.002**	0.002**	0.001*	0.000	0.000	-0.000	-0.000	-0.001	0.008***	0.007**
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)
Male caregiver's years of education	0.000	0.000	-0.000	-0.000	0.000	0.001	0.000	0.000	-0.000	-0.000	-0.001	0.000
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)
HH asset index	0.007	0.007	0.003*	0.003*	0.003	0.003	0.002	0.001	-0.000	-0.000	0.012*	0.013*
	(0.004)	(0.004)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)	(0.007)	(0.007)
Household self-identifies as very poor	0.005	-0.000	-0.004	-0.005	-0.005	-0.009	-0.008	-0.009	0.003	0.002	-0.020	-0.025
	(0.013)	(0.013)	(0.004)	(0.004)	(0.004)	(0.007)	(0.007)	(0.007)	(0.013)	(0.013)	(0.018)	(0.018)
HH Food Insecurity Access Scale	-0.005***	-0.004***	0.000	0.000	0.000	-0.000	-0.000	0.000	-0.000	0.000	0.002*	0.004***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Household size	0.008***	0.007**	0.001	0.001	0.001	0.003**	0.002	0.001	0.004	0.003	0.005	0.003
	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)	(0.004)
Child can read		-0.030			0.003		0.006	-0.009		0.024		-0.024
		(0.046)			(0.014)		(0.021)	(0.024)		(0.059)		(0.048)
Child can add numbers		0.030*		0.005	0.005			0.010		0.026*		0.027
		(0.016)		(0.004)	(0.004)			(0.007)		(0.013)		(0.019)
Child can draw a circle		0.047***		0.012***	0.011**			0.017**		0.011		0.044**
		(0.016)		(0.004)	(0.004)			(0.007)		(0.014)		(0.020)
Child can identify shapes and colors		0.044***		0.016***	0.015***			0.019***		0.022*		0.063***
		(0.013)		(0.003)	(0.004)			(0.006)		(0.011)		(0.017)
Child can count to 20 or higher		0.058***		0.013***	0.013***			0.050***		0.040***		0.024
		(0.013)		(0.004)	(0.004)			(0.006)		(0.012)		(0.017)
Child uses words to describe feelings		-0.011			0.006		0.000	-0.003		0.001		0.030
		(0.013)			(0.004)		(0.006)	(0.005)		(0.010)		(0.020)
R-sqr												
N												

Notes: *, **, and *** indicate significance at the 90, 95, and 99 percent confidence levels, respectively. Standard errors clustered at the school catchment area level are reported in parentheses. All specifications include school catchment area fixed effects. Each column is a different regression.

Figure 10. Distribution of ZAT Score by Child Age

Conclusion

The baseline results for the evaluation of Impact Network's eSchool 360 model demonstrate that the cluster-RCT was successful in creating equivalence in observable characteristics between treatment and control households. We did not find evidence for systematic statistically significant differences. Furthermore, almost none of the statistically significant differences at baseline are larger than 0.3 standard deviations. This finding indicates that the randomization will enable AIR to make causal claims about the short-term effects of the Impact Network's eSchool 360 model after the midline data collection and analysis (one year after the introduction of the model) and after the endline data collection and analysis (three years after the introduction of the model).

In addition, the analyses suggest the potential for floor effects in the estimation of program impacts on EGRA and EGMA outcomes, but we found encouraging evidence that the ZAT and the oral vocabulary test follow approximately a normal distribution. Unsurprisingly, the children in our sample scored very low on the EGRA and EGMA assessments at baseline. The midline and endline analyses will have to determine whether students improve sufficiently to mitigate concerns about floor effects in the estimation of program impacts on EGRA and EGMA outcomes. In the presence of floor effects, we will have to rely primarily on the estimation of the impacts of Impact Network's eSchool 360 model on the ZAT and the oral vocabulary test, in addition to the intermediate outcomes (e.g., school attendance and enrollment) of the theory of change.

We plan to collect midline data in November–December 2018. This will again include the collection of EGRA, EGMA, and ZAT data, as well as the collection of household-level survey data on school enrollment and attendance, student-level aspirations, and parental-level aspirations. In addition, we will collect qualitative data in two schools in each of the three treatment districts of Katete, Petauke, and Sinda. We will use three primary approaches to qualitative data collection for this evaluation: key informant interviews with community leaders, eSchool 360 model staff, teachers, and students; focus group discussions with students and parents; and classroom observations. We also plan to collect cost data to inform the cost-effectiveness analysis from November–December 2018.

References

American Institutes for Research. (2016). Zambia's Child Grant Program: 48-Month Impact Report.

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

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

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., … Wasserman, M. (2016). *Mainstreaming an effective intervention: Evidence from randomized evaluations of "Teaching at the Right Level" in India*. New York: National Bureau of Economic Research. doi: 10.3386/w22746

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.

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

Central Statistical Office Zambia, Ministry of Health Zambia, and ICF International (2014) Zambia Demographic and Health Survey 2013-14. Rockville, Maryland, USA: Central Statistical Office/Zambia, Ministry of Health/Zambia, and ICF International.

Central Statistics Office Zambia (2013). 2010 Census of Population and Housing: Population and Demographic Projections 2011-2035. Retrieved from http://www.zamstats.gov.zm/index.php/publications/category/11-2010-census

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 meta-analysis of rigorous impact evaluations* (Doctoral dissertation, Columbia University.) 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 (IDB Working Paper No. IDB-WP-304). Washington, DC: Inter-American Development Bank.

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.

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.

DeStefano, J. (2006). Meeting EFA: Zambia Community Schools. EQUIP2. Washington, DC: USAID. Retrieved from https://www.epdc.org/sites/default/files/documents/Zambia%20Community%20Schools.pdf

Dhaliwal, I., Duflo, E., Glennerster, R., & Tulloch, C. (2011). *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

Duflo, E., Kremer, M., & Glennerster, R. (2007). Using randomization in development economics research: A toolkit. (Discussion No. 6059). Centre for Economic Policy Research. Retrieved from http://economics.mit.edu/files/806

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. Newbury Park, CA: SAGE.

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. (2015, October). *Engineering scale up in humanitarian innovations missing middle*. Paper presented at the IEEE Global Humanitarian Technology Conference. doi: 10.1109/GHTC.2015.7343963

Ministry of General Education. (2014). Educational Statistical Bulletin. Zambia.

Muralidharan, K., Singh, A., & Ganimian, A. J. (2017). Disrupting education? Experimental evidence on technology-aided instruction in India (NBER Working Paper No. w22923). Cambridge, MA: National Bureau of Economic Research. doi:10.3386/222923

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

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.

Power, T., Gater, R., Grant, C., & Winters, N. (2014). *Educational technology topic guide*. HEART topic guides. London: Health & Education Advice & Resource Team (HEART).

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.

Schling, M., & Winters, P. (2017). Computer-Assisted Instruction for Child Development: Evidence from an Educational Programme in Rural Zambia. The Journal of Development Studies, 1-16.

Southern and Eastern Africa Consortium for Monitoring Educational Quality. (2010–12). SACMEQ Indicators. Retrieved from http://www.sacmeq.org/SACMEQIndicators

Snilstveit, B., Stevenson, J., Menon, R., Phillips, D., Gallagher, E., Geleen, M., ... & Jimenez, E. (2016). The impact of education programmes on learning and school participation in low-and middle-income countries.

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.

United Nations. (2017) The sustainable development goals report. New York: United nations Statistics Division. Retrieved from https://unstats.un.org/sdgs/report/2017

United Nations Educational, Scientific, and Cultural Organisation. (2014). Learning: Achieving Quality for All, EFA Global Monitoring Report 2013/14.

Winters, C., Schling, M., & Winters, P. (2013) Cost effectiveness report for the Impact Network eLearning education program in rural Zambia: Final report prepared for the Impact Network. Unpublished manuscript.

World Bank. World Development report 2018: Learning to realize education's promise. (2018). Washington, DC: International Bank for Reconstruction and Development/The World Bank.

Established in 1946, the American Institutes for Research (AIR) is an independent, nonpartisan, not-for-profit organization that conducts behavioral and social science research on important social issues and delivers technical assistance, both domestically and internationally, in the areas of education, health, and workforce productivity.

MAKING RESEARCH **RELEVANT**

AMERICAN INSTITUTES FOR RESEARCH
1000 Thomas Jefferson Street NW
Washington, DC 20007-3835 | 202.403.5000
www.air.org

LOCATIONS

Domestic: Washington, DC (HQ) | Monterey, Sacramento, and San Mateo, CA | Atlanta, GA | Honolulu, HI | Chicago and Naperville, IL Indianapolis, IN | Metairie, LA | Waltham, MA | Frederick and Rockville, MD | Chapel Hill, NC | New York, NY | Columbus, OH | Cayce, SC Austin, TX | Reston, VA

International: El Salvador | Ethiopia | Haiti | Honduras | Zambia

4117_04/18