IMAGINE

THE FUTURE OF LEARNING

Tablet-based Learning for Foundational Literacy and Math: An 8-month RCT in Malawi

Dr. Karen Levesque

Dr. Sarah Bardack

Imagine Worldwide

Dr. Antonie Chigeda

University of Malawi Chancellor College

January 31, 2020

* * * * *

Imagine's Research Advisory Group is composed of highly respected experts representing a range of academic fields. The members provided critical input at key stages of the research design, execution, analysis, and writing. Imagine is grateful for the diverse expertise and valuable input of the members. The advisory group reviewed and approved this final report.

Research Advisory Group Members

Helen Abadzi, Researcher, Education and Cognitive Neuroscience, University of Texas at Arlington

Bidemi Carroll, Senior Education Research Analyst, RTI's International Education Group

John Pane, Senior Scientist, RAND research firm

Dave Paunesku, Co-founder and Executive Director, PERTS, Stanford University

Nicola Pitchford, Professor, Faculty of Science, University of Nottingham, United Kingdom

Mark Wilson, Professor, Graduate School of Education, University of California, Berkeley

Acknowledgements

This study was made possible by the efforts of many individuals and organizations. We would like to thank the children, teachers, parents, and school and community leaders in the study schools for their willingness to participate in the research and for their strong commitment to improving literacy and numeracy in their own communities and throughout Malawi, Africa, and the world. We would also like to thank onebillion and VSO Malawi for their years of dedication in developing the software and implementation program that were at the core of this study and the researchers at the University of Nottingham and University of Malawi for their prior research, which demonstrated the promise of the intervention and set the stage for this longer term efficacy trial.

We appreciate the support of our software, implementation, and research colleagues at onebillion, VSO Malawi, and the University of Malawi, who helped to ensure the quality of implementation of the intervention, without which we would not have seen the study's positive results. We also benefitted enormously at critical stages of the research from the expert advice of our research advisory group, including Helen Abadzi, University of Texas at Arlington; Bidemi Carroll, RTI International; John Pane, RAND Corporation; Dave Paunesku, Stanford University; Nicola Pitchford, University of Nottingham; and Mark Wilson, University of California at Berkeley.

Finally, we would like to thank our colleagues at Imagine for their vision, dedication, support, and patience, including Susan Colby, CEO; Jen Welsh, Partner; Rachel Boroditsky, Associate; and Kristen Stipe, Manager of Operations and Administration. And we thank Imagine's Board of Directors, for their innovativeness, critical questioning, and unwavering support, including Joe Wolf, Chairman of the Imagine Board; Chinezi Chinioke, Nova-Pioneer Education Group; Aishetu Fatima Dozie, Bossy Cosmetics; Michael Horn, Entangled Ventures; Amy Klement, Imaginable Futures; Rapelang Rabana, Rekindle Learning; and Rebecca Winthrop, Brookings Institution.

Without the active involvement and support of the above individuals and organizations, this study and the promising learning impacts that it demonstrates would not have been possible.

Table of Contents

Tablet-based Learning for Foundational Literacy and Math: An 8-month RCT in Malawi

Executive Summary

Imagine seeks to understand whether children with few education alternatives can become literate and numerate using child-directed, technology-enabled learning. To this end, we are building an evidence base for what works, for whom, and under what conditions, in schools and out-of-school settings, in different countries and languages, and with a small collection of promising software, starting with onebillion's *onecourse*.

Study Overview

Prior research on onebillion's literacy and numeracy applications used in the Malawi government primary schools showed promising results. However, the previous randomized controlled trials (RCTs) were conducted over short periods of time (8 and 14 weeks) to establish proof of concept. To help fill the gap in knowledge about the longer-term impacts of this intervention, we conducted an 8-month RCT during the 2018-19 school year to address the following primary research questions:

1. What are the impacts over standard instruction on literacy and numeracy outcomes of using onebillion's *onecourse* software in Chichewa for 40 minutes per day for 8 months?
2. What impact does attendance in the intervention have on learning outcomes?
3. How far do children progress toward reading fluency with comprehension and comparable numeracy skills (i.e., arithmetic fluency with number sense)?
4. How are subgroup characteristics such as school, gender, and age category associated with learning outcomes?

The 8-month study tested the efficacy of onebillion's literacy and numeracy applications in the Chichewa language delivered through the Unlocking Talent implementation model. The study represented an efficacy RCT using a non-clustered, blocked individual random assignment (BIRA) design. Two government primary schools were purposively selected for the study: one urban and one peri-urban school located in the capital region.[1] Conditions in the two communities and schools are challenging. Families in both communities are very low income and face food security issues and other poverty-related challenges. Neither school has electricity and class sizes at both schools are very large (up to 100 children).

All 674 eligible Standard 2 (grade 2) learners ages 6–10 were randomly assigned independently within the two schools to treatment and control groups. The treatment groups used either the literacy or math curriculum—not both—for 40 minutes per day in an effort to maximize time on task in each subject and to isolate the impact of the two applications. A schedule was established so that children in the treatment groups stepped out of different standard classes on different days of the week to attend the learning center. Thus, the intervention represented a supplement to normal instruction in the tablet subject.[2] The

[1] The schools were selected to represent an urban and a peri-urban (more rural) environment, to meet sample size requirements, and to meet the criterion of not having used the Unlocking Talent program previously.
[2] On average, the treatment represented an estimated 40 percent of additional time in the tablet subject over standard classroom instruction in that subject.

control group students continued with standard instruction only. The study used an experimental design to ensure that differences in learning gains between the treatment and control groups can be attributed to the tablet intervention rather than to pre-existing differences among the groups. The experimental design ensures a high level of internal validity that provides rigorous estimates of the impacts at the two schools included in the study. However, due to the purposive selection of the schools, impact estimates may not be generalized to all Malawi government primary schools.

We conducted both impact analysis of the intervention on learning outcomes for the overall sample and exploratory analysis of the association of subgroup characteristics with learning outcomes. We produced two sets of impact estimates: Intent-to-Treat (ITT) estimates representing the impact of being assigned to the intervention, relative to being assigned to the control group; and Treatment-on-the-Treated (TOT) estimates representing the impact of attending the intervention at least 50% of the days that the learning center was open, relative to attending the intervention fewer or no days. Attending at least 50% of the offered days was considered minimum compliance with the treatment. About 88% of the children assigned to treatment met this attendance threshold.

Key Findings

The 8-month RCT in the two Malawi government primary schools produced the following results:

- 87% of learners persisted in the study: 89% of each treatment group and 83% of the control group persisted. After accounting for attrition, baseline equivalence of the final analytic sample was satisfied,[3] supporting the causal validity of our impact findings.
- Impacts[4]
    - The tablet literacy intervention produced a statistically significant positive effect on overall gains in literacy, with an effect size of .34 standard deviations. This translated into gains of 5.3 months of additional literacy learning over the control group for learners in the literacy intervention group, an added value of about 66%.[5]
    - The literacy intervention also produced a substantively important positive effect in reading comprehension (.25 standard deviations) and suggested positive effects in all other targeted literacy subskills (ranging from .16 to .20).
    - The tablet math intervention produced a substantively important positive effect of .29 standard deviations on gains in number identification (a key number sense skill). The math intervention also suggested a positive effect (.15 standard deviations) on gains in

[3] What Works Clearinghouse Standards Handbook (Version 4.0), page 14.

[4] Characterization of effect findings is explained in Chapter IV. We use the following convention based on the What Works Clearinghouse (Procedures Handbook, Version 4.0, pages 21-24) and on effect-size benchmarks proposed by Kraft (2018) for causal studies: we use “statistically significant positive effect” if the treatment effect is both positive and statistically significant; “substantively important positive effect” if the treatment effect is not statistically significant but is positive and equal to or larger than 0.25 standard deviations; “suggesting positive effects” if the treatment effect is not statistically significant but is between 0.15 and 0.25 standard deviations; and “indeterminate effect” if the treatment effect is not statistically significant and is between -0.15 and 0.15 standard deviations. None of the treatment effects for the study fell below -0.15 standard deviations.

[5] The added value % compares the treatment effect size to the average control group growth over the 8 months of the study for the relevant outcome. This percentage is then translated into “additional months of learning,” using 8 months as the base learning period. See Chapter IV for an explanation of this method and its limitations.

    pattern completion (another number sense skill). These effects represented an added value of 35% or higher.
  - Attending the intervention at least 50% of the time was associated with even larger positive effects on overall literacy (.40 standard deviations and an added value of 77%) and on number identification (.33 standard deviations).
- Exploratory analysis
  - The peri-urban school exhibited better treatment implementation and better treatment attendance than did the urban school during the 8-month study. While both schools were associated with statistically significant positive effects in overall literacy gains (.40 and .28 standard deviations, respectively), the peri-urban school exhibited a statistically larger gain in decoding (nonword reading) and the urban school a statistically larger gain in listening comprehension. Both skills are considered important for reading comprehension.[6]
  - Intervention effects on literacy and numeracy learning did not differ significantly by gender (male or female) or age category (6–7 years or 8–10 years).
  - Parents, teachers, and school and community leaders were overwhelmingly positive about the impact of the program on learners. More than three-fourths of the 43 adults interviewed at the end of the school year at the two pilot schools reported strong improvement in each of the following areas: children's excitement about school (88%), attendance (85%), achievement in literacy or math (81%), work effort (78%), and confidence as learners (78%).[7]

Conclusion

Findings from the pilot study are positive and encouraging. Some challenges with implementation in this pilot year may have attenuated the impact findings and we anticipate even greater learning effects when we conduct a new study in 2019-20 with a second cohort of Standard 2 learners. We are confident that this child-directed, technology-enabled learning approach can help children to become literate and numerate, but we know that it will take time and determination to achieve these goals.

---

[6] Nation (2019).

[7] The survey included the universe of Standard 2 teachers and school administrators and an opportunistic sample of parents and teachers.

I. Introduction

Study Overview

Through Imagine's planned portfolio of research, we seek to understand whether children with few education alternatives can become literate and numerate using child-directed, technology-enabled learning. To this end, we are building an evidence base for what works, for whom, and under what conditions, in schools and out-of-school settings, in different countries and languages, and with a small collection of promising software, starting with onebillion's *onecourse*.

Prior research on onebillion's literacy and numeracy applications used in the Malawi government primary schools showed promising results. However, the previous RCTs were conducted over short periods of time (8 and 14 weeks) to establish proof of concept.[8] To help fill the gap in knowledge about the longer-term impacts of technology-enabled learning, we conducted an 8-month RCT during the 2018-19 school year to address the following primary research questions:

1. What are the impacts over standard instruction on literacy and numeracy outcomes of using onebillion's *onecourse* software in Chichewa for 40 minutes per day for 8 months?
2. What impact does attendance in the intervention have on learning outcomes?
3. How far do children progress toward reading fluency with comprehension and comparable numeracy skills (i.e., arithmetic fluency with number sense)?
4. How are subgroup characteristics such as school, gender, and age category associated with learning outcomes?

The study tested the efficacy of onebillion's literacy and numeracy applications in the Chichewa language delivered through the Unlocking Talent implementation model. Unlocking Talent represents a collaboration between VSO Malawi, onebillion, and the Malawian Ministry of Education, Science and Technology. The program is currently in about 100 Malawi government primary schools and focuses on Standard 2 (grade 2) children. The *onecourse* curriculum follows accepted literacy and numeracy pedagogy and is loosely aligned to the Malawi education standards. Children progress through the tablet curriculum at their own pace. The implementation model involved building a learning center at each school that could accommodate 60 children at a time. The curriculum was delivered on 60 iPads, which were charged nightly by a battery that was powered by a solar cell installed on the roof of the learning center. The Standard 2 children in the treatment groups rotated through the learning center to use the tablets during one of four scheduled sessions per day. A rotation schedule was established so that children stepped out of different standard classes on different days of the week to attend the learning center. Thus, the intervention represented a supplement to normal instruction in the tablet subject.[9] The control group students continued with standard instruction only.

The study used an experimental design to ensure that differences in learning gains between the treatment and control groups can be attributed to the tablet intervention rather than to pre-existing differences among the groups. Attrition from the study was low (13 percent for the treatment and control groups

[8] Pitchford (2015) and Pitchford and Hubber (2017), respectively.

[9] On average, the treatment represented an estimated 40 percent of additional time in the tablet subject over standard classroom instruction in that subject.

combined). The final analytic sample met standards for baseline equivalence,[10] supporting the causal validity of our impact findings.

We collected data from three main sources: baseline and endline assessments, application usage data, and monthly monitoring visits by a University of Malawi-Chancellor College research team. To assess literacy and numeracy skills we used the Malawi adaptations of the Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA) in the Chichewa language. The same forms of the tests were used at baseline and endline. Outcome measures used for the impact analysis included overall composite scores in literacy and numeracy and selected subskills that represent key precursor skills and targeted outcomes in each subject. In literacy, we analyzed gains on the overall EGRA as well as in letter naming, decoding (nonword reading), listening comprehension, oral reading fluency, and reading comprehension. In math, we analyzed gains on the overall EGMA as well as in number identification, pattern completion, single and multiple-digit addition, and word problems. We adjusted results for multiple comparisons using the Benjamini-Hochberg correction.

We conducted impact analysis of the intervention on learning outcomes for the overall sample, producing two sets of impact estimates: Intent-to-Treat (ITT) estimates representing the impact of being assigned to the intervention, relative to being assigned to the control group; and Treatment-on-the-Treated (TOT) estimates representing the impact of attending the intervention at least 50% of the days that the learning center was open, relative to attending the intervention fewer or no days. We also conducted exploratory analysis of the association of subgroup characteristics with learning outcomes.

Report Structure

In the remainder of this report we describe the study conditions (Chapter II) and research design (Chapter III). We then present the impact findings (Chapter IV) and the results of additional exploratory analyses (Chapter V). We conclude by discussing implications of the findings (Chapter VI) and providing a reference list (References). Additional materials are provided in appendices, including information on *onecourse* for schools (Appendix A); means, standard deviations, sample sizes, and reliability for outcome measures (Appendix B); the stakeholder interview questionnaire (Appendix C); and statistical tests for subgroup differences (Appendix D).

[10] What Works Clearinghouse Standards Handbook (Version 4.0), page 14.

II. Study Conditions

This chapter describes the setting and target population for the study, the intervention, the counterfactual, and implementation.

Setting and Target Population

Two government primary schools were purposively selected for the study to represent an urban and a peri-urban (more rural) environment, meet sample size requirements, and meet the criterion of not having used the Unlocking Talent program previously. The schools were located in the Lilongwe district in the central region of Malawi, which contains the capital city. The urban school enrolled about 3,900 children and the peri-urban school about 3,200 children in Standards 1–8 (grades 1–8).

Conditions in the two communities and schools are challenging. Families in both communities are very low income and face food security issues and other poverty-related challenges. Neither school has electricity and the peri-urban school did not have an on-site source of water at the beginning of the study. The urban school serves about 10 villages (the farthest is 1 km away) and the peri-urban school serves about 18 villages (the farthest is 2 km away). Most children walk to school; road conditions are particularly bad leading to the peri-urban school and can be impassable at times during the rainy season (December to March). Class sizes at both pilot schools are very large (up to 100 children) and the achievement of learners at the beginning of the study was very low in literacy and low in mathematics. At baseline assessment, about half of the children could not recognize a single letter (54%) and about half could not add single digits (48%).

Absenteeism is a persistent problem at both schools. Common reasons for absenteeism included illness, household responsibilities, lack of food and clean clothes, bullying, failure to pay school development fees, funerals, and religious observances. At the urban school, many parents reportedly leave home early to run small businesses, so the children have to get themselves to school on their own. The urban environment also presents distractions that may lure children away from school. At the peri-urban school, the nearby tobacco industry causes seasonal migration; some families move away from the area during the growing season and return for the harvest. At both schools, some children are present at school but do not attend classes.

Table 1 provides additional information on the home environment and family resources of the children in the study. Because the information is based on baseline interviews with Standard 2 children (ages 6–10), the results may be inflated and should be interpreted with caution. However, the results indicate some consistent and some different responses across the schools. For example, the children at both schools reported similar home environments, but reported some differences in family resources: children in the peri-urban community were less likely than children in the urban community to report having electricity at home, electricity-powered appliances, or gasoline-powered vehicles, despite their more remote location.

Table 1. Baseline interview results for the study sample by school: Percentage of standard 2 learners giving the indicated response

Interview Topic	Urban School	Peri-urban school	Total
Family background / home environment			
Speaks Chichewa at home	93%	93%	93%
Father attended primary school, if applicable	75%	83%	80%
Mother knows how to read, if applicable	92%	93%	92%
Reading materials present at home	52%	54%	53%
Family Resources			
Ate food before arriving at school	73%	74%	74%
If ate, eats every day	80%	81%	80%
Electricity	52%	26%	40%
Toilet	98%	95%	97%
Radio	65%	63%	64%
Mobile phone	70%	70%	70%
TV	40%	23%	32%
Computer/video	24%	17%	21%
Refrigerator	29%	11%	20%
Bicycle	42%	42%	42%
Motorbike	17%	13%	15%
Car or truck	24%	11%	18%
Observations = 671			

NOTE: Because information is based on interviews with Standard 2 children (ages 6–10), the results may be inflated and should be interpreted with caution. However, the results indicate some consistent and some different responses across the schools.

The Intervention

The *onecourse* applications used during the 2018–19 school year in Malawi contained 1,248 instructional units in literacy and 282 units in math. Both the Malawi national curriculum and the *onecourse* curriculum were research based and generally covered the same content, although the *onecourse* curriculum covered material taught in both Standards 1 and 2, providing students with the opportunity for review before moving to new Standard 2 content. The software was delivered on iPads. Appendix A provides information in English on *onecourse* for schools including several screenshots of the literacy application in Chichewa.

Reflecting the Unlocking Talent model, VSO-Malawi constructed a learning center at each study school to accommodate the intervention. The learning centers each contained two cabinets of 30 iPads each, one designated as the literacy cabinet and the other as the math cabinet. One solar cell per cabinet was installed on the roofs of the learning centers. The solar cells charged a battery in each cabinet that in turn charged the iPads. The free-standing learning center was dedicated to the intervention and both the cabinets and center were locked when the learning center was not in session. This arrangement helped to secure the equipment and avoid any control group students "crossing over" to the treatment.

VSO-Malawi staff offered an eight-hour technical training on three afternoons to all teachers in both schools, which covered an orientation to the learning center and iPads, practice using the learner iPads and apps, an introduction to the teacher iPad, practice registering students, a model and practice *onecourse* session, troubleshooting tips, as well as an overview of the research study. The schools determined which teachers would supervise the 20 weekly learning center sessions (four 1-hour sessions per day, Monday through Friday, held during regular school hours). They also designated a learning center coordinator and established a supervision schedule. Typically, Standard 2 teachers took turns supervising the tablet sessions. No new teachers were hired to implement the program.

Students in the literacy and math treatment groups at each school were divided into four learner groups named after animals (such as antelopes, lions, etc.). The animal groups were split between the literacy and math treatment groups (up to 29 learners each). The teachers created a learning center schedule allowing the animal groups to step out of different classes on different days of the week to attend the learning center during their assigned session. Thus, the intervention represented a supplement to normal instruction in the tablet subject.[11] The supervising teachers called the appropriate animal group to the learning center during a given session.

One tablet in each cabinet was reserved as the teacher tablet, so a maximum of 29 children in the literacy treatment and 29 children in the math treatment could use the tablets during a session (58 children total per session). The supervising teachers controlled children's access to the tablets and curriculum through the teacher tablet, which used local wifi to unlock the specific curriculum for children on each side of the learning center and to select the session length (40 minutes). Children used either the literacy or numeracy app—not both—for 40 minutes per daily session in an effort to maximize time on task in each subject and to isolate the impact of the two applications. After the tablets were unlocked by the supervising teacher, the children would log in to their individual accounts by selecting the image containing their photo and name. Children progressed at their own pace through the tablet curriculum, starting each day from where they left off the previous day. After 40 minutes, the tablets locked automatically. Usage data were sent continuously from the tablets to the local server and then weekly from the server to onebillion via the internet.

[11] We estimated that the treatment represented 40 percent of additional time in the tablet subject on average over standard classroom instruction in that subject.

The Counterfactual

Standard 2 children at the two government primary schools attended classes from 7:30am to 12:30pm each day. The urban school provided extracurricular activities (such as sports, bible, debate) from 12:30-1:30pm. The school day was divided into 8 half-hour class periods, with two 15-minutes breaks and a half-hour recess/library period during the morning, for a total of 20 instructional hours per week. The daily class schedule rotated Monday through Friday. Standard 2 classes included

- Chichewa (reading), 5 hours/week
- Math, 5 hours/week
- English, 5 hours/week, and
- Additional subjects such as arts and life skills, 5 hours/week

At both schools, Standard 2 populated four classrooms with 1-2 assigned teachers and up to 100 children per classroom. Classroom assignment was fluid, and children shifted classrooms during the year due to teacher changes, following one's siblings or friends, or other reasons. Children in Standard 2 sat on the floor; desks were reserved for children in Standards 4 or higher. Children generally had notebooks and a pencil. The teacher had a chalkboard and chalk and sometimes additional materials that she or he would walk around with to show to the learners. Instruction often took the form of traditional call and response and recitation. The Standard 2 curriculum followed Malawi national standards. In literacy, the curriculum adhered to the National Reading Program (NRP), which had been implemented nationwide the prior school year. The NRP provided scripted lessons for the teachers to follow.

Implementation of the Intervention

Imagine and VSO staff visited the study schools several times to explain the intervention and the research and to meet with the school and community leaders. Before agreeing to participate in the study, leaders from both schools observed an existing implementation at a Lilongwe primary school that had been using the UT program for about 5 years.

The UT implementation model followed in this study involved a community sensitization meeting held by our VSO-Malawi partners, with Imagine staff in attendance, at both study schools prior to the launch of the intervention. The purpose of these meetings was to introduce the broad community to the intervention and study, allow the adults to use the tablets in the learning centers, answer any questions, and obtain parental verbal consent for the study. The communities welcomed the intervention and were very engaged in ongoing community meetings.

As part of the UT model, VSO staff offered the eight-hour technical training to all teachers in the two schools, with the goal of creating schoolwide support and making it possible for any teacher to step in to help in the learning center. VSO assigned one of their Education Specialists (usually a paid VSO volunteer) to each school to visit the learning center periodically and serve as a liaison for technical issues. During the 8-month intervention, VSO staff also held monthly meetings with school and community leaders to gauge engagement and troubleshoot any implementation issues.

III. Research Design

The experimental design for this study helped to ensure that differences in learning gains between the treatment and control groups can be attributed to the tablet intervention rather than to pre-existing differences among the groups. This chapter describes the research questions, randomization approach, study sample, data sources, outcome measures, and analytic methods we used.

Research Questions

Through a planned portfolio of research conducted over several years, Imagine seeks to understand whether children with few education alternatives can become literate and numerate using child-directed, technology-enabled learning. Our primary research questions for the Malawi pilot year study were:

1. What are the impacts over standard instruction on literacy and numeracy outcomes of using onebillion's *onecourse* software in Chichewa for 40 minutes per day for 8 months?
2. What impact does attendance in the intervention have on learning outcomes?
3. How far do children progress toward the ultimate learning goals of reading fluency with comprehension and comparable numeracy skills (i.e., arithmetic fluency with number sense)?
4. How are subgroup characteristics such as school, gender, and age category associated with learning outcomes?

Randomization Approach

To address the above questions about impact, Imagine and Malawi-based partners launched in October 2018 an 8-month efficacy randomized controlled trial (RCT) using a non-clustered, blocked individual random assignment (BIRA) design.[12] We purposively selected two government primary schools in the capital region to represent an urban and a peri-urban environment, meet sample size requirements, and meet the criterion that they had not previously used the Unlocking Talent program. The urban school enrolled about 3,900 children and the peri-urban school about 3,200 children in Standards 1–8 (grades 1–8). Because the two schools were purposively and not randomly selected, the resulting impact estimates represent average effects for the two study schools and do not generalize to all primary schools in Malawi.

Children were assessed at baseline prior to randomization to the treatment and control groups. We assessed all Standard 2 children in two stages a few weeks apart. The first stage assessed children who were on the initial enrollment lists provided by the schools; these lists were based on the children who had completed Standard 1 at the schools the previous year. The second stage assessed children who were not on the original lists due either to new enrollment at the schools or to re-assignment to Standard 2.

Children were considered eligible for the study if they were

(1) enrolled in Standard 2,
(2) 6-10 years old (95% of Standard 2 learners were in this age range), and
(3) present during baseline assessment.

[12] A non-clustered, blocked individual random assignment design involves conducting individual random assignment independently within non-overlapping subpopulations that comprise the entire sample (Schochet 2016). In our study, the two schools were purposively and independently selected to represent urban and peri-urban environments, then students were randomly assigned within each of the schools.

Ultimately, 674 Standard 2 learners at the two schools were determined to be eligible: 348 learners at the urban school and 326 learners at the peri-urban school. Independently within each school, the children were randomly assigned to two treatment groups (literacy and math) and one control group within four gender (male and female) and age category (6-7 years and 8-10 years) strata to ensure treatment-control group balance for each subgroup.

Study Sample

Table 2 and Figure 1 describe the characteristics of the Standard 2 learners in the study sample. Slightly more than half of the sample were boys (53 percent) and slightly more than half were aged 6-7 years (53 percent) at the beginning of the school year. The expected age of Standard 2 learners in Malawi is 7 years. Older children may have started school late, may not have passed prior end-of-year exams and were repeating the grade, or may have been retained in grade at the request of the parent(s), or for other reasons. Children in the study sample also exhibited very low baseline achievement in literacy and low baseline achievement in math. Slightly more than half of the children (54%) could not recognize a single letter at the beginning of Standard 2 and about half (48%) could not perform addition with single digits (Figure 1).

Table 2. Percentage distribution of the study sample by gender and age category

Characteristic	Urban School	Peri-urban school	Total
Column total	100%	100%	100%
Gender			
Male	53%	52%	53%
Female	47%	48%	47%
Age category			
6-7 years	58%	47%	53%
8-10 years	42%	53%	47%
Gender by age category			
Male			
6-7	54%	41%	48%
8-10	46%	59%	52%
Female			
6-7	62%	53%	58%
8-10	38%	47%	42%
Age category by gender			
6-7			
Male	50%	45%	48%
Female	50%	55%	52%
8-10			
Male	58%	57%	58%
Female	42%	43%	42%
Count (N)	348	326	674

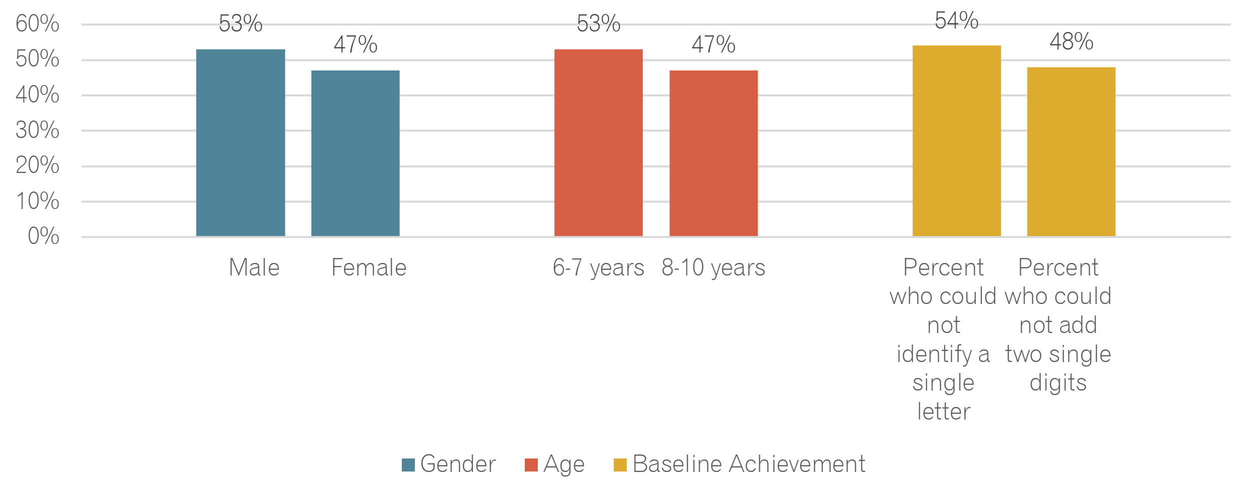


Figure 1. Characteristics of students who participated in the study

Data Sources

We collected data from three main sources:

- **Baseline and endline assessments.** These data provided the outcome measures for evaluating the impact of the intervention on literacy and numeracy learning. A University of Malawi-Chancellor College research team assessed the literacy and math skills of all Standard 2 children at the two study schools during two-week assessment periods beginning at the end of September 2018 and the end of June 2019,[13] using the Malawi adaptations of the Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA) in the Chichewa language. The same forms of the tests were used at baseline and endline. EGRA and EGMA were developed to fill a gap in international assessment of foundational and emerging literacy and numeracy skills and have become the early grade assessment standard in developing countries and development organizations.[14] The assessors recorded basic demographic information and assessment responses digitally for each child using a tablet-based data collection software, Tangerine, designed specifically for use with EGRA and EGMA.
- **Application usage data.** These data enabled analysis of the impact of attending the intervention on learning outcomes. The *onecourse* software collected data on the tablet usage of each child in the treatment groups and transmitted these data to the software developer, who generated weekly reports for Imagine. We used data on children's log-ins to the learning software to calculate daily attendance and corroborated these data with hand-written attendance registers and with additional activity data from the tablets.

---

[13] An additional week of baseline assessment was conducted in October 2018 to assess children who were not on the original Standard 2 enrollment lists.

[14] Originally developed in 2006 by RTI International with funding from USAID, EGRA has been implemented in more than 50 countries and 70 languages (Gove and Wetterberg 2011) and has been promoted in the context of the Education 2030 agenda and the Sustainable Development Goals by the Learning Metrics Task Force7, UNICEF, the World Bank, USAID, the Center for Universal Education at Brookings, and the Global Partnership for Education. EGMA was originally developed in 2008 and has been implemented in more than 14 countries (RTI International 2014).

- **Monthly monitoring visits.** These data helped to interpret treatment impacts by describing the quality and fidelity of the intervention and the nature of the counterfactual (i.e., the standard instruction received by the control group). The Chancellor College research team conducted monthly monitoring visits to each of the pilot schools to collect observational and interview data using standard data collection forms co-created by the team and Imagine. Interview questions were translated into Chichewa by the Chancellor College team. During the visits, the research team observed the implementation practices, equipment functioning, and student engagement and usage behaviors in the learning center, as well as student engagement behaviors in the regular classrooms. The team also interviewed the learning center supervising teachers and coordinator, school administrators, regular classroom teachers, and children in the treatment and control groups about implementation, engagement, and perceived impacts of the program. Part way during the year, the monitoring team also began conducting in-depth interviews with a random sample of about two dozen Standard 2 children and their parents to better understand issues affecting attendance at the schools. The team recorded monitoring results using a combination of a rating scale and notetaking. The team submitted a summary report one week following each monthly visit.[15]

In addition to the above main sources of data, we also conducted a non-experimental, end-of-year survey of 43 parents, teachers, and school and community leaders at the two schools to assess stakeholder perceptions of the program's impact.[16]

Outcome Measures

Ultimately, we want to know whether children using child-directed, technology-enabled learning can learn to read with fluency and comprehension and attain comparable numeracy skills (i.e., arithmetic fluency with number sense). However, attaining these literacy and numeracy outcomes may take more than one school year. Consequently, for the 8-month pilot study we assessed learning gains on precursor skills as well as ultimate outcomes of interest. Table 3 lists the literacy and numeracy outcome measures that we used based on EGRA and EGMA. We assessed each measure at baseline and endline and calculated the gain as the difference between the two points in time. As described in the Analytic Methods section, we used gain measures of these outcomes as the dependent variables in our analyses. Appendix B provides the means, standard deviations, sample sizes, and reliability of the outcome measures at baseline and endline.

[15] Imagine and the Chancellor College team plan to digitize the data collection process for the new study in 2019–20 to facilitate compilation and analysis of the monitoring data.
[16] While the survey included the universe of Standard 2 teachers and school administrators, it included an opportunistic sample of parents and community leaders, so the results are suggestive but not conclusive.

Table 3. Outcome measures used in the analysis

Domain	Outcome measure
Literacy	
Overall composite	EGRA average percent correct[1]
Targeted skills	Simple View of Reading
	Decoding (nonword reading) percent correct
	Listening comprehension percent correct
	Reading comprehension percent correct
	Additional targeted outcomes
	Letter naming percent correct[2]
	Oral reading fluency
Numeracy	
Overall composite	EGMA total percent correct[3]
Targeted skills	Number sense
	Number identification percent correct
	Pattern completion percent correct
	Arithmetic operations
	Addition level 1 percent correct
	Addition level 2 percent correct
	Word problems percent correct

[1]The composite EGRA average percent correct measure averaged the percent correct for each of the nine EGRA subtests administered. See discussion of composite variables below.
[2]For the new study in 2019–20, we will replace the letter naming subtest with letter sounds, the latter which is considered more appropriate for the Chichewa language (versus English).
[3]The composite EGMA total percent correct measure averaged across all items in the eight core EGMA subtests. See discussion of composite variables below.

The Malawi adaptation of the EGRA in Chichewa includes nine subtests: initial sound identification, phoneme segmentation, letter naming, syllable reading, familiar word reading, nonword reading, listening comprehension, oral reading fluency, and reading comprehension. The Malawi adaptation of the EGMA in Chichewa includes eight core subtests: number identification, quantity discrimination, pattern completion, word problems, addition level 1 (single digits), addition level 2 (multiple digits), subtraction level 1 (single digits), and subtraction level 2 (multiple digits).These EGRA and EGMA subtests reflect fundamental skills needed for the acquisition of reading and mathematics knowledge and that are predictive of later reading and math success.[17]

For both literacy and numeracy, we created an overall composite measure that combined results on all available subtests. The composite measures provided a summary measure of how much the tablet intervention contributed to overall learning across the subskills measured by EGRA and EGMA. For literacy, we created a composite EGRA *average* percent correct measure by averaging the percent correct for each of the nine EGRA subtests administered. For numeracy, we created a composite EGMA *total* percent correct measure by averaging across all items in the eight core EGMA subtests administered. For

[17] RTI International (March 2016 and March 2014, respectively).

EGRA, we created the *average* percent correct measure, which weights each *subtest* equally, rather than a *total* percent correct measure, which weights each *item* equally, because the wide variation in number of items on the EGRA subtests (ranging from 5 to 100 items) does not reflect the relative importance of the subtests. In contrast, for EGMA, we created the *total* percent correct measure, because the number of items on the core EGMA subtests were more similar (ranging from 4 to 10 items) and gave more equal weight to the subtests.[18]

In addition to the EGRA and EGMA composite measures, we also targeted for analysis a subset of literacy and numeracy outcome measures that reflect important precursor skills as well as ultimate outcomes of interest. For literacy, we selected five EGRA subtests that reflect both reading progress and attainment. Three of these subtests are incorporated in the Simple View of Reading,[19] which asserts that reading comprehension is largely the product of decoding and linguistic comprehension. We used the EGRA nonword reading, listening comprehension, and reading comprehension subtests to represent these constructs. We also targeted two additional literacy skills: letter naming as an early precursor reading skill (to be replaced in the next study with letter sounds) and oral reading fluency.[20] For numeracy, we selected five EGMA subtests that reflect the numeracy learning goal of arithmetic fluency with number sense. Two subtests (number identification and pattern completion) measure fundamental number sense skills. The other three subtests (addition with single digits (level 1), addition with multiple digits (level 2), and word problems) measure arithmetic operations skills.

Analytic Methods

We conducted our study in two main parts: (1) analysis of the impact of the intervention on learning outcomes for the overall sample and (2) exploratory analysis of the association of subgroup characteristics with learning outcomes. We describe both sets of analyses below and present respective findings in the next two chapters.

Impact Analysis

We present two sets of estimates of the impact of the intervention on literacy and numeracy outcomes:

- Intent-to-Treat (ITT)[21] impact estimates: Representing the impact of being assigned to the intervention, relative to being assigned to the control group.
- Treatment-on-the-Treated (TOT)[22] impact estimates: Representing the impact of attending the intervention at least 50% of the days that the learning center was open, relative to attending the intervention fewer or no days. Attending at least 50% of the offered days was considered minimum compliance with the treatment. About 88% of the children assigned to treatment attained this attendance rate.

[18] In future, we plan to investigate a more nuanced approach to creating the composite measures, such as using Item Response Theory to scale the items. Such composite EGRA and EGMA measures do not currently exist.
[19] Nation (2019).
[20] Letter sounds is considered more appropriate than letter naming in the Chichewa language. The proposed Malawi benchmark for oral reading fluency in Standard 2 is reading 40+ correct words per minute of connected text (i.e., passage reading) (USAID, Proposing Benchmarks for Early Grade Reading in Malawi).
[21] See What Works Clearinghouse Standards Handbook Version 4.0, page 47. Intent-to-Treat is also referred to as Intention-to-Treat.
[22] Ibid. TOT impact estimates are also referred to as Complier Average Causal Effects (CACE).

The ITT analysis estimates the effect of being offered the intervention, while the TOT analysis estimates the effect of actual take-up of the intervention at a minimum threshold of "compliance." We use the ITT analysis to address our first research question about the impacts of the intervention over standard instruction on literacy and numeracy outcomes. We use the TOT analysis to address our third research question about the impact of attendance in the intervention on learning outcomes.

General Approach to Impact Analysis

For both impact analyses, we followed standard practice for estimating impacts from a non-clustered, blocked individual random assignment design (BIRA) evaluation. Because we purposively selected the two schools in the study, and randomly assigned learners independently within each school, the schools represented independent samples. We conducted the impact analysis as a multi-site randomized trial, averaging separately derived site-level impacts and assuming fixed site effects.[23] To produce the average treatment effect for each outcome measure, we estimated the treatment impact and associated effect size separately for the two schools and then averaged the estimates. We calculated standard errors for the averages by pooling the site-level standard errors.

To obtain more precise school-level impact estimates, we adjusted for baseline student characteristics in an ordinary least squares regression model that used the gain score for each outcome measure as the continuous dependent variable. We used gain scores instead of endline outcomes as the dependent variables to avoid attenuation bias due to measurement error in the baseline measure. In the regression model we included fixed effects for the gender-age category strata as well as baseline covariates for gender, age category, the relevant outcome measure, and the opposite-subject composite outcome measure. We made no adjustments for data nonresponse (which was extremely low);[24] cases with missing data were deleted from the specific relevant analysis.

To estimate site-level ITT impacts, we used a regression model that compared the mean gains of the relevant treatment group (literacy or numeracy) to those of the control group on the outcome measures shown in Table 3, allowing the impact estimates to vary for each site. The basic form of the ordinary least squares model was:

$$(1)\ y_{isj} = \alpha_j + \beta X_{isj} + \delta_j T_{isj} + \lambda_s + \varepsilon_{isj}$$

where

y_{isj} was the simple gain score (endline score minus baseline score) on the outcome of interest in Table 3 for student *i* in strata *s* in site *j*,

α_j was a site-specific intercept,

X_{isj} was a vector of baseline characteristics of student *i* in strata *s* in site *j*; the control variables included

gender (male vs. female)

age category (6–7 years vs. 8–10 years)

baseline score on the relevant outcome measure

[23] Bernstein et al. (2009); Gleason et al. (2010); Dong & Maynard (2013); and Schochet (2016).

[24] The number of individual cases with missing data ranged from 0 to 3, depending on the outcome measure.

baseline score on the opposite-subject composite measure (overall EGRA average percent correct for numeracy outcomes and overall EGMA total percent correct for literacy outcomes),

T_{isj} was a binary variable for treatment status, indicating whether student *i* in strata *s* was assigned to the relevant treatment (literacy or numeracy) in site *j*,

λ_s represented fixed effects for the gender-age category strata,

ε_{isj} was a random error term that reflects the influence of unobserved factors on the outcome,

β was a vector of parameters to be estimated for the control variables, and

δ_j was the estimated coefficient on treatment status in site *j* and represented the impact of participating in the treatment at site *j*.

For each site-level impact estimate (δ_j), we also computed the associated effect size, which reflected the magnitude of the impact relative to the variation in the outcome measure in the sample (the treatment and control groups combined). Site-level effect sizes were calculated using Cohen's *d*. To produce the overall average treatment effect (δ) for each outcome measure, we averaged the site-level estimate for the treatment impact (δ_j) and the associated effect size produced by the regression in (1) above. We calculated standard errors for these averages by pooling the site-level standard errors. Statistical significance of the treatment effects was adjusted for multiple treatment-control comparisons using the Benjamini-Hochberg correction, as described later in this chapter.

Treatment-on-the-Treated (TOT) Impact Analysis

To examine the impact of attendance in the intervention on learning outcomes, we set a minimum threshold of compliance with the treatment at 50% attendance on the days the learning centers were open. While most children assigned to the intervention attended at least 50% of the time that their learning center was open, 12% did not. Because the number of days the learning centers were open differed by school—with the peri-urban center open a total of 130 days and the urban center open a total of 106 days—we calculated the 50% threshold separately for each school. To investigate the impact of the intervention on those who attended at least 50% of the available time, we followed a standard approach for estimating TOT impacts. First, we used treatment status as the "instrumental variable" to predict attendance. We then used the predicted attendance (met threshold, did not meet threshold) as the treatment indicator (T_{ij}) in the regression model (1) above. Low attendance may be due to factors that are correlated with learning outcomes, such as lack of motivation or lower academic ability. By predicting attendance using a random variable (assignment to treatment), rather than using actual attendance, the predicted attendance is highly correlated with actual attendance but uncorrelated with other student characteristics. Therefore, estimating the relationship between predicted attendance and learning outcomes produces an unbiased estimate of the impact of attending the intervention at the threshold level.

Adjustment for Multiple Comparisons

Analysis of multiple outcomes and population subgroups can yield false positive results: the more comparisons that are conducted, the greater the likelihood that a positive result is produced simply by chance. To avoid "false discoveries," we limited our confirmatory analyses to treatment-control comparisons among the overall sample only and to the outcome measures we designated as the most important only (those shown in **Table 3**). All subgroup analyses were pre-determined to be exploratory. We then applied formal adjustments for the multiple treatment-control comparisons using the Benjamini-Hochberg correction.[25] We based our main conclusions in Chapter IV on the results after adjustment for multiple hypothesis testing. However, because of lack of agreement on the need for such an adjustment, or the most appropriate method for making it, we indicated in the results tables where results were statistically significant both before (†) and after (*) the adjustment.

Reporting Impact Findings

When describing the strength of a treatment effect, it is important to consider both its statistical significance and its magnitude (effect size). In this report we used the convention shown in **Table 4** for describing the strength of our effect findings, based on What Works Clearinghouse guidelines and additional recommendations by Kraft (2018).

Table 4. Characterization of effect findings

Description of effect	Explanation
Statistically significant positive effect	The estimated treatment effect (δ) is positive and statistically significant.[1]
Substantively important positive effect	The estimated effect (δ) is not statistically significant but is positive and the effect size is 0.25 standard deviations or larger.[1]
Suggests a positive effect	The estimated effect (δ) is not statistically significant but is positive and the effect size is between 0.15 and 0.25 standard deviations.[2]
Indeterminate effect	The estimated effect (δ) is not statistically significant and the effect size is between -0.15 and 0.15 standard deviations.[1, 2]

[1]Based on What Works Clearinghouse Procedures Handbook (Version 4.0), pages 21-24.
[2]Based on effect-size benchmarks proposed by Kraft (2018) for causal studies. An effect size of between 0.15 and 0.25 corresponds to Kraft's "medium" effect category, adjusted for the lower elementary grades.
NOTE: None of the study's treatment effects were negative and statistically significant, nor were any effect sizes -0.15 standard deviations or lower, so comparable categories for negative effects were not applicable.

[25] See What Works Clearinghouse Procedures Handbook (Version 4.0), pages F-3 to F-6, on multiple outcome measures within the same domain, tested with a single comparison group.

In addition to reporting the statistical significance and magnitude (effect size) of a treatment effect, we also translated the effect sizes in three ways to help with interpretation:

1. The "% added value" compares the treatment effect size to the average control group growth over the 8 months of the study for the relevant outcome. The ratio of the standardized treatment growth to control group growth is expressed as a percent. We used this translation to help interpret the practical importance of the effect size. However, this method has several important limitations.[26]
2. We translated the % added value into "additional months of learning," using 8 months as the denominator. The same limitations that apply to method #1 above also apply to this measure.
3. Following recommendations in Baird and Pane (2019), we also translated the effect sizes into percentile growth. Assuming a normal distribution, the translation estimates the change in percentile rank that would be expected for an average (median) student who received the treatment.

Attrition and Baseline Equivalence

As mentioned, attrition from the study after 8 months was only 13% overall, with 11% for the treatment groups and 17% for the control group. Following What Works Clearinghouse guidelines,[27] we calculated standardized mean differences between the literacy treatment group and the control group, and between the numeracy treatment group and the control group, on measurable baseline characteristics including gender, age category, and baseline achievement in literacy and math. All of the standardized mean differences met WWC group design standards. Baseline equivalence was satisfied without the addition of covariates for differences $\leq .05$ and with the addition of covariates for differences $>.05$ and $\leq .25$.[28] This equivalence result supports the causal validity of our impact findings for the overall sample.

Statistical Power

During the design phase for the study, we conducted power calculations to determine minimum detectable effects (MDEs) in order to plan an adequate sample size and to detect anticipated impacts. We subsequently revised the calculations based on the final analytic sample size. Both power calculations used a two-tailed t-test, a 5 percent critical value for assessing statistical significance, and an 80 percent level of statistical power. In each case, we assumed fixed site effects, reflecting the fact that the two study schools were purposively selected and the study's estimates are representative of the participating schools only. Our final analytic sample provided a high probability of detecting effect sizes for the literacy and numeracy interventions as small as .27 before adjustment for multiple comparisons. Impacts of this

[26] As explained in Baird and Pane (2019), the method assumes that additional learning is linear and would occur at the same rate as during the study period, which may not be true. In addition, small denominators can produce implausibly large results: the measure should always be interpreted with caution and particularly large estimates (indicated with a !) should be interpreted with extreme caution. Finally, the translation introduces additional statistical uncertainty; it cannot be assumed that because a treatment effect was statistically significant this means that the % added value translation is also statistically significant.
[27] What Works Clearinghouse Standards Handbook (Version 4.0), page 14.
[28] Standardized differences between the literacy treatment group and control group were: gender (.02), age category (.09), baseline EGRA average percent correct (.08), and baseline EGMA total percent correct (.02). Standardized differences between the numeracy treatment group and control group were: gender (.02), age category (.00), baseline EGRA average percent correct (.12), and baseline EGMA total percent correct (.08).

size would be smaller than the EGRA and EGMA impacts found in prior short-term studies of onebillion's software in Malawi (.42 for literacy and .36 for math).[29]

Exploratory Analyses

We were also interested in examining whether impacts differed at the two government primary schools and whether gender and age category were associated with outcomes. Specifically, we wanted to know whether the better implementation and better attendance we observed at the peri-urban school were associated with better outcomes than at the urban school. We also hypothesized that the "blind" nature of tablet instruction—where the software does not differentiate users based on their gender or age—may generate more equitable learning outcomes for subgroups than some classrooms.

We estimated school-level impacts in the process of calculating average study effects, as described in the Impact Analysis section above. For the subgroup analyses related to gender and age category, we followed the same analytic approach as for the overall sample described in that section, adding an interaction term for the subgroup and treatment along with the main variables in the regression model (1). We estimated impacts for boys and girls and for younger and older children at each school and then averaged the school-level estimates. We then conducted t-tests of the differences between the subgroups.

Because of the limitations in statistical precision when conducting multiple subgroup comparisons, we pre-determined that these analyses would be exploratory only. Thus, the estimated relationships between the factors examined in the exploratory analysis and treatment outcomes cannot be interpreted as causal. Rather, the estimated relationships are suggestive of factors that could contribute to intervention success and may be worthy of further research.

---

[29] Pitchford et al. (2017) and Pitchford (2015), respectively.

IV. Impacts

Through a planned portfolio of research, Imagine seeks to understand whether children with few education alternatives can become literate and numerate using child-directed, technology-enabled learning. For this Malawi pilot study, our specific research questions included

1. What are the impacts over standard instruction on literacy and numeracy outcomes of Standard 2 children using the *onecourse* software for 40 minutes per day for 8 months?
2. What impact does attendance in the intervention have on learning outcomes?
3. How far do children progress toward targeted learning outcomes?

In this chapter, we present results of our analyses of the tablet intervention's impacts on gains in EGRA and EGMA test scores for the overall sample and we compare how far children in the treatment and control groups progressed on reading and math benchmarks. We also present the impacts of attending the intervention at least 50% of the time that it was offered. Finally, we present results from a non-experimental survey of stakeholder perceptions of the intervention impacts at the two schools. In the next chapter we present results from our exploratory subgroup analyses.

Impact of Being Selected for the Intervention

This section presents our Intent-to-Treat (ITT) impact estimates for the overall sample (not subgroups). The ITT estimates represent the impact of being assigned to the intervention, relative to being assigned to the control group. As seen in Table 5 for the full treatment group

- **The tablet literacy intervention had a statistically significant impact on overall gains in literacy with an effect size of .34 standard deviations.**

Learners in the literacy intervention group gained the equivalent of 5.3 months of literacy learning over the control group during the 8-month intervention, an added value of about 66%. Furthermore, an average student's percentile rank would be expected to increase by 13 percentiles due to the intervention. The literacy intervention also produced a substantively important positive effect in reading comprehension of .25 standard deviations and suggested positive effects in all other targeted literacy subskills (ranging from .16 to .20).

- **The tablet math intervention had a substantively important positive effect of .29 standard deviations on gains in number identification (a key number sense skill).**

Learners in the math intervention group gained the equivalent of 3.1 months of additional number identification learning over the control group, an added value of about 39%. And an average student's percentile rank would be expected to increase by 11 percentiles due to the intervention. The math intervention also suggested a positive effect of .15 standard deviations on gains in pattern completion (another number sense skill).

Impact of Attending the Intervention at least 50% of the Time

This section presents our Treatment-on-the-Treated (TOT) impact estimates for the overall sample (not subgroups). The TOT estimates represent the impact of attending the intervention at least 50% of the days that the learning center was open, relative to attending the intervention fewer or no days. Because the number of days the learning centers were open differed by school, we calculated the 50% threshold separately for each school. About 88% of the children assigned to treatment attained this attendance rate. As seen in Table 5

- For children who attended at least 50% of the time, the tablet literacy intervention had a statistically significant impact on overall gains in literacy, with an effect size of .40.

Learners in the literacy intervention group who met the minimum attendance threshold gained the equivalent of 6.2 months of literacy learning over the control group during the 8-month intervention, an added value of about 77%. Furthermore, an average student's percentile rank would be expected to increase by 15 percentiles due to the intervention. For this subset of learners, the literacy intervention also produced a substantively important positive effect in reading comprehension of .29 standard deviations and suggested positive effects in all other targeted literacy subskills (ranging from .18 to .24).

- For children who attended at least 50% of the time, the tablet math intervention had a substantively important positive effect of .33 standard deviations on gains in number identification (a key number sense skill).

Learners in the math intervention group who met the minimum attendance threshold gained the equivalent of 3.5 months of additional number identification learning over the control group, an added value of about 44%. And an average student's percentile rank would be expected to increase by 13 percentiles due to the intervention. The math intervention also suggested a positive effect of .17 standard deviations on gains in pattern completion (another number sense skill).

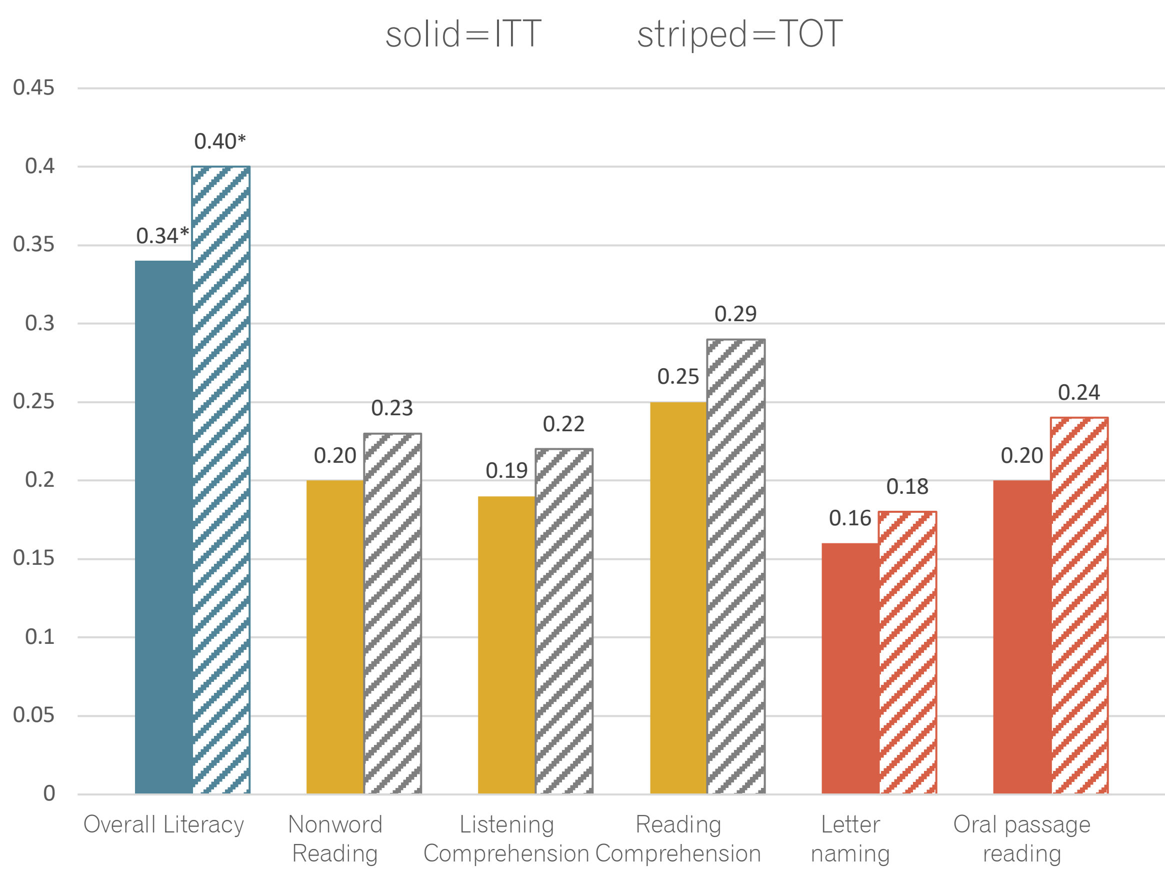
Figures 2 and 3 following **Table 5** illustrate the ITT and TOT effect sizes. Attending the intervention at least 50% of the time was generally associated with larger positive effects than being assigned to treatment.

Table 5. Effect sizes, percentile growth, and added value for the full treatment group (ITT) and those who attended at least 50% of the treatment days (TOT)

	Full treatment group (ITT)				Those with >=50% attendance (TOT)[1]			
			Added value[2]				Added value[2]	
Targeted Outcome Measures	Effect size	Pctl grwth[3]	%	Mos.	Effect size	Pctl grwth[3]	%	Mos.
Literacy								
Overall EGRA average percent correct	.34*	13	66%	5.3	.40*	15	77%	6.2
Subtest percent correct								
Simple view of reading								
Nonwords/Invented Words	.20	8	51%	4.0	.23	9	59%	4.8
Listening Comprehension	.19	8	264%!	21.1!	.22	9	297%!	23.7!
Reading Comprehension	.25	10	234%!	18.7!	.29	11	270%!	21.6!
Additional targeted outcomes6								
Letter Naming	.16	6	30%	2.4	.18	7	34%	2.7
Oral Reading Fluency	.20	8	49%	4.0	.24	9	58%	4.6
Numeracy								
Overall EGMA total percent correct	.07	3	7%	0.6	.07	3	8%	0.6
Subtest percent correct								
Number sense								
Number Identification	.29†	11	39%	3.1	.33†	13	44%	3.5
Pattern Completion	.15	6	35%	2.8	.17	7	38%	3.1
Arithmetic operations 0								
Addition Level 1	-.01	3	-2%	-0.1	-.02	3	-3%	-0.2
Addition Level 2	.00	0	0%	0.0	.00	-1	0%	0.0
Word Problems	.07	0	13%	1.1	.08	0	14%	1.1
Observations (EGRA/EGMA)	382/382				382/382			

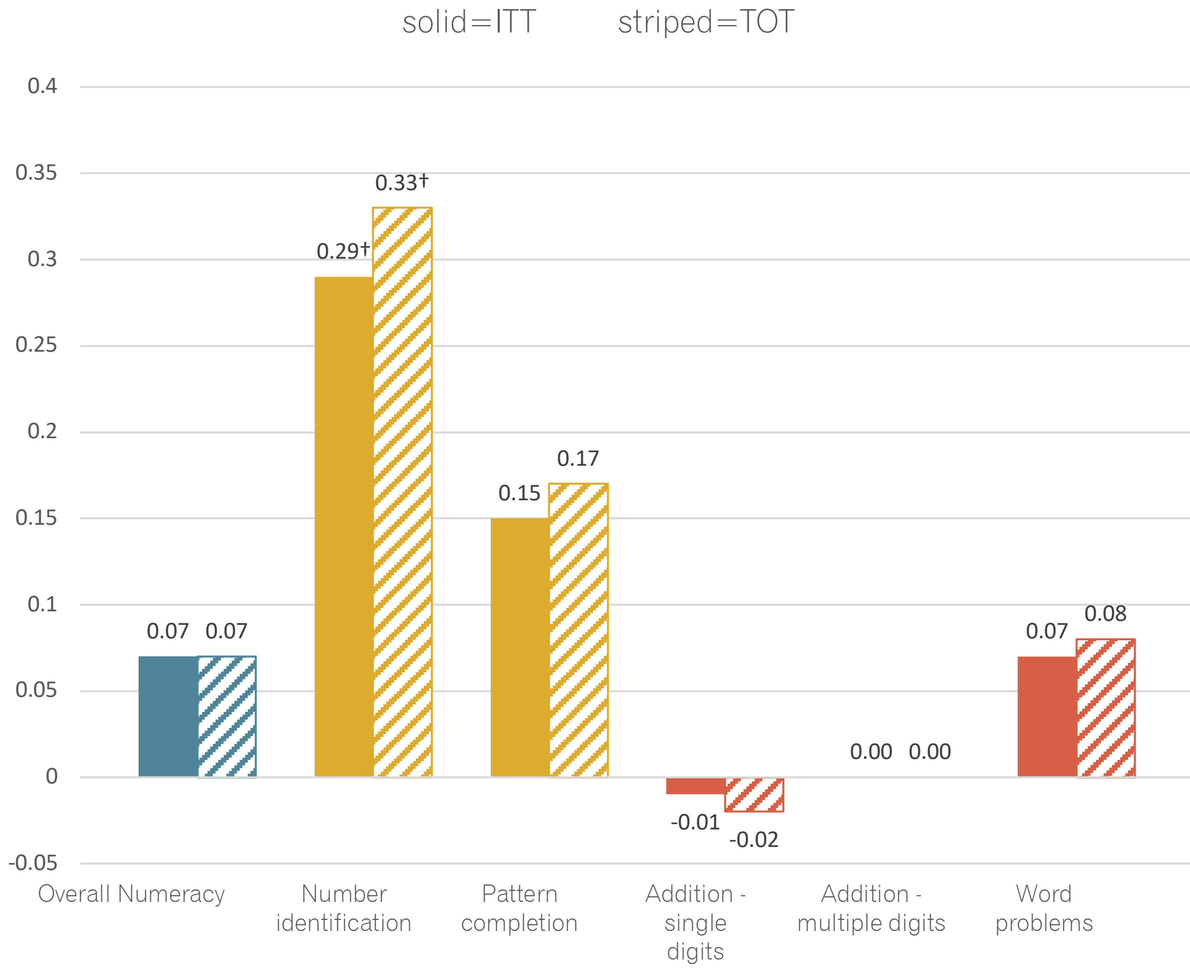
* Indicates a statistically significant result after correction for multiple hypotheses.
† Indicates statistical significance before correction for multiple hypotheses.
! Indicates that the estimate should be interpreted with extreme caution. See note 2 below.
[1]About 88% of the children in the treatment groups attended at the 50% rate. The number of days the learning centers were open differed by school, with the peri-urban center open a total of 130 days and the urban center open a total 106 days.
[2]The % added value compares the treatment effect size to the average control group growth over the 8 months of the study. This ratio of the standardized treatment growth to control group growth is expressed as a percent. This percent is then translated into added months of learning. The added value measures should always be interpreted with caution and particularly large estimates (indicated with !) should be interpreted with extreme caution.
[3]Percentile growth indicates the change in percentile rank that the average (median) student would be expected to experience if they received the treatment.

Figure 2. Effect sizes for literacy for the full treatment group (ITT) and those who attended at least 50% of treatment days (TOT)



*Indicates a statistically significant result after correction for multiple hypotheses.

Figure 3. Effect sizes for numeracy for the full treatment group (ITT) and those who attended at least 50% of the treatment days (TOT)



†Indicates statistical significance before correction for multiple hypotheses.

How far did children progress toward targeted learning outcomes?

Imagine's ultimate research goal is understanding whether children can become literate and numerate using child-directed, technology-enabled learning. Consequently, in addition to understanding what the relative impacts of the intervention were over standard instruction in our pilot study, we also wanted to learn how far children progressed toward reading fluency with comprehension and comparable numeracy skills after using the *onecourse* software over 8 months of the 2018–19 school year.

To address this question, we compared progress of the treatment and control groups on published benchmarks for reading and mathematics based on the oral passage reading and pattern completion subtests of EGRA and EGMA, respectively. Specifically, we examined movement between baseline and endline along the following reading and math continuums:

- Oral reading fluency benchmarks[30]
  - Non-reader = 0 correct words per minute (cwpm) of reading connected text
  - Low reader = 1 to 19 cwpm
  - Emergent reader = 20 to 39 cwpm
  - Fluent reader = 40+ cwpm
- Pattern completion benchmarks[31]
  - Starting mathematician = 0% correct on pattern completion subtest
  - Low mathematician = 1–29% correct
  - Emergent mathematician = 30%–59% correct
  - Mathematician = 60%+ correct

Table 6 shows the percentage distributions of the treatment and control groups.

- After 8 months, statistically larger proportions of the treatment groups than the control group increased at least one benchmark level in reading and math. Specifically, 23% of the literacy treatment group compared with 15% of the control group increased at least one level along the oral reading fluency continuum. And 44% of the numeracy treatment group compared with 32% of the control group increased at least one level along the pattern completion continuum.

Despite these statistically larger gains, only 8% of the literacy treatment group attained emergent reader status at the end of the study and 1% attained fluent reader status. In math, 18% of the numeracy treatment group attained emergent mathematician status at the end of the study and 1% attained mathematician status. These findings suggest that it will take longer than one school year of intervention for children to attain ultimate learning goals for literacy and numeracy.

[30] The Malawi Ministry of Education, Science and Technology has proposed a Standard 2 benchmark for oral reading fluency of 40 correct-words-per-minute (cwpm) (USAID 2015a). The Zambian government set additional interim Standard 2 benchmarks for the Chichewa ("Nyanja") language (USAID 2015b).

[31] The Zambian government set the Standard 2 benchmarks for pattern completion based on EGMA conducted in the Chichewa ("Nyanja") language (USAID 2015b).

Table 6. Percentage distribution of study participants at baseline and endline by reading and math continuum category and the percentage in each category who increased or decreased at least one level on the continuum between baseline and endline

	Count (N) at Baseline	% distribution at Baseline	% distribution at Endline	Baseline to endline % who Increased	Baseline to endline % who Decreased
Reading continuum					
Study participants, all	384	100%	100%	19%	1%
Non-reader	366	95%	80%	17%	—
Low	15	4%	12%	53%	33%
Emergent	3	1%	7%	67%	—
Fluent	0	0%	1%	—	—
Literacy treatment group, all	197	100%	100%	23%*	1%
Non-reader	189	96%	77%	21%	—
Low	6	3%	14%	77%	17%
Emergent	2	1%	8%	100%	—
Fluent	0	0%	1%	—	—
Control group, all	187	100%	100%	15%	2%
Non-reader	177	95%	84%	14%	—
Low	9	4%	10%	44%	44%
Emergent	1	1%	6%	0%	—
Fluent	0	0%	0%	—	—
Mathematics continuum					
Study participants, all	384	100%	100%	38%	12%
Starting mathematician	190	51%	28%	62%	—
Low	172	44%	56%	17%	19%
Emergent	20	5%	15%	0%	55%
Mathematician	2	1%	1%	—	100%
Math treatment group, all	198	100%	100%	44%*	11%
Starting mathematician	97	49%	22%	69%	—
Low	89	45%	60%	22%	13%
Emergent	10	4%	18%	0%	70%
Mathematician	2	1%	1%	—	100%
Control group, all	186	100%	100%	32%	13%
Starting mathematician	93	52%	35%	65%	—
Low	83	42%	52%	11%	25%
Emergent	10	4%	12%	0%	40%
Mathematician	0	0%	1%	—	—

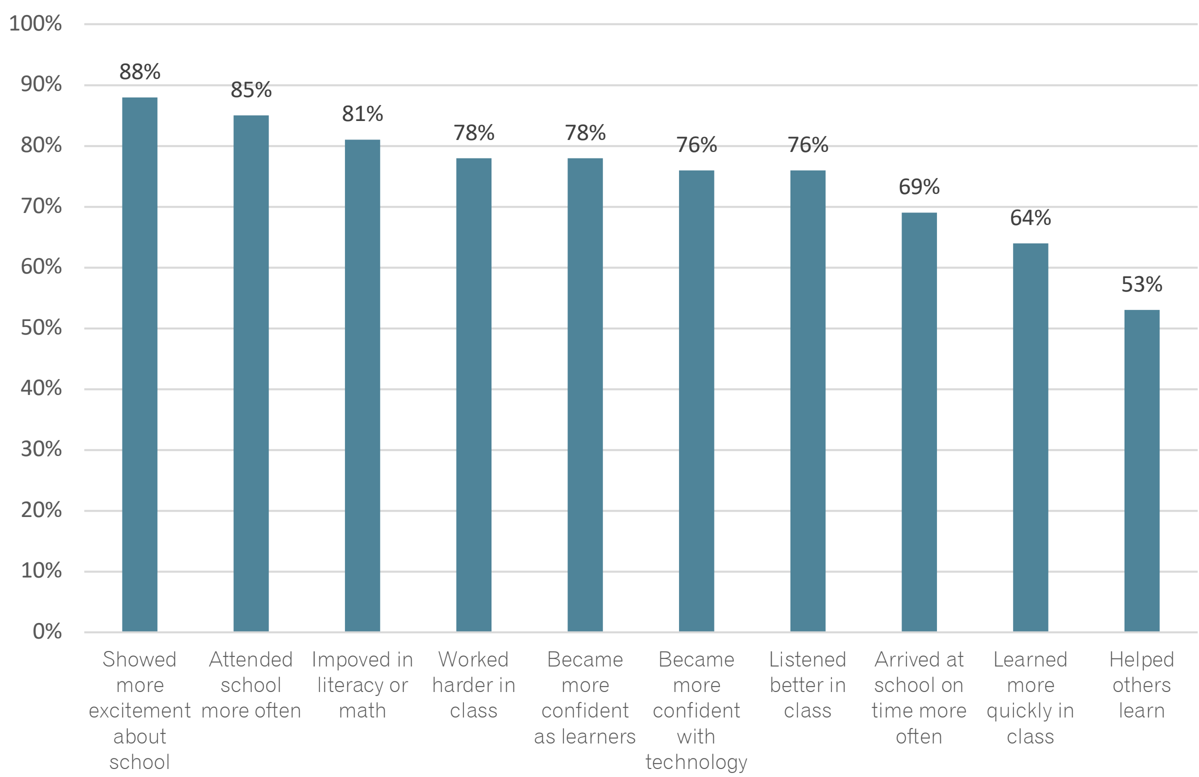
*Indicates that the treatment group increase was statistically greater than the control group increase.
—Not applicable.
NOTE: The benchmarks are based on the oral passage reading and pattern completion subtests of EGRA and EGMA, respectively. The pattern completion subtest has fewer items than the oral passage reading subtest, so performance on the above math benchmarks is likely to be more variable than on the reading benchmarks between baseline and endline due to potentially greater measurement error in the pattern completion subtest.

Stakeholder Perceptions of Program Impacts

In addition to analyzing gains on EGRA and EGMA, we also conducted a non-experimental end-of-year survey of 43 adult stakeholders to understand their perceptions of program impact. Appendix C contains the English version of the interview protocol, which was translated into Chichewa by our University Malawi research partners. The survey included the universe of Standard 2 teachers and school administrators and an opportunistic sample of parents and teachers, so the results are suggestive but not conclusive.

As seen in **Figure 4**, the parents, teachers, and school and community leaders who participated in the interviews were overwhelmingly positive about the impact of the program on learners. More than three-fourths of the 43 adults interviewed at the end of the school year at the two pilot schools reported that learners showed “a lot” of improvement in the following areas because of the tablet program: children's excitement about school (88%), attendance (85%), achievement in literacy or math (81%), work effort (78%), and confidence as learners (78%).

Figure 4. Percentage of adult stakeholders* reporting "a lot" of improvement in learners due to the tablet program



*Includes 43 parents, teachers, and school and community leaders at the two study schools.

V. Exploring Subgroup Results

We were also interested in exploring whether impacts differed at the two government primary schools and whether gender and age category were associated with outcomes. In this chapter, we present results of the exploratory analyses of the tablet intervention's impacts on gains in EGRA and EGMA test scores for these subgroups, focusing on the ITT analysis. Appendix D shows the statistical test results for the subgroup differences.

Differences at the Two Study Schools

We were interested in exploring whether impacts differed at the two schools. Specifically, we wanted to know whether the better implementation and better attendance we observed at the peri-urban school (discussed in Chapter VI) were associated with better outcomes than at the urban school.

As seen in Table 7

- The peri-urban school exhibited a statistically significant positive effect in overall literacy learning (.40 standard deviations) and the urban school exhibited a substantively important positive effect in overall literacy (.28 standard deviations) for the full treatment group (ITT).

The difference between the two schools was not statistically significant.

- The peri-urban school exhibited a statistically larger positive effect than the urban school in nonword reading (.35 vs. .04 standard deviations) and the urban school exhibited a statistically larger positive effect than the peri-urban school in listening comprehension (.33 vs. .05 standard deviations) for the full treatment group (ITT).

These were the only literacy subtests with statistically significant differences between the two schools.

- The peri-urban school exhibited a (marginally) statistically larger positive effect in number identification (.43 vs. .15 standard deviations) for the full treatment group (ITT).

This was the only numeracy subtest with a statistically significant difference between the two schools.

The above results suggest that the better implementation and better attendance at the peri-urban school may have contributed to greater impacts in both literacy and numeracy at this school. However, the urban school exhibited a stronger impact in listening comprehension.

Table 7. Effect sizes by school for the full treatment group (ITT) and for those who attended at least 50% of the treatment days (TOT)

Targeted Outcome Measures	Full treatment group (ITT)		Those with >=50%[1] attendance (TOT)	
	Urban	Peri-urban	Urban	Peri-urban
Literacy				
Overall EGRA average percent correct	.28†	.40*	.31†	.48*
Subtest percent correct				
Simple view of reading				
Nonwords/Invented Words	.04	.35*	.05	.42
Listening Comprehension	.33†	.05	.37*	.06
Reading Comprehension	.20	.30†	.23	.36†
Additional targeted outcomes				
Letter Naming	.21	.11	.23	.13
Oral Reading Fluency	.12	.29†	.13	.34†
Numeracy				
Overall EGMA total percent correct	.17	-.03	.18	-.03
Subtest percent correct				
Number sense				
Number Identification	.15	.43*	.16	.50*
Pattern Completion	.24	.07	.26	.08
Arithmetic operations				
Addition Level 1	.11	-.13	.11	-.15
Addition Level 2	-.12	.11	-.13	.13
Word Problems	.16	-.01	.17	-.01
Observations (EGRA/EGMA)	201/202	181/180	201/202	181/180

* Indicates a statistically significant result after correction for multiple hypotheses.
† Indicates statistical significance before correction for multiple hypotheses.
[1]About 88% of the children in the treatment groups attended at the 50% or higher rate. The number of days the learning centers were open differed by school, with the peri-urban center open a total of 130 days and the urban center open a total 106 days.

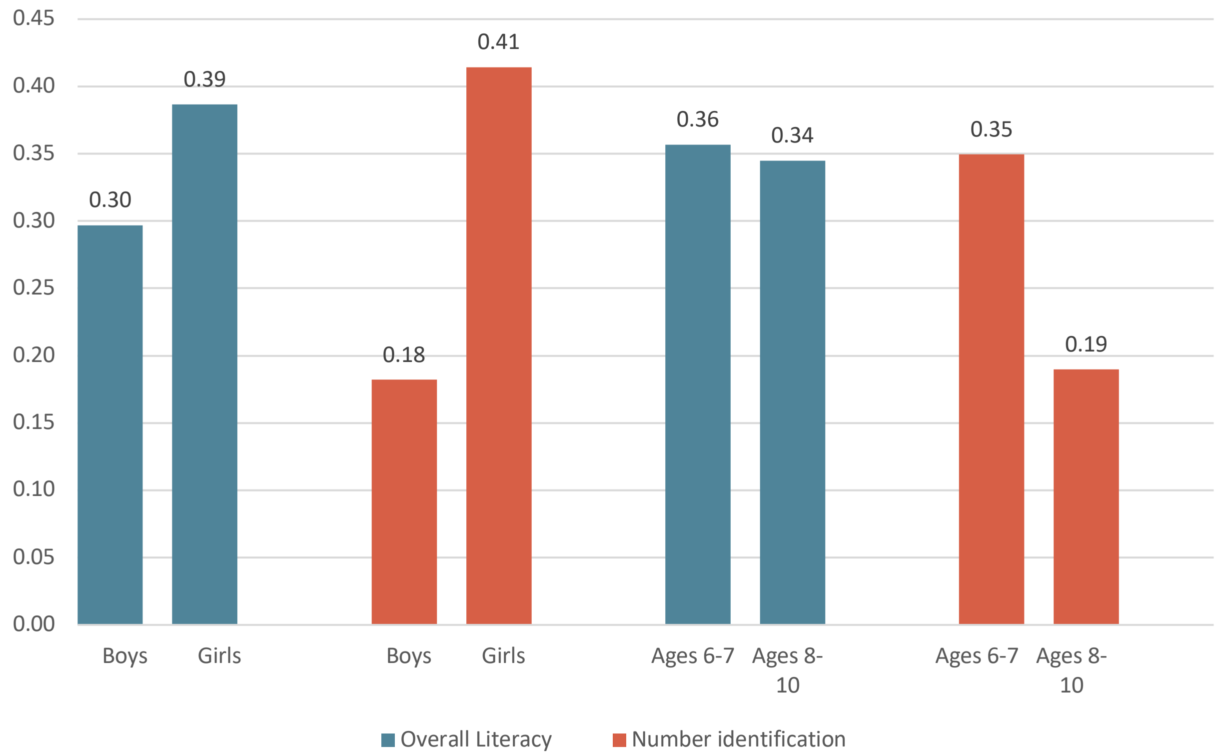
Gender and Age Category Differences

We were interested in examining how gender and age category were associated with outcomes, hypothesizing that the "blind" nature of tablet instruction—where the software does not differentiate users based on their gender or age category—may generate more equitable learning outcomes for subgroups than some classrooms would do. We pre-determined that these analyses would be exploratory only and focused on the ITT analysis.

- Intervention effects on literacy and numeracy learning were not statistically different for either gender or age category.

Figure 5 depicts the effect sizes for boys and girls and for younger and older children in overall literacy and in number identification. As shown in Appendix D, none of these differences was statistically significant.

Figure 5. Comparing effect sizes by gender and age category for selected outcomes for the full treatment group (ITT)



VI. Discussion

Our analyses produced statistically significant and substantively important positive effects in overall literacy and in number identification. The analyses also suggested consistently positive effects in all targeted literacy areas and in both number sense skills in mathematics. However, some challenges with implementation during the pilot study may have contributed to somewhat lower-than-expected learning gains, particularly in mathematics.

Implementation Challenges

With our VSO, onebillion, and school partners, we made a concerted effort to implement the intervention with fidelity. Despite our best efforts, however, we encountered some implementation challenges that may have attenuated the impact results from this pilot study. As described in Chapter III, we monitored the fidelity of implementation through regular review of tablet usage data and through monthly monitoring visits conducted by the University of Malawi's Chancellor College. These monitoring efforts uncovered the following implementation challenges.

Implementation challenges included

- Software limitations. The *onecourse* math application in use in 2018–19 in the UT schools across Malawi, including our two study schools, did not include the periodic quizzes that had previously been included in the apps. Prior research had been done on the version containing quizzes.[32] (The literacy application had not previously contained quizzes.) Without the math quizzes, children proceeded quickly through the available curriculum. Because the number of instructional units available in math was lower than in literacy (282 versus 1,248 units), this resulted in most children repeating the limited math curriculum over the 8 months of the study. This could have contributed to null findings among the arithmetic operations skills and in overall numeracy: as children in the treatment group repeated the tablet math content, the children in the control group may have caught up. Nevertheless, the math intervention produced a substantively important positive effect in number identification and suggested a positive effect in pattern completion. Updated versions of both the math and literacy applications that include periodic diagnostic assessments (with opportunities for remedial work) as well as significantly more content in both subjects were released in summer 2019. We are using these updated versions for a new study with a second cohort of Standard 2 children in 2019–20.
- Lower time on task than anticipated. We designed the study to provide 90 hours of tablet session time during the 8-month intervention period.[33] However, challenges with high absenteeism and lost instructional time resulted in an average accumulation of 53 hours of tablet session time. In addition to the common reasons for absenteeism described in Chapter II, the 2018-19 school year also witnessed Cyclone Idai and additional storms that brought significant flooding to Malawi and made it unsafe for children to walk to school, especially in the more rural area. In addition, the learning centers were sometimes closed during exam periods (as many as nine weeks of instructional time could be lost to exam preparation, conduct, and grading alone) and during the first week of each term as classes would resume. This lost instructional time was a particular

[32] Pitchford (2015); Pitchford et al. (2017).
[33] This maximum time-on-task target assumed an 80% attendance rate on days the schools were officially in session.

problem at the urban school. Furthermore, 2019 was a general election year for Malawi and schools were closed on election day in May and a number of subsequent days due to post-election unrest, especially in the urban area. Ultimately, the learning centers were open for 130 days at the peri-urban school and 106 days at the urban school out of an official 195-day school year. Children attended the learning center on average 93 days at the peri-urban school and 79 days at the urban school.

- Additional technical issues. We also faced miscellaneous technical issues, including challenges obtaining reliable, strong, and inexpensive headphones (broken headphones created a noisier-than-ideal learning center environment until they were replaced); a few malfunctioning learning units prevented students from progressing until the units were deactivated; and occasionally the iPads would freeze and need to be restarted, resulting in some minutes of lost learning time.
- Differential implementation at the two study schools. In addition to keeping the learning center open for more days, the peri-urban school also had stronger learning center management, which resulted in quicker identification and resolution of the technical issues described above.

Implications

The implementation challenges described above may have attenuated impact results from the pilot study, but they also suggest that improved implementation could produce even greater impacts. Specifically, the updated *onecourse* applications that contain diagnostic assessments and more content could produce greater learning impacts during the new study that we are conducting with a second cohort of Standard 2 children in 2019–20. This new study will also allow us to improve the implementation in other ways as well, such as procuring sturdier headphones and addressing other technical issues.

We sought to conduct our research under normal, challenging conditions to understand better what impacts could be anticipated at scale. Barriers to accumulating time on task in these contexts are daunting. And, as indicated by the benchmark analysis in Chapter V, while more children in the treatment groups than in the control group made progress on reading and math benchmarks, few progressed beyond the lowest reading and math levels. These findings may require adjusting expectations about the dosage, lapsed time, or delivery model that may be required to attain learning goals.

Conclusion

The current study contributes to the body of evidence on the impacts of education technology. Specifically, the study adds rigorous evidence about the longer-term impacts of a child-directed, technology-enabled learning intervention (i.e., onebillion's *onecourse* software) on literacy and numeracy learning over 8 months. The experimental design for this study ensured that differences in learning gains between the treatment and control groups could be attributed to the tablet intervention rather than to pre-existing differences among the groups. And baseline equivalence of the final analytic sample supported the causal validity of the study.

The ITT impact analysis demonstrated statistically significant and substantively important positive effects in overall literacy and in number identification. The analysis also suggested positive effects in all other targeted literacy areas and in pattern completion. And the TOT analysis showed that meeting a minimum threshold of attendance in the intervention had a statistically significant impact on outcomes. Additional exploratory analyses suggested that the better implementation and better attendance at the peri-urban school may have produced stronger impacts than at the urban school. And differences in intervention effects by gender and age category were not statistically significant.

Some challenges with implementation during the pilot study year may have attenuated the impact findings. Building on lessons learned this year, we anticipate better implementation and greater learning impacts from a new study with a second cohort of Standard 2 children during the 2019-20 school year. However, we also understand that challenges with time on task in these contexts may require adjusting expectations about the dosage, lapsed time, and delivery model that may be needed to attain ultimate learning goals.

We are very excited about the positive effects emerging from this rigorous pilot study. We look forward to applying lessons learned from the pilot year in our new study with a second cohort in 2019–20 and to investigating further what works, for whom, and under what conditions. We continue to be confident that child-directed, technology-enabled learning can help children to become literate and numerate, although know that it will take time and determination to achieve these goals.

References

Baird, M. D., & Pane, J. F. (2019). Translating standardized effects of education programs into more interpretable metrics. *Educational Researcher*, *48*(4), 217–228.

Bernstein, L., Rappaport, C. D., Olsho, L., Hunt, D., & Levin, M. (2009). Impact Evaluation of the US Department of Education's Student Mentoring Program. Final Report. NCEE 2009-4047. *National Center for Education Evaluation and Regional Assistance*.

Clearinghouse, W. W. (2017). Standards handbook (version 4.0). *Washington, DC: Institute of Education Sciences*.

Dong, N., & Maynard, R. (2013). PowerUp!: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasi-experimental design studies. *Journal of Research on Educational Effectiveness*, *6*(1), 24–67.

Gleason, P., Clark, M., Tuttle, C. C., & Dwoyer, E. (2010). The Evaluation of Charter School Impacts: Final Report. NCEE 2010-4029. *National Center for Education Evaluation and Regional Assistance*.

Gove, A., & Wetterberg, A. (2011). *The Early Grade Reading Assessment: Applications and Interventions to Improve Basic Literacy.* ERIC.

Kraft, M. A. (2018). Interpreting effect sizes of education interventions. *Brown University Working Paper*.

Pitchford, N. J. (2015). Development of early mathematical skills with a tablet intervention: a randomized control trial in Malawi. *Frontiers in Psychology*, *6*, 485.

Pitchford, N. J., Hubber, P. J., & Chigeda, A. (2017). Unlocking Talent: Improving Learning Outcomes of Primary School Children in Malawi. Unpublished report.

RTI International. (2014). EGMA Toolkit. Washington, DC: Research Triangle International.

RTI International. (2016). Early grade reading assessment toolkit (2nd ed.). North Carolina: Research Triangle Park.

Schochet, P. Z. (2016). Statistical theory for the RCT-YES software: Design-based causal inference for RCTs (NCEE 2015–4011). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education.

USAID. (2015a). *Proposing Benchmarks for Early Grade Reading in Malawi*.

USAID. (2015b). *Proposing Benchmarks and Targets for Early Grade Reading and Mathematics in Zambia*.

IMAGINE

THE FUTURE OF LEARNING

Tablet-based Learning for Foundational Literacy and Math:
An 8-month RCT in Malawi

Appendix A: Information on *onecourse*

- This appendix provides selected pages from the *onecourse* handbook, which can be retrieved in full at https://ob-assets.netlify.com/onecourse_2018_digital_handbook-9a827155f0c479bb70018016f31d286d6c854746a7d0bfec0256f76a762ef3fb.pdf

- Imagine's pilot study used the "*onecourse* for schools" version of the app.

- The appendix also provides several screenshots from the literacy app in Chichewa.

onecourse

ONE APP

ALL OUR CHILDREN

NUMERATE

AND WITH A

LOVE

OF READING

onecourse
handbook

by onebillion

onebillion learners is a registered charity in England and Wales N° 1159480

onebillion.org

CONTENTS

01

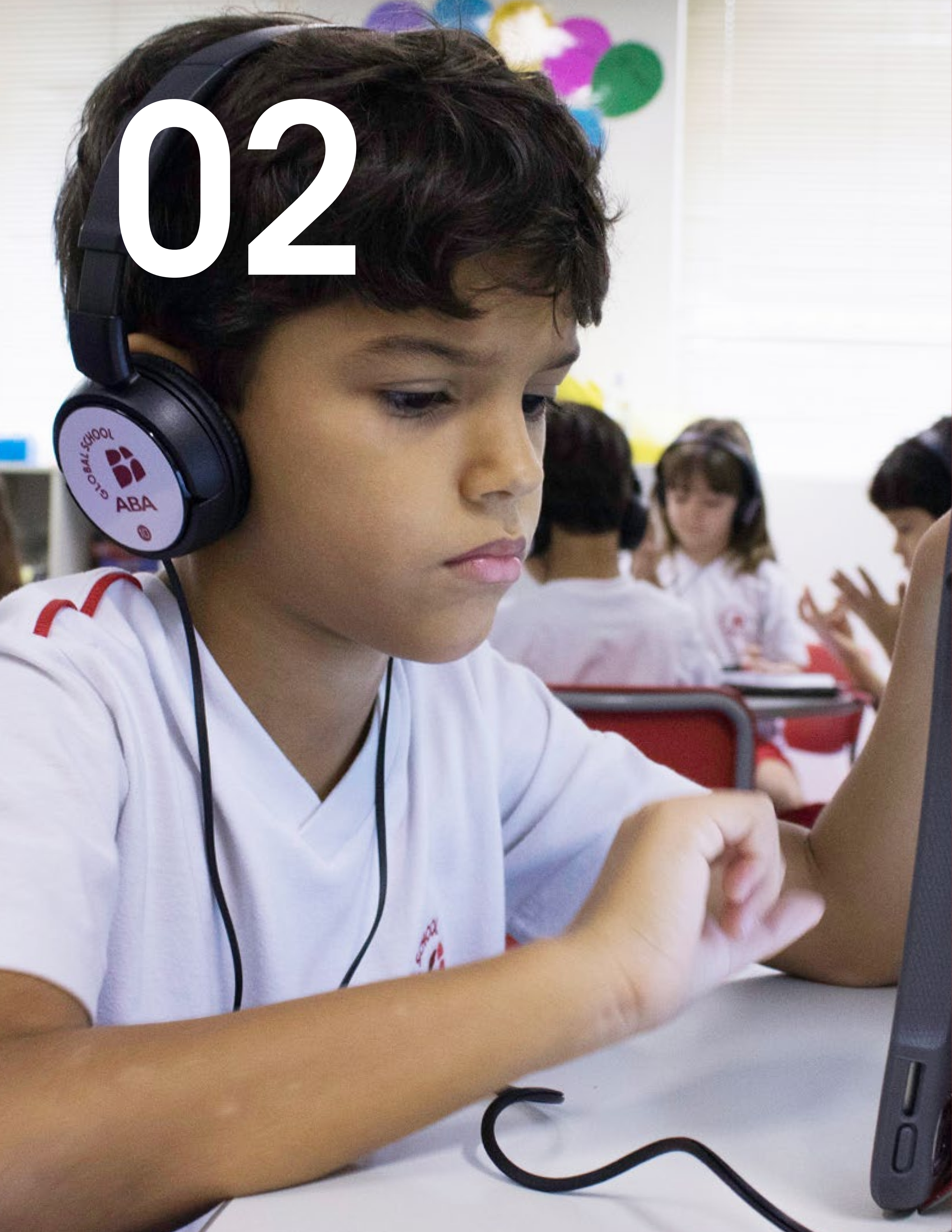
onecourse

one app designed for every child

- in the child's own language
- for use in both schools and in a wider community setting
- works on tablets and smartphones, both iOS and Android
- built by international non-profit **onebillion,** based in London
- designed to take a child from zero fluency to reading with comprehension
- delivers numeracy and a love of reading

This handbook gives a more detailed view of the content of **onecourse,** and the growing evidence base that shows significant learning outcomes.

02



An overview of onecourse

onecourse is our response to the global education challenge: a comprehensive, personalised learning software, which enables children anywhere in the world to become literate and numerate in their own language.

Onecourse software is adaptable to different contexts and cultures, and compatible with tablets and smartphones running both iOS and Android operating systems.

The two versions

There are two versions of onecourse: onecourse for schools and onecourse for communities.

1. onecourse for schools

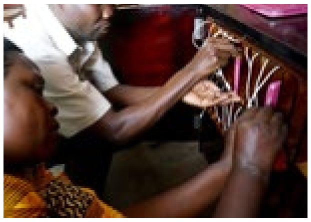
onecourse is implemented in a school environment with support from the Ministry of Education.

Key Features

- Teachers register and group children according to age or ability, as in a normal classroom.
- Literacy and numeracy are presented as separate strands. Lessons in each are planned and allocated by qualified teachers and teaching assistants.
- Children's progress is monitored via a low-powered local server.
- An administrative tool allows teachers to track progress and give instant support.

The numeracy material is presented in the form of graded topics. The literacy material is divided into ten levels. Children work through short learning units at their own pace.

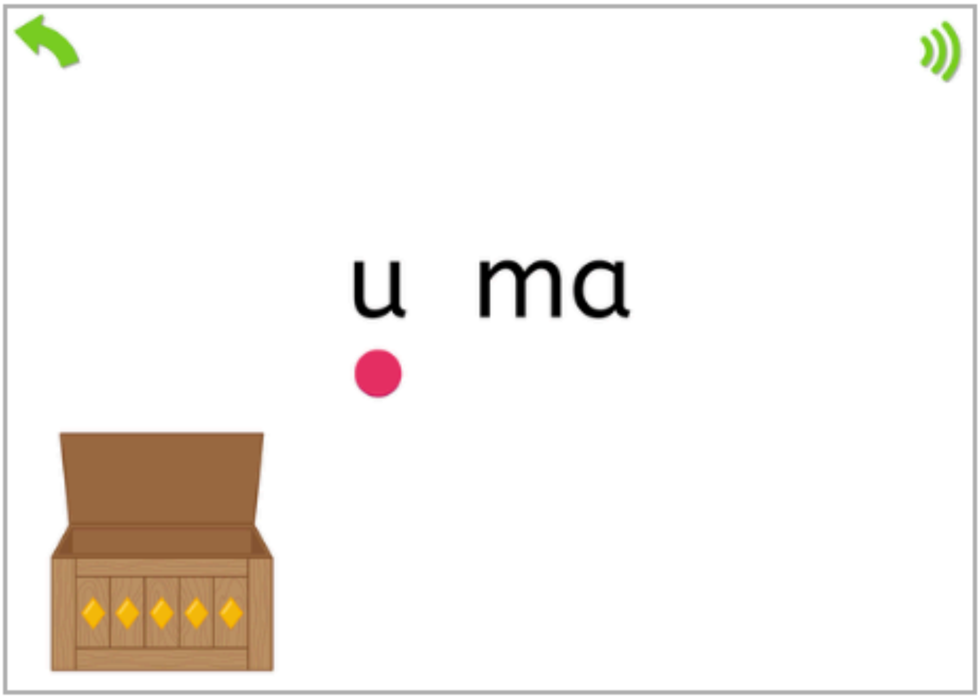
onecourse schools version is being used under the Unlocking Talent initiative; a joint programme between onebillion and Voluntary Service Overseas (VSO International).

Screenshots of the Literacy Application in Chichewa

<u>Example short units for literacy include the following:</u>

1. Core lessons, such as making syllables. In the activity shown below, words fly out of the box, and the child taps the button to hear the syllables.



2. Practice activities such as making words that the child hears. In the practice activity shown below, the child drags words to complete phrases.

3. Stories are initially read to the child with only the story title in text. Gradually, the app moves through different story modes, until finally the child can read alone with the option to touch and hear words they find difficult.

IMAGINE

THE FUTURE OF LEARNING

Tablet-based Learning for Foundational Literacy and Math:
An 8-month RCT in Malawi

Appendix B: Descriptive Statistics for Outcome Measures

Descriptive statistics for EGRA baseline percent correct scores

	Chichewa Group		Control Group		Combined Sample		Cronbach's Alpha
	Mean	SD	Mean	SD	Mean	SD	α
Baseline Decoding (Nonword reading) Percent Correct	0.01	0.061	0.01	0.070	0.01	0.066	0.98
Baseline Listening Comprehension Percent Correct	0.40	0.206	0.40	0.238	0.40	0.222	0.47
Baseline Reading Comprehension Percent Correct	0.00	0.040	0.00	0.019	0.00	0.031	0.72
Baseline Letter Naming Percent Correct	0.04	0.080	0.04	0.076	0.04	0.078	0.97
Baseline Oral ReadingFluency	0.01	0.062	0.01	0.048	0.01	0.055	0.98
Baseline Composite EGRA Average Percent Correct M	0.11	0.066	0.12	0.065	0.12	0.065	0.75
N	222		226		448		448

Descriptive statistics for EGRA endline percent correct scores

	Chichewa Group		Control Group		Combined Sample		Cronbach's Alpha
	Mean	SD	Mean	SD	Mean	SD	α
Endline Decoding (Nonword reading) Percent Correct	0.06	0.137	0.04	0.116	0.05	0.127	0.96
Endline Listening Comprehension Percent Correct	0.48	0.215	0.42	0.225	0.45	0.222	0.37
Endline Reading Comprehension Percent Correct	0.02	0.082	0.00	0.025	0.01	0.062	0.55
Endline Letter Naming Percent Correct	0.11	0.133	0.10	0.125	0.11	0.129	0.97
Endline Oral ReadingFluency	0.06	0.151	0.04	0.119	0.05	0.137	0.98
Endline Composite EGRA Average Percent Correct Me	0.19	0.109	0.16	0.100	0.18	0.105	0.85
N	199		187		386		386

Descriptive statistics for EGMA baseline percent correct scores

	Masamu Group		Control Group		Combined Study Sample		Cronbach's Alpha
	Mean	SD	Mean	SD	Mean	SD	α
Baseline Number Identification Percent Correct	0.22	0.156	0.23	0.157	0.23	0.156	0.70
Baseline Pattern Completion Percent Correct	0.09	0.112	0.08	0.105	0.08	0.108	0.60
Baseline Word Problems Percent Correct	0.23	0.262	0.22	0.256	0.22	0.259	0.59
Baseline Addition Level 1 Percent Correct	0.33	0.400	0.31	0.375	0.32	0.388	0.89
Baseline Addition Level 2 Percent Correct	0.08	0.143	0.06	0.140	0.07	0.142	0.54
Baseline Composite EGMA Percent Correct Measure	0.18	0.144	0.18	0.132	0.18	0.138	0.80
N	223		225		448		448

Descriptive statistics for EGMA endline percent correct scores

	Masamu Group		Control Group		Overall Study Sample		Cronbach's Alpha
	Mean	SD	Mean	SD	Mean	SD	α
Endline Number Identification Percent Correct	0.41	0.222	0.36	0.205	0.39	0.215	0.78
Endline Pattern Completion Percent Correct	0.16	0.128	0.13	0.129	0.14	0.129	0.64
Endline Word Problems Percent Correct	0.39	0.265	0.36	0.258	0.38	0.262	0.55
Endlline Addition Level 1 Percent Correct	0.58	0.407	0.59	0.394	0.59	0.400	0.87
Endline Addition Level 2 Percent Correct	0.19	0.168	0.17	0.184	0.18	0.176	0.48
Endline Composite EGMA Percent Correct Measure	0.34	0.189	0.32	0.177	0.33	0.183	0.86
N	199		187		386		386

Descriptive statistics for EGRA baseline fluency rates

	Chichewa Group		Control Group		Combined Sample	
	Mean	SD	Mean	SD	Mean	SD
Baseline Decoding (Nonword reading) Correct Nonwo:	0.57	3.07	0.68	3.51	0.63	3.30
Baseline Oral Reading Fluency Correct Words Per Min	0.62	3.76	0.56	2.91	0.59	3.35
N	222		226		448	

Descriptive statistics for EGRA endline fluency rates

	Chichewa Group		Control Group		Combined Sample	
	Mean	SD	Mean	SD	Mean	SD
Endline Decoding (Nonword reading) Correct Nonwor	3.03	6.85	2.07	5.78	2.56	6.37
Endline Oral Reading Fluency Correct Words Per Minu	3.94	9.23	2.59	7.23	3.29	8.34
N	199		187		386	

IMAGINE

THE FUTURE OF LEARNING

Tablet-based Learning for Foundational Literacy and Math:
An 8-month RCT in Malawi

Appendix C: Stakeholder Interview Questionnaire

Interview Questionnaire

Interviewee Name______________________________ Date________________________________

Role____ School_______________________________

#	*Question*	*Response*	*Notes/Examples*
	Because of the tablet program, did your learner(s)….		
1	Attend school more often?	No Yes-A little Yes-A lot	
2	Arrive at school on time more often?	No Yes-A little Yes-A lot	
3	Work harder in class?	No Yes-A little Yes-A lot	
4	Listen better in class?	No Yes-A little Yes-A lot	
5	Pay less attention in class?	No Yes-A little Yes-A lot	
6	Improve in literacy or maths?	No Yes-A little Yes-A lot	
7	Learn more quickly in class?	No Yes-A little Yes-A lot	
8	Fall behind in class?	No Yes-A little Yes-A lot	
9	Show more excitement about school?	No Yes-A little Yes-A lot	
10	Become more confident as learners?	No Yes-A little Yes-A lot	
11	Become more confident with technology?	No Yes-A little Yes-A lot	
12	Help others learn?	No Yes-A little Yes-A lot	

Did the tablet program have any other impacts on your learner(s), good or bad?

IMAGINE

THE FUTURE OF LEARNING

Tablet-based Learning for Foundational Literacy and Math:
An 8-month RCT in Malawi

Appendix D: Subgroup Statistical Tests

Summary of Subgroup Differences for EGRA Outcomes (ITT)

	School Differences (Urban Versus Rural)		Gender Differences (Male versus Female)		Age Category Differences (Younger versus Older)	
EGRA Outcomes	Difference in Effect Sizes	Impact Estimate P-value	Difference in Effect Sizes	Impact Estimate P-value	Difference in Effect Sizes	Impact Estimate P-value
Decoding (Nonword reading)	-0.31	**0.03**	-0.24	0.12	0.03	0.88
Listening Comprehension	0.28	**0.02**	0.00	0.96	-0.02	0.91
Reading Comprehension	-0.10	0.88	-0.14	0.57	0.10	0.62
Letter Naming	0.09	0.56	-0.31	0.10	-0.09	0.48
Oral Reading Fluency	-0.17	0.27	0.00	0.68	0.08	0.73
Composite EGRA Average Percent Correct Measure	-0.13	0.41	-0.09	0.61	0.01	0.92

Bolded values indicate differences that were statistically significant at the p<.05 level.

Summary of Subgroup Differences for EGMA Outcomes (ITT)

	School Differences (Urban Versus Rural)		Gender Differences (Male versus Female)		Age Category Differences (Younger versus Older)	
EGMA Outcomes	Difference in Effect Sizes	Impact Estimate P-value	Difference in Effect Sizes	Impact Estimate P-value	Difference in Effect Sizes	Impact Estimate P-value
Number Identification	-0.28	*0.07*	-0.23	0.32	0.16	0.39
Pattern Completion	0.17	0.26	-0.02	0.96	-0.23	0.31
Word Problems	0.17	0.22	-0.16	0.45	-0.01	1.00
Addition Level 1	0.23	0.92	-0.16	0.49	-0.03	0.90
Addition Level 2	-0.23	0.13	-0.01	0.98	0.05	0.82
Composite EGMA Percent Correct Measure	0.20	0.19	-0.09	0.20	-0.05	0.85

Italics indicate a difference that was statistically significant at the p<.10 level.