FINAL EVALUATION REPORT

THE KIRIBATI EDUCATION IMPROVEMENT PROGRAM (KEIP) TABLET TRIAL IMPACT EVALUATION

DATE: APRIL 2019 PRESENTED TO: COFFEY PRESENTED BY: NORC

NORC at the UNIVERSITY of CHICAGO

Table of Contents

Table of Tables

Table of Figures

Acronyms

CEFR	Common European Framework Reference
COR	Classroom observation rubric
DFAT	Department of Foreign Affairs and Trade
DQR	Data Quality review
EDR	Evaluation Design Report
EGMA	Early Grade Math Assessment
EGRA	Early Grade Reading Assessment
ESL	English as a Second Language
FGD	Focus Group Discussions
GDP	Gross Domestic Product
ICT	Information and Communications Technology
IPW	Inverse Probability Weighting
ITC	Information Technology Communication
KEIP	Kiribati Education Improvement Program
KII	Key Informant Interviews
KTT	Kiribati Tablet Trial
MDES	Minimum Detectable Effect Size
MoE	Ministry of Education
PCA	Principal Component Analysis
PRIM	Primary Math and reading Initiative
RCT	Randomized Control Trial
RQ	Research question
STS	School-To-School International
TOEFL	Teaching of English as a Foreign Language
USD	United States Dollar
VAM	Value-Added Models
WPM	Words per Minute
WWC	What Works Clearinghouse

Executive Summary

This report presents the results of the impact evaluation of the Kiribati Tablet Trial (KTT), an intervention that used preloaded tablet computers to provide teachers and students with curriculum aligned learning materials, in an effort to improve teaching and learning. The trial ran during the 2018 school year – from February to November. For this evaluation we conducted a randomized control trial (RCT), where five schools were randomly assigned to treatment and five to control groups.

The study finds that the program had positive impacts on students' English reading skills. Out of the five subtasks evaluated, we find positive and statistically significant effects for three of them, namely reading made-up words, oral reading fluency and reading comprehension. For made-up words, students in the treatment group were able to read 4.8 wpm more than in the control group, while the impact for oral reading fluency was 5.1 wpm. For reading comprehension, students in the treatment group answered correctly 0.53 more questions than their counterparts in the control group. We do not find evidence that the program affected students' math skills. Out of the eight evaluated subtasks, we found positive impacts for six of them, but a statistically significant effect for only the multiplication subtask.

We also assessed the impact of the program on teachers' English skills. Albeit the small sample size for this part of the analysis prevented a more robust study of this aspect of the program, we find evidence that suggests a positive impact for three of the four conducted English assessments. In addition, teachers reported that the tablet helped them improve their English skills.

We analyze whether the program had a stronger effect for struggling readers than for the rest of students, defining struggling readers as students with reading skills below or equal to the median baseline. We do not find evidence that the program had a differential effect by achievement level at baseline.

In terms of program implementation, KTT was successful in addressing two key conditions for tablet programs to work. First, tablet content was curated so there was a correspondence with the curriculum. Second, teachers incorporated student tablets and other project resources in class activities.

There are two key limitations to this study. First, the 10 evaluation schools were purposefully selected. In effect, one of the main reasons why these schools were selected was their easy accessibility. This implies that the evaluation results cannot be extrapolated to other schools. Second, given the small sample size, the program needed to have a large impact for the design to measure it with conventional precision. This could explain why we did not find impacts for students' math skills. In other words, it is possible that the

lack of significant results for math are a consequence of the small sample size, and not that the program did not have an impact on students' math skills.

Despite the lack of results for math, the positive impacts for English suggest that an expansion of the program to other areas of the country would promote English learning. However, to fully answer this question, further research should determine what the cost-effectiveness of expanding the tablet program would be, and how this compares to alternative policy options.

1. Introduction[1]

Under the Kiribati Education Improvement Program (KEIP), the Kiribati Tablet Trial (KTT) used preloaded tablet computers to provide teachers and students with curriculum related, content-rich learning materials in classrooms, in an effort to improve teaching and learning. Coffey, the managing contractor of KEIP, engaged NORC to conduct an external impact evaluation of the tablet trial. The trial ran during the 2018 school year – from February to November.

In Kiribati, English is meant to be the primary language for instruction starting in Year 4 of primary school.[2] However, English is still very much a foreign language in Kiribati, although both Te-Kiribati (the local language) and English are official languages. Most of the population communicates in te-Kiribati. In many parts of the country, almost no English is used outside of the school environment. Kiribati also experiences a shortage of printed and other learning materials for students, together with the need for resources and examples of spoken English. These needs led KEIP to develop the Kiribati Tablet Trial (KTT). The tablets used in the trial were intended to supplement and strengthen teaching and learning in English in the target Kiribati classrooms.

The trial is designed to enable the Ministry of Education (MoE) and Australia's Department of Foreign Affairs and Trade (DFAT) to better understand the potential of tablets in improving teaching and learning to make informed policy decisions, including the possibility of scaling up the use of tablets for teachers and students. More specifically, the objective of the KTT was to analyze the impact of tablets on three key educational classroom needs:

1. Support for the classroom teachers who need to develop and maintain a high level of proficiency in English as the language of instruction as students progress to upper primary,

2. Support to students' subject comprehension in upper primary, and

3. Provision of increased access to learning options and materials for those Year 4 students in the trial schools who struggle with literacy.

[1] This section follows program implementation documents (KEIP, 2017a; KEIP, 2017b).
[2] The education system in Kiribati consists of six years of primary education (Year 1-6), followed by three years of junior secondary school (Form 1-3). These first nine years of education are compulsory and free. Senior secondary school consists of four additional years of schooling (Form 4-7).

Each treatment school was provided with one kio-kit, which is a package that includes 40 tablets and a docking station to house and charge the kio-kit tablets, as well as the kio-kit motherboard. The motherboard stores all pre-loaded data to the kio-kit system, and wirelessly transmits the data to each of the kio-kits. The tablet trial also provided each treatment school with a pre-loaded teacher tablet and a wireless projector that allowed teachers to show content from their teacher tablets. Finally, the tablet trial constructed solar charging stations at each of the treatment schools, so schools could charge the kio-kit tablets, the teacher tablets, and the projectors.

Teacher and student tablets were pre-loaded with resources aligned with the national Year 4 curriculum outcomes and content, based on the trial objectives.[3] Previous research on information and communications technology interventions (ICT) has found that, when content is not well-aligned with the corresponding curriculum, the impacts on student achievement tend to be negligible (Beuermann et al., 2015; Kremer et al., 2013).[4] Therefore, that tablet content was aligned to the curriculum was an important consideration for program design.

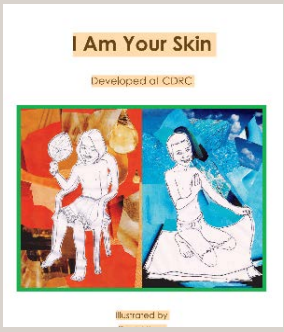
The student tablet content, designed to address classroom needs 2 and 3 above, included both (1) curated lesson materials following the curriculum of each term, and (2) resource materials to reinforce the lesson materials. The lesson materials included resources for English and math, and comprised Year 4 weekly lessons throughout the school year. The resource materials included readers, learning activities and games for English and math, supplementary reading materials, and audio and video media, including talking books, aimed at strengthening oral English pronunciation and confidence. Overall, the tablet content fits into four broad categories: reading books, talking books, English-language-specific content, and math-specific content. Box 1 provides a brief description of each of these content categories. While the content categories primarily focus on English and math subjects, some of the content addressed other Year 4 subjects. For example, the reading and talking books, which were always in English, often presented topics related to environmental sciences and healthy living.

[3] The materials also had to meet the following criteria: (1) culturally appropriate, (2) free and unencumbered with copyright, and (3) able to be downloaded on to the kio-kits (i.e. the tablet kits).
[4] A review of recent research on ICT interventions in schools is available in Annex I.

Box 1. Contents of Student tablet

Reading book

Books published by the Ministry of Education. The reading follow the curriculum of the school year and highlight words and concepts from the week's lesson plans.



English specific

These covered a wide range of resources outside of reading and talking books, including videos on English grammar concepts, vocabulary songs and games.

Talking book

Audio versions of the weekly books outlined above, read by native English speakers. The audio reads out loud while the child reads.

Math specific

These also cover a wide range of resources, including videos on math concepts, math songs and math games.

Content for each new school term was uploaded at the start of the term, with content from previous terms left available to students and teachers. The student tablets could only access the content that was preloaded on to the system, and were not able to access the internet at large. Student tablets were linked to the motherboard at a range of 20 meters, so they could only be used within that range of the docking station. In practice, this meant that students were only able to access their tablets in school. Moreover, students needed permission of the Year 4 teacher to use the tablets.

The teachers' tablets were independent from kio-kits. Teachers were able, and encouraged, to take their tablets home after the school day to continue using them, and they were able to connect to the internet for

a wider breadth of resources. The teacher tablet content included (1) all materials loaded on to the students' kio-kits, (2) additional content for English language strengthening, (3) resource materials, and (4) resources for teachers' recreational use, especially in relation to their mastery of English language. These resources included television shows and movies in English, along with English novels and other reading materials.

Program success rested on teachers incorporating tablets in day-to-day lessons. Previous research has found that teacher use of resources is an important determinant of ICT interventions' success (Barrera and Linden, 2009). Therefore, training teachers so they were able to incorporate the resources in daily activities, was a key component of the program. KEIP provided three teacher trainings over the course of the intervention. The first one took place in South Tarawa over the course of one week prior to the start of the school year. Year 4 teachers of treatment schools attended training on how to use the tablet content with students, including ways to use tablets as part of their day-to-day teaching. To monitor implementation, teachers were instructed to record tablet use daily in addition to other monitoring activities conducted directly by KEIP.[5] Principals and Year 5 teachers, who often fill in for Year 4 teachers when they are absent, also attended this training session. The second training was for Year 4 teachers only. This training session was also conducted in South Tarawa and lasted 1 week, between the first and second terms. During the second training, Year 4 teachers were invited to share their successes and challenges from the first term and worked towards improving the use of tablet resources in their classrooms. The third and final training consisted of a personalized session in which a KEIP team member observed each of the treatment classrooms, and provided feedback to teachers on tablet use in the classroom, which could be described as a coaching session more than a traditional training session.

2. Evaluation Purpose and Evaluation Questions

The main purpose of the KTT and its evaluation is to test the hypothesis that access to additional instructional resources and opportunities will improve students' achievement in English and math, and teachers' English proficiency. This information is critical for guiding the MoE in the interpretation of teaching and learning results and maximizing the effectiveness of any larger investment.

[5] In particular, KEIP staff conducted classroom observations over the school year.

The specific research questions (RQ) this evaluation addresses are:[6]

1. Does the consistent use of tablets, loaded with curriculum-linked content and used in a supportive and supported classroom, have a significant effect on student learning in the Kiribati context?
2. Does access to a tablet loaded with English language resources improve the level of English language proficiency among teachers?
3. Does the use of tablets loaded with curriculum-linked content support the learning of struggling readers?
4. How is the trial affecting classroom pedagogy, interactions and perceptions of learning at school?

A secondary purpose of the KTT was to contribute to the evidence on tablet use as an instrument to improve students' achievement, and to add to the literature on the effectiveness of using ICT in classrooms in the Pacific region.

3. Evaluation Design

3.1 Sampling Strategy

The Ministry of Education and the KEIP team selected the four target islands where the trial was implemented, namely Abaiang, Maiana, Marakei and North Tarawa. These four islands have a total population of 16,978, roughly 15 percent of the total population in Kiribati.[7] All 10 mono-grade schools in these islands were part of the trial evaluation (in mono-grade schools there is a least one class for each grade, while in multi-grade schools Year 3 and 4 curricula are taught together).[8]

Two objectives were considered when choosing these islands for the evaluation. First, schools should be located in relatively isolated places so the evaluation documents the impact of fielding the program in remote schools. There are 85 primary schools in Kiribati, most of them located in small, isolated islands with no more than three mono-grade schools. Hence, schools in these islands were chosen because they

[6] Questions 1-3 are the first three questions in KEIP (2017a). A fourth question in KEIP (2017a) reads "What are the costs and logistical issues associated with providing the underpinning hardware, pedagogical and content support to use tablets in Kiribati classrooms?" We do not include this question here as it will be addressed by the project team. In effect, the project team will "prepare a report that summarizes the considerations, issues, lessons learned and costs incurred in supporting the hardware, the teachers and the content acquisition." KEIP (2017a).

[7] Source: Own calculations using 2015 census data (National Statistics Office, 2016).

[8] While English is the primary language for instruction in Year 4, it is not in Year 3. Thus, multi-grade classes were excluded from the trial because it would be difficult to field the program in Year 3-4 multi-grade classes.

are representative of most schools in Kiribati. Second, these four islands are relatively close to South Tarawa and can be reached within a day, facilitating tablet trial support and monitoring.

The 10 mono-grade schools were randomly assigned to treatment and control groups using islands as blocks, such that there were schools in both treatment and control groups on each island.[9]

Table 1 shows the schools selected for treatment and control.

Table 1 Treatment and control schools

Island	Primary School	Status
Abaiang	Sunrise	Control
Abaiang	Waakam	Control
Abaiang	St Paul	Treatment
Maiana	Urintebura	Control
Maiana	Abitabu	Treatment
Marakei	Uaboong	Control
Marakei	Nikierere	Treatment
Marakei	Taiti	Treatment
North Tarawa	BT Uekera	Control
North Tarawa	Neintebwara	Treatment

3.2 Empirical Methods

To evaluate the impact of the program, we compare students' English and math endline results. We conducted a randomized control trial (RCT), where five schools were randomly assigned to treatment and five to control groups. Randomization ensures that, on average, characteristics of the treatment and control groups are statistically identical, with the only difference being their participation in the intervention. Therefore, any measured difference in outcomes between the groups over time can be attributed to the program. We also adjust or isolate students' baseline characteristics in the analysis, in particular achievement at baseline, to reduce the error or noise on the estimation of the program impact. A similar analysis was conducted to evaluate the impact of the program on teachers' English skills, and

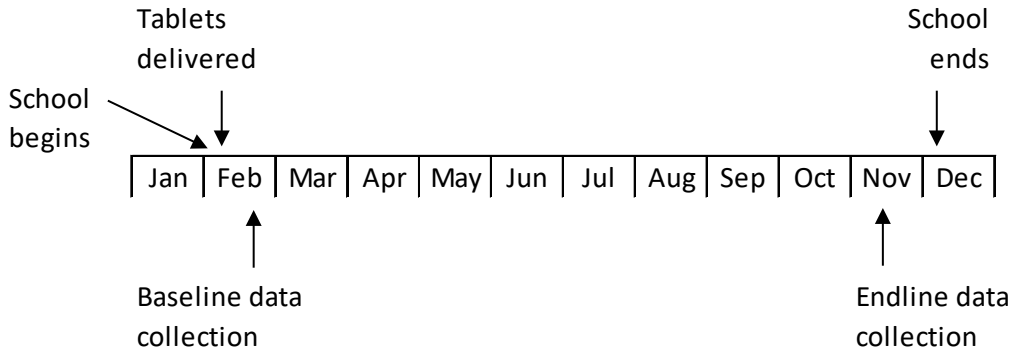
[9] When randomization was conducted (December 2017), there were 11 mono-grade schools in these four islands. Out of these 11 schools, six schools were selected for treatment. After randomization had been conducted, KEIP and MoE informed NORC that school Otin Maboran Taai was going to be turned into two multi-grade schools, which implied that it had to be dropped from the trial. As a result, the final number of trial schools was reduced to 10 (five treatment and five controls).

results for struggling readers. We also collected qualitative data to analyze different aspects of project implementation. A more detailed description of the evaluation methods can be consulted in Annex II.

3.3 Data Collection

Data was collected in February 2018 (baseline), and November 2018 (endline). Quantitative data was collected during each wave of data collection in both treatment and control schools. These included student assessments for English and math, and a teacher English assessment. Students and teachers also answered a brief sociodemographic survey. At endline, students' and teachers' surveys also included items related to KTT (e.g. views on the program, use of resources). Figure 1 shows the timeline for data collection and program implementation.

Figure 1. Program and data collection timeline - School year 2018



All Year 4 students in the 10 selected schools were surveyed, at both baseline and endline. At baseline 261 students were surveyed, out of which 202 were surveyed at endline, for an attrition rate of 22.6 percent.

Qualitative data was collected at endline only, and only in treatment schools, because qualitative data mostly pertains to the implementation, adequacy, requirements, and challenges of the tablet trial. In each treatment school, we interviewed the principal and each Year 4 teacher, and conducted classroom observations in each Year 4 classroom. Focus groups discussions (FGDs), each with 8-12 Year 4 students, were conducted in three treatment schools, and in two of these schools we conducted FGDs with Year 4 parents and one key informant interview (KII) with a community stakeholder. All qualitative data was collected in Te-Kiribati and translated to English for analysis.

Annex III reviews in detail the data collection process, including instrument content and design, training and fieldwork.

4. English and Math Baseline Results

In this section we review baseline results for the students' math and English assessments. The main purpose of this section is to discuss differences between treatment and control groups. This distinction is important because having a balanced sample – one where differences between treatment and control groups are negligible – is key to guarantee that the evaluation results reflect the causal effect of KTT, and not pre-existing differences between treatment and control groups.

We find that the differences between treatment and control groups across English and math subtasks, at baseline, are small and not significant.

Table 2 shows baseline scores for the subtasks in the student English assessment. The results reflect a relatively low performance level among students in both treatment and control groups. For example, for oral reading fluency students in the control group read an average of 13.6 words per minute (wpm), while students in the treatment group read 16.4 wpm. Moreover, students correctly answered 0.45 reading comprehension questions in the control group and 0.65 in the treatment group, on average. As a point of reference, most countries have defined reading at 60 wpm and answering at least 0.8 comprehension questions as the desired level of reading comprehension (RTI International 2009).

As explained before, in Kiribati, students start reading and writing in English in Years 3 and 4, so the observed low scores for English are to be expected, especially considering that the school year was just starting when baseline data was collected. In terms of balance, students in the treatment group had slightly higher scores than those in the control group for four of five subtasks, although the differences are not statistically significant.

Table 2. Baseline English scores

	Control (N=124) (1)	Treatment (N=137) (2)	Diff (3)=(1)-(2)
Familiar Word Fluency	7.9	10.3	-2.4
Made-up words	3.4	4.4	-1.0
Reading Fluency	13.6	16.4	-2.8
Reading Comp Correct	0.45	0.65	-0.20
Listening Comp Correct	0.95	0.93	0.02

* p<0.1 ** p<0.05 *** p<0.01
Source: Baseline student survey.

Table 3 shows results for the math assessment by treatment status. Students were able to correctly answer more items on addition than on subtraction or multiplication. For six out of the eight subtasks, the treatment students outperformed the control students, although the differences are small and not significant.

Table 3. Baseline math scores

	Control (N=124) (1)	Treatment (N=137) (2)	Diff (3)=(1)-(2)
Missing Number	5.1	5.4	-0.3
Addition 1	9.5	9.6	-0.1
Subtraction 1	6.3	6.4	-0.1
Multiplication 1	1.9	1.8	0.2
Measurement	3.2	3.4	-0.3
Addition 2	2.4	2.9	-0.5
Subtraction 2	1.6	1.6	0.0
Word Problems	2.4	2.6	-0.2

* p<.1 ** p<.05 *** p<.01
Source: Baseline student survey.

In sum, comparing English and math results between treatment and control students at baseline indicate that the differences between the two groups are small and not significant. This suggests that there are no meaningful pre-existing differences between the two experimental groups, hence any difference we find at endline can be attributed to the program.

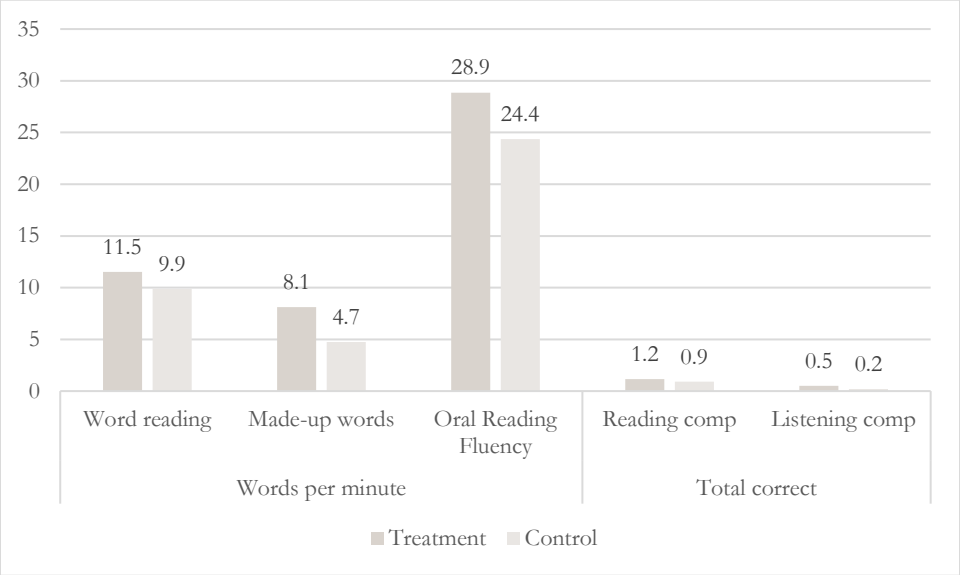
A more detailed review of students' baseline characteristics is available in Annex IV.

5. Impact Evaluation Results

RQ1: Does the consistent use of tablets, loaded with curriculum-linked content and used in a supportive and supported classroom, have a significant effect on student learning in the Kiribati context?

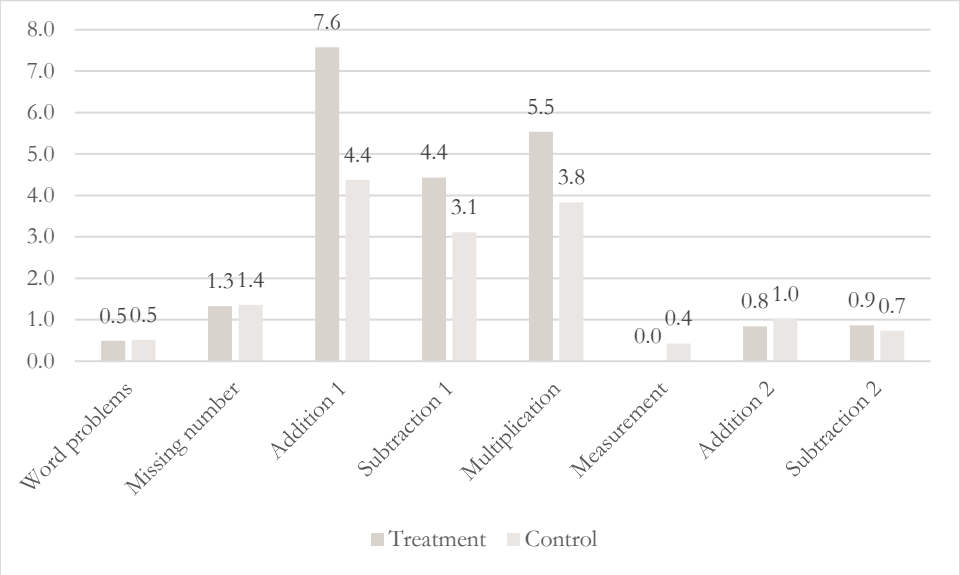
The evaluation results indicate that the program had a positive impact for English but not for math skills. To illustrate these results, Figure 2 shows changes between baseline and endline for the treatment and the comparison groups for reading skills. Across all five reading subtasks, progress was higher for the treatment than for the control group.

Figure 2. Change in reading skills between baseline and endline, by treatment status



For math, the results are much less clear, as can be seen in Figure 3**.** While progress for the treatment group was higher for addition 1, subtraction 1, multiplication, and subtraction 2, the opposite is observed for word problems, missing number, measurement, and addition 2.

Figure 3. Change in math skills between baseline and endline, by treatment status



The analysis presented in Figure 2 and Figure 3 has a few limitations worth highlighting. First, it does not establish if the differences between study groups are statistically significant or not. In other words, we would like to know if these differences are large enough so a case can be made that they not are simply the product of random variation in the students' performance, but an actual consequence of KTT. Second, we are not isolating for other characteristics that can affect the results. Third, the analysis focuses on students that were observed both at baseline and endline, but given that the attrition rate was relatively high (22.6 percent), it would be important to incorporate attrition in the analysis.

To overcome these limitations we ran regression models where the outcome of interest is each reading or math subtask score at endline, and the control variables are the corresponding baseline scores, and a set of sociodemographic characteristics.[10] We also correct for attrition by giving more weight in the analysis to observations that were observed at endline, but were similar to the observations that attrite from the sample; this way, we try to replicate the impact that would had been estimated if no attrition had happened.[11]

Figure 4 shows results for English subtasks. To denote significance we use stars, so impacts with one or more stars are considered statistically significant, and not a product of the natural variation of students' performance. The result for word reading indicates that students in treatment schools read 1.3 words per minute more than in the control group, but the coefficient is not statistically significant. For reading made-up words, there is a positive and significant impact; **children in the treatment group read 4.8 more made-up words per minute than in the control group**. We also find positive and significant impacts for oral reading fluency and reading comprehension. For oral reading fluency, we find that **children in the treatment group read 5.1 more words per minute more than in the control group.** For reading comprehension, the coefficient indicates that **children in the treatment group are able to**

[10] More formally, we estimate:

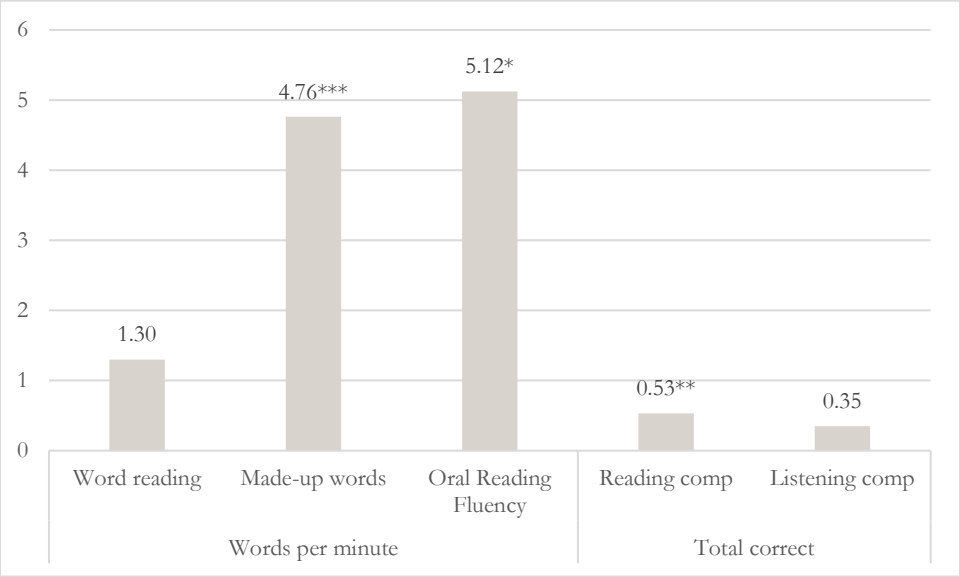
$$A_{si1} = \alpha + \beta D_s + \gamma A_{si0} + u_{si1}$$

where A_{si1} and A_{si0} are the outcomes of interest (i.e. students' test scores) for individual *i* in school *s* at endline and baseline, respectively; D_s is a dummy variable for treatment status; u_{si1} is an error term, and α, β and γ are parameters to be estimated. The coefficient of interest is β, that captures the effect of the program. Furthermore, because treatment is randomized, β can be considered an unbiased estimate of the program impact, as students in treatment and control groups should be observationally equivalent at baseline, so any difference at endline can be attributed to the program. We also include students' sociodemographic characteristics as control variables, namely children's age and gender, an asset index that is the first principal component of households' ownership of a clock, a radio, a television, a computer, a bicycle, a motorcycle, and a car, and binary variables for mother and father literacy and for whether the child had breakfast that day in order to improve the precision of the estimated program impacts.

[11] To control for attrition we use inverse probability weighting (IPW). Annex V describes this methodology and how we implemented it.

answer correctly half a question more than in the control group. Finally, for listening comprehension the coefficient is positive but not significant.

Figure 4. Impact of KTT on reading skills (N=199(a))



Notes: Baseline data included as covariates are the value of each subtask baseline, children's age and gender, an asset index that is the first principal component of households' ownership of a clock, a radio, a television, a computer, a bicycle, a motorcycle, and a car, and binary variables for parental literacy (one for the mother's and another for the father's) and for whether the child had breakfast that day. All regressions are weighted to control for attrition. Standard errors clustered at the school level in parentheses. Regressions coefficients for all the covariates included can be consulted in Annex VI.
(a)While the number of students we surveyed at both baseline and endline was 202, for the analysis we needed to drop three of these students due to item-specific missing data.
* p<0.1 ** p<0.05 *** p<0.01

How do these results compare to other interventions in the education literature? To put these results in context, it is useful to estimate the corresponding effect sizes, which are calculated as the treatment effect shown in Figure 4, over the standard deviation of the corresponding endline scores. For the three subtasks for which statistically significant impacts are found, the effects sizes are 41 for made-up words, 19 for oral reading fluency, and 40 for reading comprehension. In general, program impacts between 20 and 40 percent of standard deviation are considered large. The What Works Clearinghouse (WWC), a US based repository of high-quality impact evaluations, characterizes impacts greater than 25 percent of a standard deviation as "substantively important" (WWC, 2014 p. 23). Along these lines, the results for made-up words and reading comprehension are substantial.

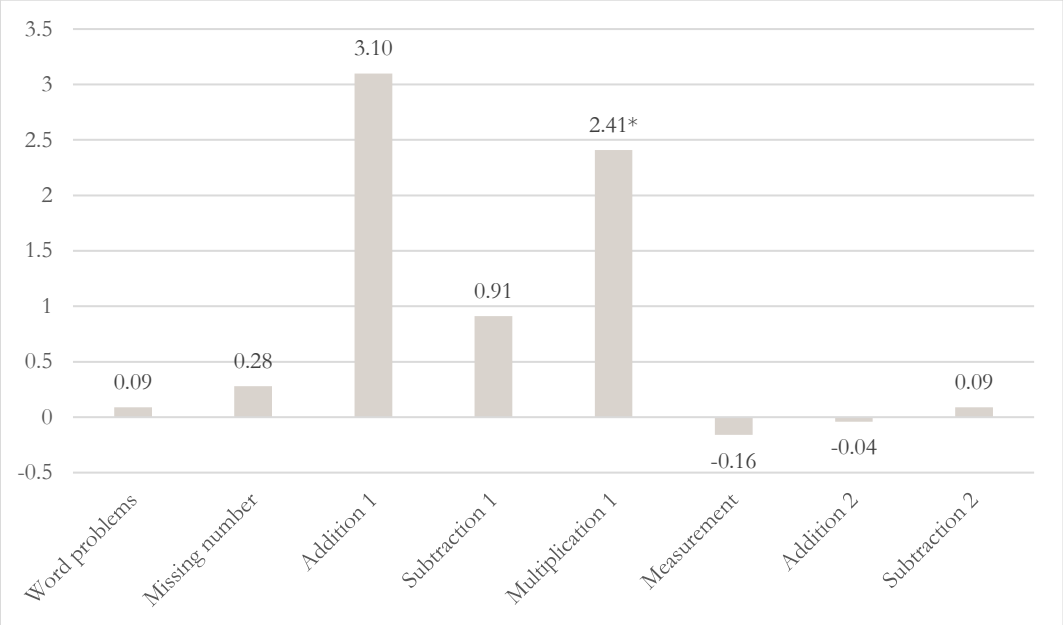
When we look at the actual changes in terms of words per minute, or questions correctly answered, the conclusions are more nuanced. For reading comprehension, we see that the students in the treatment

group answer correctly only half a question more than students in the control group, which seems small. However, regarding oral reading fluency, the 5.1 wpm impact is roughly 8-10 percent of the 45 to 60 wpm benchmark required to read with comprehension (RTI International 2009), indicating that the program is helping students reach this threshold in a meaningful way.

Results for math subtasks are shown in Figure 5. We do not find strong evidence that the program had a positive impact on math skills. While the effects are positive for 6 out of 8 evaluated subtasks, the effect is statistically significant only for multiplication.

It is possible that KTT had relatively small-medium size impacts on math skills, that the available sample is not able to detect because the effects are not large enough. This argument would be more credible if the estimated program effects were positive across all eight subtasks, but in fact the results are negative for measurement and addition 2. Therefore, while it is possible that the program has a positive (and modest) impact on math skills that we cannot estimate with precision because of the small sample, the fact that the estimated program impacts are not all positive suggest that the program does not have an impact on math skills.

Figure 5. Impact of KTT on math skills (N=199(a))



Notes: Baseline data included as covariates are the value of each subtask baseline, children's age and gender, an asset index that is the first principal component of households' ownership of a clock, a radio, a television, a computer, a bicycle, a motorcycle, and a car, and binary variables for parental literacy (one for the mother's and another for the father's) and for whether the child had breakfast that day. All regressions are weighted to control for attrition. Standard errors clustered at the school level in parentheses. Regressions coefficients for all the covariates included can be consulted in Annex VI.
(a)While the number of students we surveyed at both baseline and endline was 202, for the analysis we needed to drop three of these students due to item-specific missing data.
* p<0.1 ** p<0.05 *** p<0.01

The evaluation results indicate that the program had a positive impact for English but not for math skills. This subject-driven difference is rather puzzling. One explanation could be that the sample size was too small to identify treatment effects for math skills with enough precision. It is also possible that the tablets were filling a gap that was more critical for learning English than math. Reading resources are scarce in Kiribati, so access to reading materials is probably critical to learn English but not math. If learning math is less hampered by lack of books, this could explain why the tablets did not have a substantial impact on math skills.

RQ2: Does access to a tablet loaded with English language resources improve the level of English language proficiency among teachers?

To analyze the results for teachers' English skills, we calculated the changes for teachers in the treatment group and compare them with changes in the control group. We also interviewed teachers to record their own views on their English skills. Both the English assessment and qualitative data suggests that the program had a positive impact on teachers' English skills.

Table 4 shows the results of this analysis. In the first two columns, mean scores for control and treatment groups at baseline are displayed, and the differences between the two groups can be seen in column (3). Columns (4) to (6) show the scores and their differences at endline. Column (7) shows the double-difference. Notably, teachers in the treatment group outperformed teachers in the control group across all assessments, and both at baseline and endline, as results in columns (3) and (6) show. The results in column (7) indicate that the differences at endline are larger than the differences at baseline for three of the four assessments, namely reading, writing and speaking, as well as for the average score. This suggests that the program had a positive impact on three of the four English assessments. However, given the small sample size, we refrain from producing standard errors for this analysis, hence we cannot say if these impacts are statistically significant.

It is also interesting to note that teachers are improving their reading skills across the board. The average score for the control group went from 55.7 at baseline to 64.5 at endline, suggesting that the control group is also improving reading skills.

Table 4. Impact on teachers' English assessment

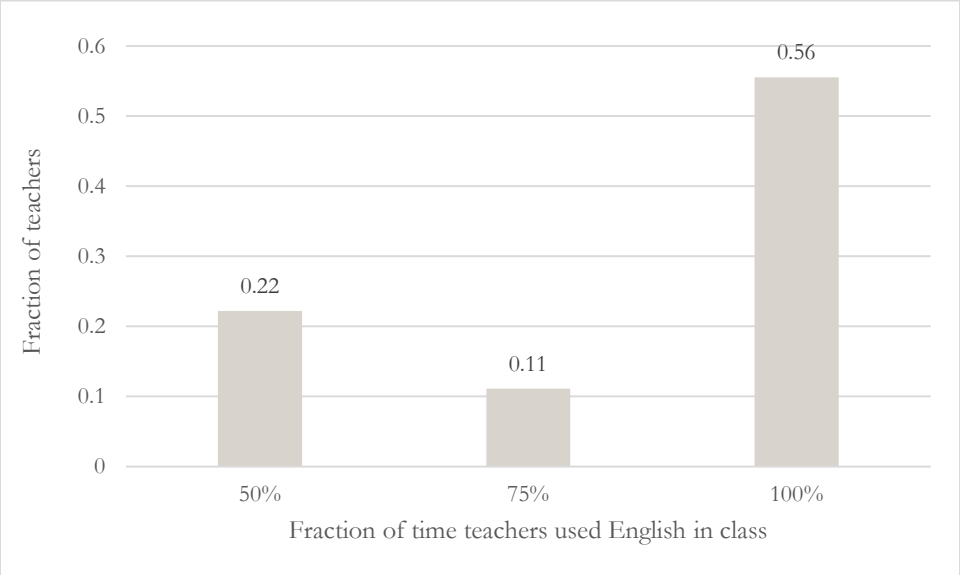
	Baseline			Endline			
	Control (1)	Treatment (2)	Diff. (3)=(2)-(1)	Control (4)	Treatment (5)	Diff. (6)=(5)-(4)	Diff-in-diff (7)=(6)-(3)
Reading	65.0	69.5	4.5	73.5	82.5	9.1	4.6
Listening	71.8	83.8	11.9	82.8	88.1	5.3	-6.6
Writing	41.3	45.1	3.8	39.0	51.3	12.3	8.5
Speaking	47.0	54.5	7.5	62.7	77.5	14.8	7.3
Average	55.7	63.0	7.4	64.5	74.9	10.4	3.0

Note: Sample sizes are 14 (6 control and 8 treatment teachers) for each wave of data collection. Given the small sample available, we do not produce standard errors for this analysis.

We also interviewed teachers to record their own views on their English skills, their perceptions of the program, and how it helped (or didn't help) improve their English skills. These data show that four out of eight teachers believe they have the English skills necessary to teach Year 4 classes, while two said they do not believe they have the English skills necessary to teach a Year 4 class. One of the teachers who felt their English was not strong enough to teach Year 4 explained that she spoke English too slowly, while the other teacher said that her students often did not understand her. Over the course of the year, all six teachers who spoke to this question felt their English had improved thanks to the tablet trial evaluation. Three of those teachers said their vocabulary had improved, with one teacher specifically mentioning a built-in dictionary as the primary source of the new vocabulary. Two teachers mentioned their improved pronunciation attributed to watching videos on the tablets, and one teacher mentioned her improved grammar.

Finally, through classroom observations we measured the extent to which teachers used English in class. Figure 6 shows that all teachers used English at least 50 percent of the time, and that five of the eight observed teachers used English all the time. Note that teachers may decide to use English more when data were collected because they were being observed, hence English use could probably be lower in a regular day.

Figure 6. Proportion of time the teacher spent instructing in English during observed lessons (%)



Source: Endline classroom observations.

In conclusion, both the English assessment and qualitative data suggest that the program had a positive impact on teachers' English skills, although the small sample prevents estimating confidence intervals for the assessment results.

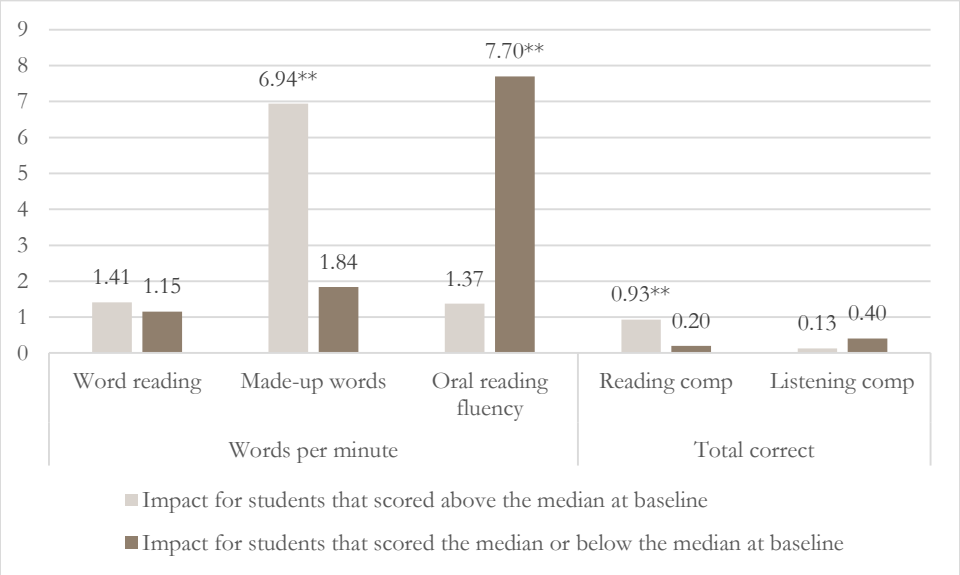
RQ3: Does the use of tablets loaded with curriculum-linked content support the learning of struggling readers?

To analyze whether there are differential impacts for struggling readers, we estimate again the effect of the program, but this time we split the sample between students that scored above and below the median in their English assessment at baseline. The results do not suggest that the program had a particularly stronger impact on struggling readers.

Figure 7 shows the results for this exercise. For word reading, the effect for both groups is quite similar: small and not significant. For made-up words, the results indicate a larger impact for students that scored above the median at baseline than for the rest of students. In effect, for students that scored above the median at baseline, there was a statistically significant impact of 6.9 wpm, while for students that scored the median or below the median, the program impact is much smaller and not significant. For oral reading fluency, the results are the opposite than for made-up words. In this case, the program impact for students that scored above the median at baseline is small and not significant, while the impact for students that scored the median or below the median at baseline is relatively large (7.7 wpm) and statistically

significant. For reading comprehension, the pattern is similar to the one observed for made-up words, that is, the program effects are larger for students that scored above the median at baseline than for the rest of students. Finally, for listening comprehension, while the coefficient for students that scored the median or below the median at baseline is larger than that for the rest of students, neither of the estimates are statistically significant.

Figure 7. Heterogeneity by baseline performance



Notes: Baseline data included as covariates are the value of each subtask baseline, children's age and gender, an asset index that is the first principal component of households' ownership of a clock, a radio, a television, a computer, a bicycle, a motorcycle, and a car, and binary variables for parental literacy (one for the mother's and another for the father's) and for whether the child had breakfast that day. All regressions are weighted to control for attrition. Standard errors clustered at the school level in parentheses.
* p<0.1 ** p<0.05 *** p<0.01

The results on heterogeneity are ambiguous in terms of whether the program benefited struggling readers more or not. While the results for oral reading fluency suggest that the program did help more struggling readers, the opposite can be concluded when we look at the results for made-up words or reading comprehension. In conclusion, the collected data does not suggest that the program had a particularly stronger impact on struggling readers, at least using the proposed definition of struggling reader. Finally, we split the sample to conduct this analysis, so it is possible that we fail to find significant difference across baseline performance because the design is underpowered.

RQ4: How is the trial affecting classroom pedagogy, interactions and perceptions of learning at school?

Regarding classroom pedagogy, we find that teachers used program resources quite intensively in class, including the projector and their own tablet, while students' use of tablets was less intensive. This finding suggests that the program helped teachers organize class content and structure, without having children using the tablets excessively. We also find that the program was well received by the community. For example, teachers, principals, parents, and community members reported that children in treatment schools were more interested in attending school; also, parents reported that children were learning skills that helped them use other devices owned by the household, suggesting potentially important spillover effects.

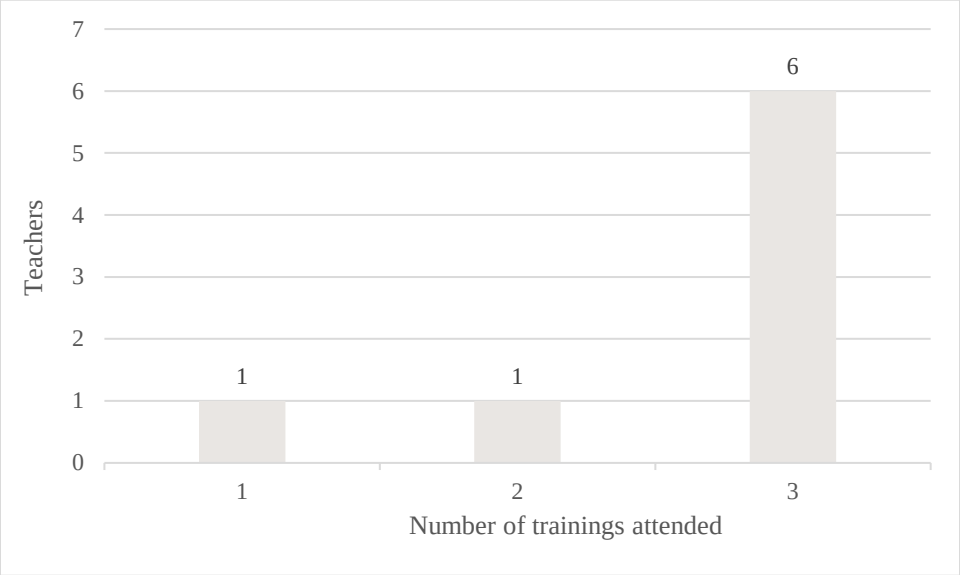
In this subsection, we discuss different aspects of program implementation including teacher training, program take-up, and how tablets were incorporated in class. We also discuss available resources in schools that could be alternatives to tablets, to describe the extent to which tablets are filling a resource gap, or competing/supplementing existing resources. Finally, we summarize challenges faced during program implementation.

Teacher Training

As explained in Section 1, three teacher trainings were conducted by KEIP with Year 4 treatment school teachers. Despite some logistical challenges during training, most teachers reported attending all training sessions, as Figure 8 shows.[12] When asked whether training sessions had been sufficient to help them learn how to use the tablets in the classroom, all eight teachers in the treatment group answered yes.

[12] Throughout the trainings there were logistical limitations that led to an uneven amount of training for the treatment teachers. The first limitation is that only the Year 4 teachers were invited to the second week-long training in South Tarawa. After the second training, one of the Year 4 treatment teachers left her school and a Year 5 teacher took her place. This Year 5 teacher attended the first training, but did not attend the second one. Another limitation is that during the second training in South Tarawa, many of the treatment teachers were invited to attend a different educational training at the same time as the tablet trial training. Many teachers split their time between the two trainings shortening their overall training time for the tablet trial program. Due to these limitations, not all treatment teachers received the same amount of training as part of the tablet trial.

Figure 8. Teacher attendance to tablet training (number of teachers)



Source: Endline teacher survey.

Satisfaction with the training sessions was confirmed during qualitative data collection. Each of the eight teachers interviewed agreed that the trainings were effective to learn how to use the tablets in their classroom. All teachers agreed in general terms that the trainings expanded their knowledge on tablet use in the classroom, while some provided specific examples. One teacher mentioned learning how to help students learn via the apps, saying "Yes [the trainings] are really helpful, by the use of the tablet in our teaching as well as in students' learning by using the installed apps." Another three teachers specifically said that learning how to use Power Point was the most beneficial part of the training. In addition, each of the five interviewed principals believe that the teachers were sufficiently trained to use the kio-kit tablets in the classroom, with one principal saying, "Yes, the training/workshop is really good for them as it helps them understand how to operate these new teaching aids."

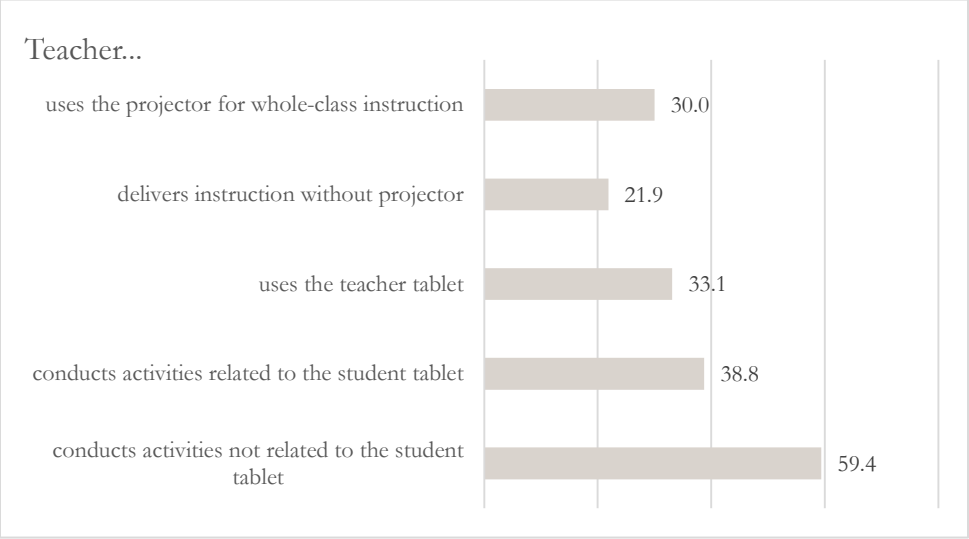
In conclusion, the data indicate that teachers' training attendance was quite high, and that teachers were satisfied with the skills they learned.

Tablet Take-Up and Classroom Pedagogy

We collected data on tablet use through a classroom observation rubric (COR), and student and teacher surveys. Classroom observations reveal a relatively high use of program resources. Note that these observations were conducted at endline, so teachers had had the whole year to get used to incorporating program resources in class.

During classroom observations, enumerators recorded pre-coded classroom activities in 6-minutes intervals over 60-minute class periods. Every six minutes, enumerators recorded activities conducted by teachers, as well as activities conducted by students. Figure 9 shows the results for teacher activities. On average, teachers spent 30 percent of time using the projector for whole-class instruction, 21.9 percent of the time delivering instruction without using the projector, 33.1 percent of the time using the teacher tablet, 38.8 percent of the time conducting activities related to the students' tablets, and 59.4 percent of the time conducting activities not related to the students' tablets.

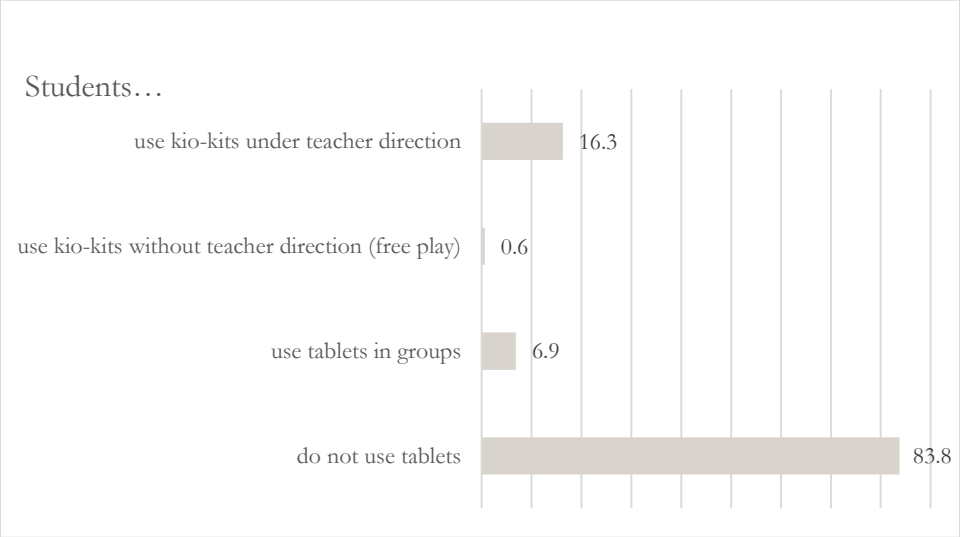
Figure 9. Teacher activities during class (%)



Note: No disciplining instances were reported, which is why the category is not shown. More than one activity can be selected/taking place at the same time
Source: Endline classroom observations.

Figure 10 shows students' activities during classroom observations. According to these data, students used their tablets under teacher direction 16.3 percent of the time, students used the tablets without teacher direction 0.6 percent of the time, students used the tablets in groups 6.9 percent of the time, and students did not use the tablets 83.8 percent of the time.

Figure 10. Students' activities during class (%)

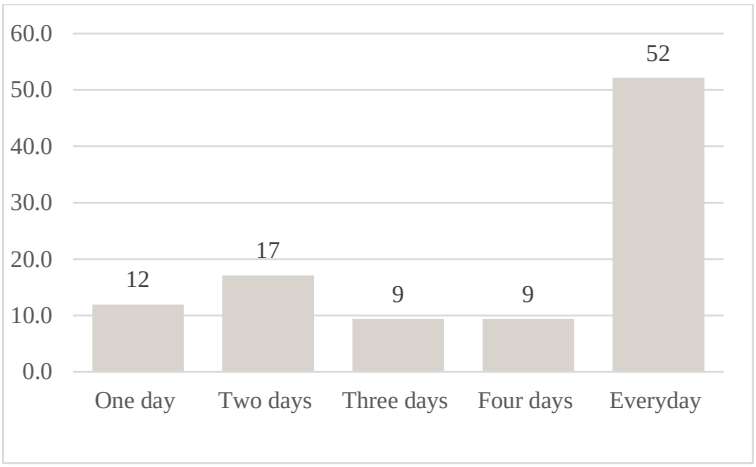


Source: Endline classroom observations.

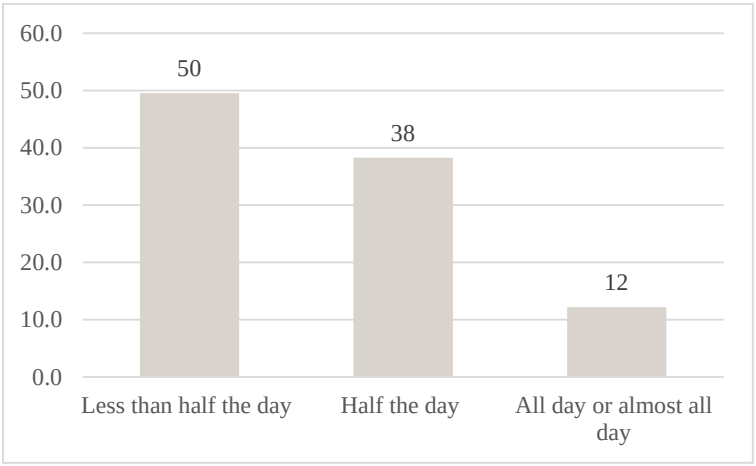
Students' and teachers' surveys also included items to record tablet use. As Figure 11 shows, students in the treatment schools reported using the tablets regularly. Panel A shows that over half the students said they used the tablet every day during the previous week. Panel B shows how much students used the tablets the last day they used them; 50 percent of students used the tablets less than half the day, 38 percent used them half the day and 12 percent used them all day or almost all day.

Figure 11. Use of tablets as reported by students (% of treatment group)

A. Number of days tablets were used the previous week (N=117)



B. How long tablets were used the last day (N=115)

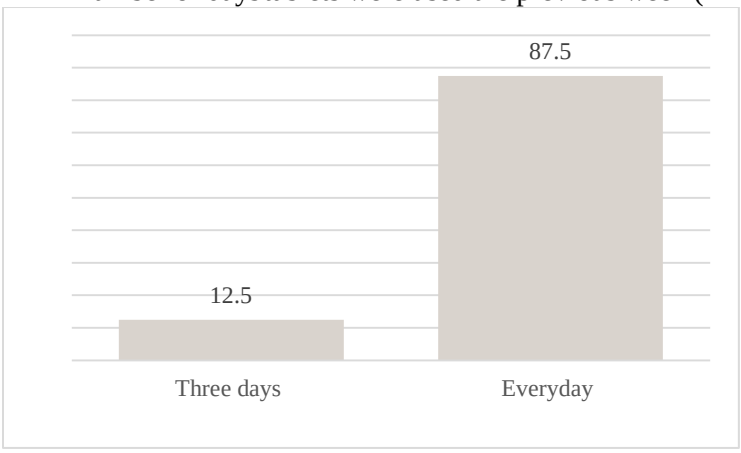


Source: Endline student survey.

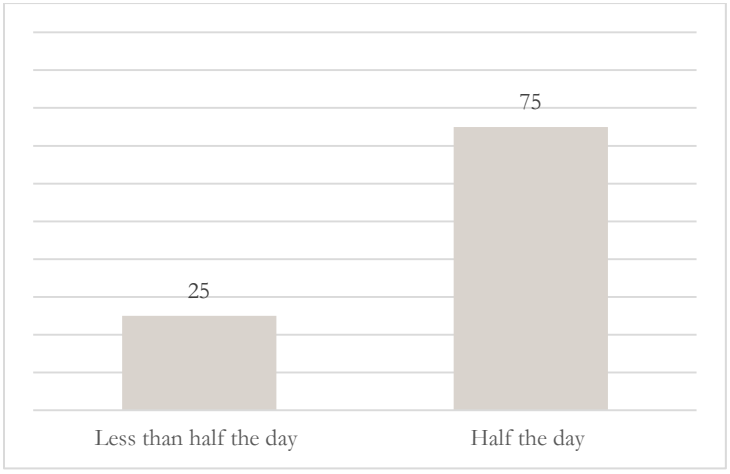
Teachers also reported high tablet take-up figures. As Figure 12 shows, the majority of teachers in the treatment group used the tablet every day the week of class prior to data collection; and three fourths of them used it for half a day the last day they used it.

Figure 12. Use of tablets as reported by teachers (% of treatment group)

A. Number of days tablets were used the previous week (N=8)



B. How long tablets were used the last day (N=8)



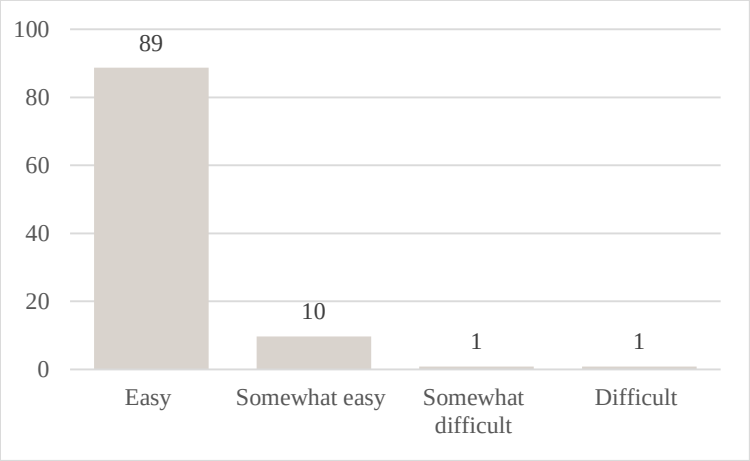
Note: Corresponds to the use of tablets in general in the classroom, without discriminating between teacher of student tablets.
Source: Endline teacher survey.

In sum, teachers used program resources quite intensively in class, including the projector and their own tablet. At the same time, students' use of tablets was less intensive, especially when we look at the classroom observations results. These results suggest that the program helped teachers organize class content and structure without having children use the tablets excessively. Notably, tablet usage data from both students' and teachers' surveys (shown in Figure 11 and Figure 12, respectively) suggest more intensive use of tablets than what the COR found. It is possible that students' and teachers' reports are affected by recall error, or tend to overstate tablet usage due to social desirability, which is why we consider the COR results a more reliable measure of tablet use in class, than students' or teachers' reports;

despite the fact that, also due social desirability, it is possible that teachers used the tables more the day of the observation than they did regularly.

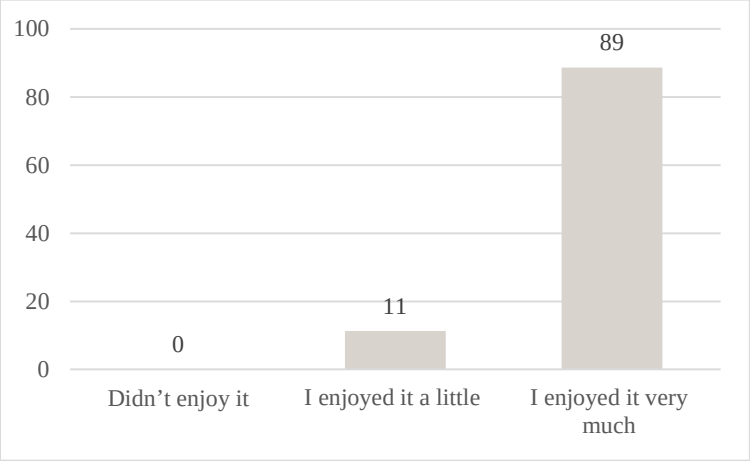
Regarding how students and teachers interacted with program resources, survey data indicate that both students and teachers found the resources useful and easy to use. In effect, as Figure 13 shows, 89 percent of students found it easy to use the tablets; and as Figure 14 shows, 89 percent of students enjoyed using the tablets very much. Similarly, all eight surveyed teachers said the teacher tablets were easy to use.

Figure 13. How easy was it to use the tablets (% of treatment group) (N=124)



Source: Endline student survey.

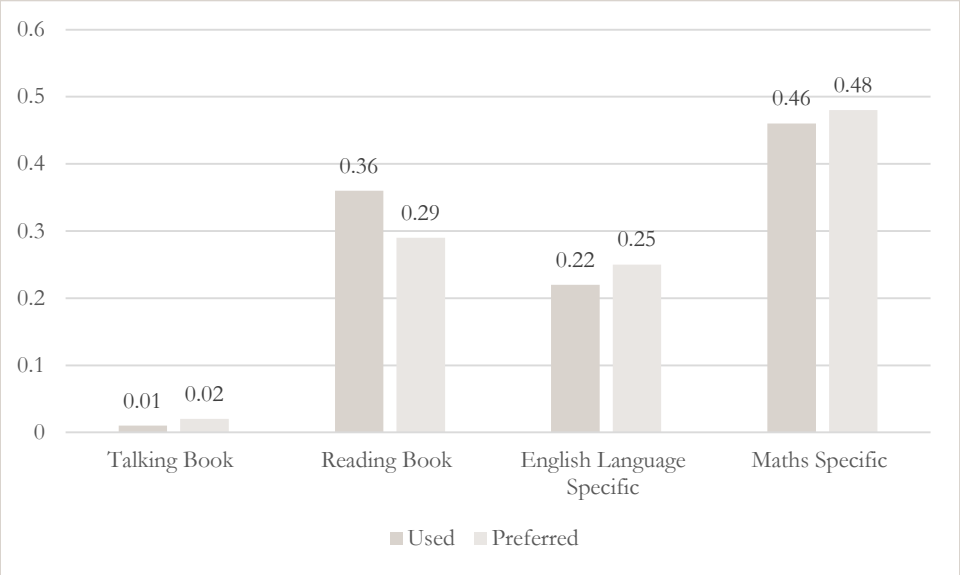
Figure 14. How much students enjoy using the tablet (% of treatment group) (N=124)



Source: Endline student survey.

We also asked students which types of applications they used the most, and which applications they preferred. Figure 15 shows the responses to these questions. Math applications were most used by students, followed by reading books and English-specific applications, while talking books were rarely used. Similar responses were recorded for students' application preferences. Note that while math applications were more used than other types of applications, this is not true if we compare math applications to the aggregate of English/reading applications (Talking books + reading books + English-specific language).

Figure 15. Applications (% of treatment group) (N=130)



Note: For both questions (which applications students used the most and which they preferred) multiple responses were possible, which is why fractions add to a little more than one.
Source: Endline student survey.

During FGDs, students reported that their favorite uses of the tablets were reading, songs, math games, and movies. The reasons for liking using the tablets were a mix between entertainment and education. One student said "It is fun and gives us knowledge and skills like reading in English and understanding English words." When students were asked how the tablets helped them improve their English and math skills, students specifically mentioned that the tablets helped them learn how to speak English. One student said, "[the tablet] enriches our vocabulary and helps us on how to pronounce words." For their math skills, multiple students mentioned that they learned math operations through games preloaded onto the tablets.

Teachers also reported general satisfaction with the tablets. During KIIs, teachers reported using the teacher and kio-kit tablets for all Year 4 subjects, including English, math, environmental science, healthy

living, and in some cases, Te-Kiribati. The eight teachers interviewed believe that the content loaded on to the tablets aligns with the Year 4 curriculum. Teacher satisfaction with tablets is multifaceted and is best represented by one teacher who said, “This implementation has been very helpful to students in learning English speaking and numeracy and also to us teachers. It also helps us manage or discipline our students well. Just like when their in-class activities are finished, they will have time to use the tablets and that really makes them attentive, punctual, and participatory”.

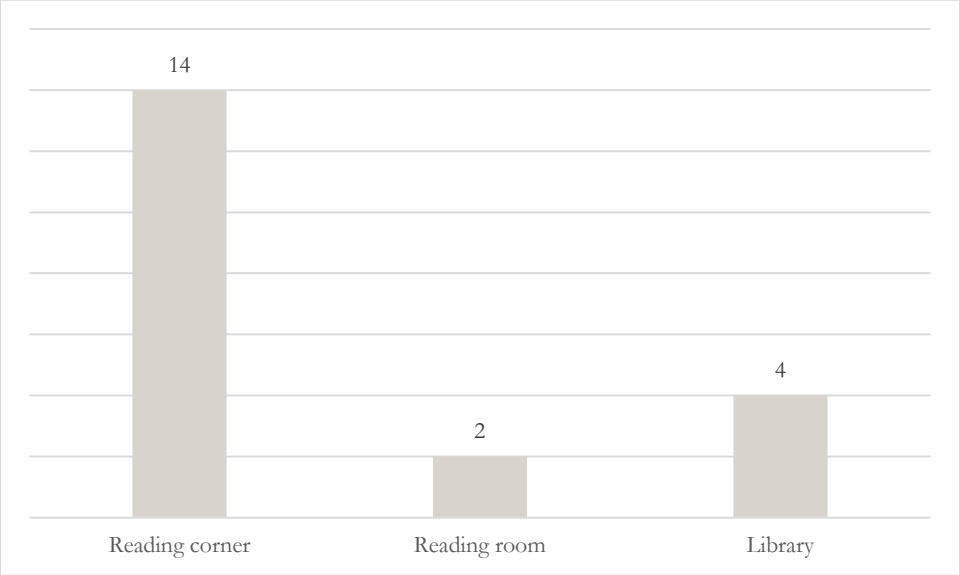
When asked about the tablets as a teaching resource, six out of the eight teachers said that they prefer using the tablets as a teaching resource over the resources used in the previous school year, which included big books, posters, and pictures. The teachers primarily mentioned that the tablets were easier to use than other materials, especially considering all the resources were housed in one place. Some teachers also mentioned that the tablets were better resources for student learning, and one teacher mentioned that they are good motivational tools that keep students engaged.

In sum, both quantitative and qualitative data indicate that teachers and students found it easy and useful to use the tablets.

Alternative Resources

In order to document the extent to which tablets were filling a resource gap versus supplementing existing materials, the endline survey included items on availability of books at school. According to the teacher survey, all Year 4 classrooms have a reading corner where children can read and borrow books. In addition, two teachers said there was a reading room in their school, and four said there was a library, as can be seen in Figure 16.

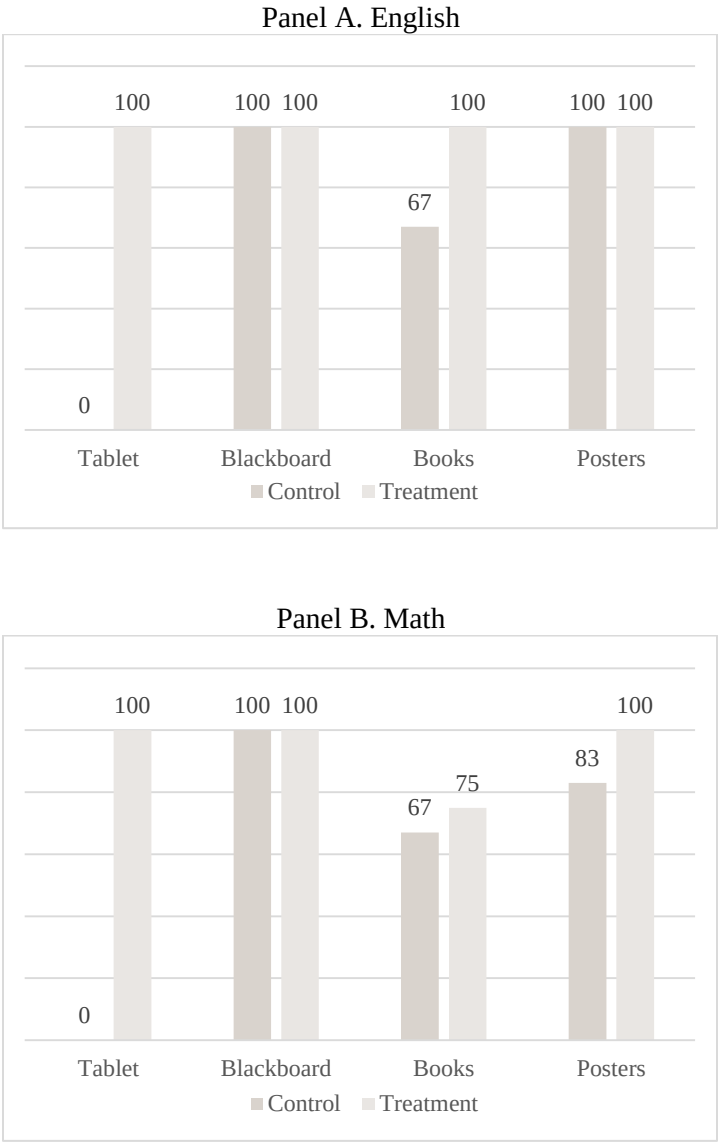
Figure 16. Access to reading resources as reported by teachers (N=14)



Source: Endline teacher survey.

Figure 17 shows the resources teachers most used in class. Panel A shows results for English class. All eight surveyed teachers in the treatment schools reported using tablets, while no teacher in the control group did. All teachers in the treatment group also reported using blackboards, books, and posters. All teachers in the control group reported using blackboards and posters, and 67 percent reported using books. For math (Panel B) all teachers in the treatment group reported using tablets, blackboards, and posters, and 75 percent reported using books; while all teachers in the control group reported using blackboards, 67 percent reported using books, and 83 percent reported using posters.

Figure 17. Resources used in class as reported by teachers (%)



Source: Endline teacher survey.

Although it is clear that there are multiple resources available for teachers, student-level resources are much scarcer. The MoE does not provide text books to students, hence there are no text books for these subjects in the country. Moreover, all teachers in the sample reported that no student has math or English textbooks. This finding shows that tablets were filling a need for materials to study English and math, rather than competing with or supplementing other available resources.

Perceptions of Learning at School

As part of the qualitative fieldwork, the evaluation team held two FGDs with parents of Year 4 students in treatment schools and interviewed two School Improvement Committee representatives. Among the parents, there was consensus that the use of tablets in the classroom was a good thing. Many parents reported improvements in their children's English skills, with one parent telling a story of their child returning from school and asking for the definitions of words they learned from the kio-kit tablets. Another interesting change is that several parents reported their children were using their knowledge of the kio-kits to use mobile devices owned by the household.

The two School Improvement Committee representatives were aware of the tablet trial, but had limited knowledge of the details of the intervention. One representative said, "I have a bit of understanding about the tablet, as what I have seen as a community leader, I noticed that it is really good for use in schools." Both representatives believe that the schools were well equipped with all the resources required for the tablet program to be effective, and believe that the tablet trial is a positive thing.

Community and parent enthusiasm was corroborated through the Principal interviews. All five of the interviewed principals mentioned that the community was proud to be participating in the tablet trial.

One of the most referenced changes in perception of learning at school is that the Year 4 students were eager to go to class because of the kio-kit tablets. Teachers, principals, parents, and community members mentioned that the students looked forward to school because they were able to use the kio-kit tablets. One parent mentioned that their child did not want to miss school, even when sick, in order to use the kio-kit tablet. When discussing student discipline, one teacher said "If a student is annoyed in class, they tell him not to do tablet anymore, and at once that child changes very easily." Where there may have been concerns that the tablets would be a distraction and create disciplinary issues in the classroom, it seems that they became a tool to promote class attendance and participation.

Implementation Challenges

Each of the qualitative interview participants were asked about challenges facing the tablet trial implementation, and how they think those challenges should be addressed. The challenges mentioned can be separated into two categories – usability challenges and technical issues. In terms of usability, some teachers and principals mentioned that the tablets were difficult to use, for both teachers and students, when the tablet trial began. However, teachers also reported that these difficulties were resolved as teachers and students used the tablets and became more comfortable with them. This change is reflected in the endline quantitative data, where most students said that tablets were easy to use (Figure 13).

Moreover, during classroom observations, enumerators found that neither teachers nor students appeared to find it difficult to operate the project equipment.

Technical issues seemed to be more common and persistent throughout the tablet trial. One principal mentioned that the tablets did not charge properly, while a teacher mentioned that the solar charging station was contingent on the weather. If the weather was cloudy, the tablets would not charge fully. Another challenge mentioned is the connectivity between the kio-kits themselves and the housing unit that contains the motherboard. One principal acknowledged that the kio-kits perform better when the housing unit is placed in the classroom, but also mentioned that the unit is very heavy and difficult to move from the main office to the classroom every day. Aside from these issues, it is important to highlight that during classroom observations, no damaged equipment was found (this includes the projector and the teacher and student tablets).

6. Final Comments

6.1 Limitations

There are two key limitations to this study, both of which are associated with the sample of schools that was selected for the evaluation. First, the 10 evaluation schools were purposefully selected. In effect, one of the main reasons why these schools were selected was their easy accessibility. This implies that the evaluation results cannot be extrapolated to other schools. Limited external validity is a pervasive feature of the evaluation literature, in particular in the RCT literature, as this type of evaluation is often limited to a specific set of selected individuals or organizations. That being said, schools were selected so they were relatively similar to other schools in Kiribati, at least in terms of demographics, school environment and staffing. Along these lines, the evaluation results could inform the potential program effects on other similar islands in the country.

Second, given the small sample size, the program needed to have a large impact for the design to measure it with conventional precision. This could explain why we did not find impacts for students' math skills. In other words, it is possible that the lack of significant results for math are a consequence of the small sample size, and not that the program did not have an impact on students' math skills. This limitation is even more apparent when we consider the results for teachers' English skills, as the sample size was only 14. This is why the quantitative results for teachers were supplemented with qualitative data.

Trial duration is another factor that must be considered. The program could have larger impacts if it was implemented for more than a year, as teachers could improve how they integrate the program resources in

class, and students could have access to more resources over their primary education. However, it is also possible that some of the results we documented would vanish as excitement about the tablets diminishes over time.

In terms of program implementation, one observation worth highlighting is that when baseline data was collected, the student tablets had already been delivered in schools, which could threaten the validity of the baseline; however, students had been exposed to tablets only for a few days, so we do not believe this compromised the achievement results recorded at baseline in a meaningful way.

Student attrition was relatively high. This has implications for power, and potential impacts on external validity and program impact bias. We used inverse probability weighting to control for attrition based on observable characteristics.

Social desirability bias or other related biases could affect some of the results documented in this report. For example, respondents (students, teachers, principals, parents, or other stakeholders) could overstate the benefits of the program to make the case that the program should be expanded. The results from classroom observations could also be affected by these biases if teachers planned their lessons in a special way so when they were observed tablets were used more. Collecting academic achievement data was important, as it acted as a measure of program success that should be purged from any of these types of biases.

Finally, albeit teachers reported using tablets not only to teach math and English, but also science and healthy living, the achievement data collected for the evaluation focuses on English and math only. Impacts on other subjects are not considered in the analysis. Likewise, the available data do not allow us to document the impact on IT literacy.

6.2 Conclusions and Recommendations

In this report, we present the results of the impact evaluation of the Kiribati Tablet Trial. Given the proposed research questions, these are main conclusions:

RQ1: Does the consistent use of tablets, loaded with curriculum-linked content and used in a supportive and supported classroom, have a significant effect on student learning in the Kiribati context?

- The program had positive impacts on students' English reading skills. Out of the five subtasks evaluated, we found positive and significant effects for three of them, namely reading made-up words, oral reading fluency and reading comprehension. For made-up words, we found that

students in the treatment group were able to read 4.8 wpm more than in the control group, while the impact for oral reading fluency was 5.1 wpm. For reading comprehension, we found that students in the treatment group answered correctly 0.53 more questions than their counterparts in the control group.

- We did not find evidence that the program affected students' math skills. Out of the eight evaluated subtasks, we found positive impacts for six of them, but a statistically significant effect for only the multiplication problems subtask.

RQ2: Does access to a tablet loaded with English language resources improve the level of English language proficiency among teachers?

- The small sample size for the analysis on teachers' English skills prevented a more robust study of this aspect of the program. However, the estimated results suggest a positive impact for three of the four conducted English assessments. In addition, teachers did report that the tablet helped them improve their English skills.

RQ3: Does the use of tablets loaded with curriculum-linked content support the learning of struggling readers?

- We did not find strong evidence that the program had a differential effect by achievement level at baseline. When we split the sample by performance at baseline, we found that for oral reading fluency the program benefited more students that scored the median or below the median at baseline than students that scored above the median, but the opposite pattern was observed when we looked at the results for made-up words and reading comprehension. Moreover, we did not find differential effects across baseline performance for any of the other three English subtasks.

RQ4: How is the trial affecting classroom pedagogy, interactions and perceptions of learning at school?

- Tablets and project resources were used in class. This finding was confirmed through classroom observations and surveys with both students and teachers. The results suggest that teachers used tablets and projectors to organize class and content.

- Stakeholders had positive views on the tablet program. During FGDs, parents reported improvements in their children's English skills, and an increase in children's interest in going to school. Teachers also reported being satisfied with the tablet program, and six out of the eight teachers said they prefer using the tablets as a teaching resource over the resources used in the previous school year. Students also reported enjoying using the tablets, and said that the tablets

helped them understand words in English, among other benefits. Other stakeholders less involved in the day-to-day program, namely principals and community members, also had positive views on the program.

- The two main implementation challenges highlighted by stakeholders were that sometimes charging the tablets was challenging (because of the weather), and that connectivity between the docking station (where the motherboard is located) and tablets was sometimes compromised by the fact that the docking station would be housed outside the classroom.

The present study contributes to the literature on computers and learning. More specifically, this study documents the impact of using technology in schools in a context where no other learning resources are available, other than classroom-level resources like posters, blackboards and teachers' books.

KTT was successful in addressing two key conditions for tablet programs (and ICT interventions in general) to work. First, tablet content was curated so there was a correspondence with the curriculum. This success is reflected in that all treatment teachers reported that the content loaded on the tablets was aligned with the Year 4 curriculum. Previous interventions involving the use of computers in schools have found that when content is not aligned with the curriculum, these interventions tend to have negligible impacts. Any expansion of the KTT to other grades should incorporate this consideration into the design.

Second, teachers incorporated student tablets and other project resources in class activities. The factors that made this integration possible are surely multifaceted, but teacher training most likely played an important role. Teachers' training attendance was quite high, and teachers reported that training was effective in teaching them how to use the tablets in class. Again, any replication or expansion of KTT in Kiribati should consider the importance of adequate teacher training.

More generally, it is important to highlight that KTT had staff dedicated to curating the tablet content, training teachers, and overseeing every aspect of the program; replicating this level of dedication at scale or in other contexts could be challenging.

Future replications of the tablet program should also consider the differential effects we found for English and math. While we cannot discard that a larger sample would allow estimation of significant effects for math, the fact that we do not find at least treatment effects with positive sign across all math subtasks suggest that the program did not have an impact on these skills. These results indicate that future replications of the program should consider either rethinking of the math component of the intervention,

or focusing on English rather than math. Along these lines, it would also be worth it to explore impacts on other subjects (e.g. natural sciences, social sciences).

An unintended spillover effect of the program seems to have been an improvement in how students use mobile devices owned by the households. Given the importance of IT literacy, this outcome should probably be considered in future replications of the program.

Despite the lack of results for math, the positive impacts for English suggest that an expansion of the program to other areas of the country would promote English learning. However, to fully answer this question, further research should determine what the cost-effectiveness of expanding the tablet program would be, and how this compares to alternative policy options.

References

Bando, Rosangela, Francisco Gallego, Paul Gertler, Dario Romero Fonseca (2017). Books or laptops? The effect of shifting from printed to digital delivery of educational content on learning. Economics of Education Review 61 (2017) 162–173.

Berlinski, S., & Busso, M. (2017). Challenges in educational reform: An experiment on active learning in mathematics. Economics Letters, 156, 172-175.

Beuermann, Diether W., Julian Cristia, Santiago Cueto, Ofer Malamud, and Yyannu Cruz-Aguayo (2015). One Laptop per Child at Home: Short-Term Impacts from a Randomized Experiment in Peru. American Economic Journal: Applied Economics 2015, 7(2): 53–80.

Bulman, George, and Robert W. Fairlie (2016). Technology and Education: Computers, Software, and the Internet. Handbook of the Economics of Education 5: 239-280.

Council of Europe (2018). Common European Framework of Reference for Languages (CEFR), Global scale - Table 1 (CEFR 3.3): Common Reference levels. Retrieved from: https://www.coe.int/en/web/common-european-framework-reference-languages/table-1-cefr-3.3-common-reference-levels-global-scale

National Statistics Office (2016). 2015 Population and housing census, Volume 1: Management Report and Basic Tables. Bairiki, Tarawa.

Habyarimana, James P. and Shwetlena Sabarwal (2018). Re-Kindling Learning: E-readers in Lagos. World Bank Policy Research Working Paper No. 8665.

KEIP (2017a). Kiribati Education Improvement Program Phase II - The Kiribati Tablet Trial Implementation Plan.

KEIP (2017b). State of Play - Trial on the use of tablets in 6 Kiribati classrooms.

Kremer, M., Brannen, C., & Glennerster, R. (2013). The challenge of education and learning in the developing world. Science, 340 (6130), 297–300 2013 Apr 19. doi: 10.1126/science.1235350.

Malamud, O. , & Pop-Eleches, C. (2011). Home computer use and the development of human capital. Quarterly Journal of Economics, 126 , 987–1027 .

McEwan, P. (2015). Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments. Review of Educational Research 85(3)353–394.

Muralidharan, Karthik, Abhijeet Singh, and Alejandro J. Ganimian (Forthcoming). Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India. American Economic Review.

NORC at the University of Chicago (2018a). Impact Evaluation of READ CO Ethiopia, Endline Evaluation Report.

NORC at the University of Chicago (2018b). USAID Impact and Performance Evaluation of the Waache Wasome Let Them Learn Project (Tanzania).

Piper, Benjamin, Stephanie Simmons Zuilkowski, Dunston Kwayumba and Carmen Strigel (2016). Does technology improve reading outcomes? Comparing the effectiveness and cost-effectiveness of ICT interventions for early grade reading in Kenya. International Journal of Educational Development 49 (2016) 204–214.

RTI International (2009). Early Grade Reading Assessment Toolkit. World Bank Office of Human Development.

RTI International (2014). Early Grade Math Assessment Toolkit. World Bank Office of Human Development.

SolarSPELL (2018). SolarSPELL 2017 Annual Impact Report.

Telecommunications and Radiocommunications Regulator of Vanuatu (2018). Universal access policy (UAP). Stakeholders update report no. 9. Update on UAP implementation.

What Works Clearinghouse (2014). WWC procedures and standards handbook (Version 3.0). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, What Works Clearinghouse.

World Bank. 2015. Conducting classroom observations: analyzing classrooms dynamics and instructional time - using the Stallings 'classroom snapshot' observation system: user guide (English). Washington, D.C. : World Bank Group.

Annex I. Previous Research on ICT Interventions in Schools

A large body of literature documents the use of computers in education, with varying results in terms of impacts on academic achievement. McEwan (2015) reviewed the experimental evidence of using computers or technology in schools and found a significant average impact on academic achievement of 0.15 standard deviations; however, Bulman and Fairlie (2016) reviewed studies analyzing ICT interventions and computer-aided instruction, and found mixed results. Some studies have even found negative impacts of the use of computers on academic achievement (Malamud & Pop-Eleches, 2011).

The variation of impacts associated with the use of technology is hardly surprising given that interventions can vary greatly across key dimensions. For example, how linked computer content is to the school curriculum, and how computers are integrated into a classroom's day-to-day activities, are major determinants of their effectiveness in increasing student achievement (Kremer et al., 2013). In effect, interventions where content is not well coordinated with school curriculum tend to have no impacts on academic achievement. Beuermann et al. (2015) evaluated the One Laptop Per Child Program in Peru, where computers were preloaded with standard applications not directly related to the curriculum (e.g. word processor, games, and programming environments); they found no impact on student academic achievement. Similarly, Barrera and Linden (2009) evaluated a school-based computer program in Colombia, *Computadores para Educar*, and also failed to find positive impacts on students' academic results. In that case, data collected by the study suggest that the program was not successful because teachers did not incorporate computer use in class. In contrast, Banerjee et al. (2007) evaluated a computer-assisted learning program in India where students solve math problems that matched their ability level, and found large and significant impacts on math test scores.

More recent evaluations have also found mixed results on the effects of computers on education, even when computer content is well-aligned with school curriculum. Berlinski & Busso (2017) analyzed the impact of a multi-arm intervention implemented in Costa Rica aimed at improving student learning in mathematics. Schools were divided into four treatment groups and one control group. In one of the treatment groups, teachers were trained in 'active learning' techniques, which involved teaching approaches that encourage student interactive participation, group work, and practical exercises. The other treatment groups received the same active learning training plus interactive whiteboards, active learning training plus a computer lab, or active learning training plus one computer per student. Notably, students in the control group outperformed students in all treatment groups. Moreover, the computer per student treatment arm had the worst performance.

Piper et al. (2016) analyzed an experiment in Kenya, where three different ICT resources were evaluated through a multi-treatment design. Schools were divided into four groups; group 1 received e-readers for students, group 2 received tablets for teachers, and group 3 received tablets for instructional supervisors from the Ministry of Education. These three groups also received an intervention package previously fielded in Kenya called the Primary Math and Reading Initiative (PRIMR), which included teacher and head teacher training in math and reading, books for every child, and ongoing instructional support for teachers. A fourth group received no intervention. The authors found positive impacts for all three treatment groups compared to the control group, but due to study design, it is hard to disentangle whether the driver of these impacts were the ICT components or PRIMR. When the authors compared the results of the evaluation with those obtained for PRIMR's original implementation, they did not find evidence that any of the ICT components added much value to student achievement.

Bando et al. (2017) analyzed the impact of substituting books with laptops in a sample of schools in Honduras. They found no major differences in academic (math and Spanish) and non-academic outcomes between the two groups. The authors argue that, while initially it is more costly to provide laptops than books, the marginal cost of providing more content through laptops is cheaper, hence laptops could be more cost-efficient if enough books were replaced with them.

Muralidharan et al. (forthcoming) analyze an intervention in India that provides vouchers for students to attend an after-school program, where they had access to computer-aided instruction tailored to their ability level. The study finds that the program had positive and significant impacts on academic achievement.

Habyarimana and Sabarwal (2018) studied a multi-treatment intervention in Nigeria, where students were given e-readers with different types of content. In treatment arm 1, e-readers were preloaded with non-curriculum content; in arm 2, e-readers were preloaded with non-curriculum content and curriculum textbooks; and in arm 3, they were loaded with all the content in arm 2 plus supplementary curriculum-related material. A fourth treatment arm was used as a control. Researchers did not find significant results on academic achievement for any treatment group; however, they did find positive and significant effects for reading in treatment arm 2 among students who did not have textbooks at baseline, suggesting that the intervention (arm 2) was successful for students with no other access to curriculum materials.

In sum, the empirical literature on the use of computers and other technology in schools shows mixed results. One conclusion that can be derived is that alignment with the curriculum is a requisite for a program's success, as well as incorporating ICT resources in day-to-day class. The present study

contributes to the literature on computers and learning where computer content is curated according to the school curriculum, and measures were taken so teachers would incorporate project resources in their class activities, mostly in the form of adequate teacher training, including a coaching session. Moreover, within this literature, this study documents the impact of using technology in a context where relatively few alternative learning resources are available in school, other than classroom level resources like posters, blackboards and teachers' books. Other evaluations compare computers with books (Bando et al., 2017), or are fielded in environments where books and computers are supplements (Piper et al., 2016; Habyarimana and Sabarwal, 2018). The evaluation of KTT considers the case of using tablets in a context where instructional materials at the student level are scarce or nonexistent. The second contribution of this study is that it documents the impact of using ICT in classrooms in the Kiribati context; albeit there have being other education-related ICT interventions in the Pacific region (e.g. solar-powered wi-fi hotspots to disseminate educational content in Vanuatu, Samoa and the Federated States of Micronesia (SolarSPELL, 2018), and computer labs and tablets for students in Vanuatu (Telecommunications and Radiocommunications Regulator, 2018), to our knowledge this is the first impact evaluation studying the use of ICT in schools in the Pacific region.

Annex II. Empirical Methods

Answering research questions 1 to 3 requires an impact evaluation of the program. We used a randomized design to evaluate KTT. Ten public schools with mono-classes were purposefully selected, and then half were randomly assigned to treatment and the other half to the control group. At baseline (February 2018), in all 10 schools, we assessed all Year 4 students and the Year 4 teacher (or teachers if there were more than one Year 4 class in the school). We assessed the same students and teachers at endline (November 2018).

Collecting longitudinal data on students allows us to estimate the impact of the program controlling for their baseline characteristics. Specifically, we estimate:

$$A_{si1} = \alpha + \beta D_s + \gamma A_{si0} + u_{si1} \quad (1)$$

where A_{si1} and A_{si0} are the outcomes of interest (i.e. students' test scores) for individual *i* in school *s* at endline and baseline, respectively; D_s is a dummy variable for treatment status; u_{si1} is an error term, and α, β and γ are parameters to be estimated. The coefficient of interest is β, that captures the effect of the program. Furthermore, because treatment is randomized, β can be considered an unbiased estimate of the program impact, as students in treatment and control groups should be observationally equivalent at baseline, so any difference at endline can be attributed to the program.

We use the model described in equation (1) to answer research question 1. In this case, the outcome of interest (A_{si1}) is students' test scores on math and English. We also include students' sociodemographic characteristics as control variables (e.g. students' age, gender, parental literacy) in order to improve the precision of the estimated program impacts.

One limitation that needs to be considered when using longitudinal data is attrition. Attrition can have three different impacts on the evaluation results. First, the fewer observations, the lower the power to identify treatment effects. Second, unless attrition is random, it could reduce the external validity of the results. For example, if disadvantaged students tend to be more likely to drop out from the sample, the estimated impacts will be valid for relatively fewer disadvantaged students, not the entire population targeted at baseline. Third, and perhaps more importantly, attrition can bias the evaluation results if different types of students tend to abandon the sample contingent on their treatment status. For example, suppose that low performing students in the control group tend to drop out from the sample more than low performing students in the treatment group. This will introduce an imbalance because, in the absence of treatment, students in the control group would outperform students in the treatment group; therefore,

the estimated treatment effects would underestimate the true impact of the project. To control for attrition, we use inverse probability weighting (IPW). This is a common technique used to address attrition and other missing data problems. Annex V describes this methodology and how we implemented it. All the regressions in the report are weighted using IPW.

To answer research question 2, we use a simpler approach. Given the small sample size of teachers that we have to answer this question, we simply estimated the double-difference of the teachers' assessments.[13] That is, we calculated the changes for teachers in the treatment group and compare them with changes in the control group. We supplement this quantitative analysis for teachers with qualitative data. Specifically, we conducted Key Informant Interviews (KIIs) with teachers in the treatment group to record their views on the effect of tablets on their English proficiency level, as well as on their students' achievement.

To answer research question 3, we used the model in equation (1), but splitting the sample between students that scored above and below the median in their English assessment at baseline. Comparing impact effects on English skills by stratifying the sample this way allows us to analyze if there are differential impacts by English performance at baseline.

We address research question 4 using both quantitative and qualitative data collected at endline. Using classroom observations, we document how the program has affected classroom pedagogy. We then use KIIs with teachers and FGDs with students to explore how the program affected teaching practices in the classroom and how students are learning. Finally, we use data from FGDs with students, FGDs with School Improvement Committee representatives, and KIIs with teachers and principals to examine their views on the program, its main benefits, implementation challenges, and ways to improve the program. Table 5 summarizes the methods used to answer each research question and the corresponding data sources.

[13] In the original design we had proposed using regression analysis similar to the one described by equation (1) to examine the impact of the program on teachers' English skills. Upon further analysis, and given the fact that the small sample size precludes from detecting effects, we concluded that the regression analysis could be misleading. Hence, we opted for the simpler approach of double-difference, and refrained from producing standard errors, to avoid using this small sample to discard program effects that could actually be significant if the sample was large enough.

Table 5. Evaluation design matrix

Research question	Indicators	Data source	Empirical method
1. Does the consistent use of tablets preloaded with curriculum-linked content and used in a supportive and supported classroom, have a significant effect on student learning in the Kiribati context?	Student test scores in math and English	Student assessments in math and English	RCT
2. Does access to a tablet loaded with English language resources improve the level of English language proficiency in teachers?	Teachers' English proficiency test scores in English assessment Teachers' views on whether or not, and how, tablets helped improve their English proficiency	Teacher survey (Includes English proficiency assessment) KII with teachers	Quantitative: RCT Qualitative: Analysis of KII with teachers
3. Does the use of tablets loaded with curriculum-linked content support the learning of struggling readers?	Student test scores in math and English	Student assessments in math and English	RCT with interaction terms
4. How is the trial affecting classroom pedagogy, interactions and perceptions of learning at school?	Use of English during class, use of tablets during class, question items addressing interactions and perceptions on the program	Classroom observations and FGD with students. Interviews with teachers and principals and community stakeholders.	Analysis of qualitative data

Annex III. Data Collection

Data were collected in February 2018 (baseline), and November 2018 (endline). Quantitative data were collected during each wave of data collection in both treatment and control schools. Qualitative data were collected at endline only, and only in treatment schools, because qualitative data mostly pertains to the implementation, adequacy, requirements, and challenges of the tablet trial.

Below is a description of each of the instruments that were fielded:

Quantitative Data

a) *Student assessments:* These assessments cover English and math skills. The student English and math assessments were designed by the KEIP office, based on the widely used Early Grade Reading Assessment (RTI International 2009) and Early Grade Math Assessment (RTI International 2014), respectively. The English assessment includes subtasks on familiar words, made-up words, oral reading, reading comprehension, and listening comprehension. The math assessment includes subtasks on missing numbers, addition, subtraction, multiplication, measurement, and word problems. The KEIP team introduced modifications and added subtasks to adapt the assessments to Kiribati's Year 4 curriculum. For example, the words used in the Familiar Word subtask were taken from the Year 4 curriculum and not from a previously fielded EGRA. Additionally, the multiplication and measurement subtasks in the math assessment were added by KEIP based on the Year 4 math curriculum. While the assessments were contextualized to the Year 4 curriculum, the final versions of the assessments were not formally adapted or pre-tested as is customary when designing new reading and math assessments. However, once the assessments were designed, NORC and our partner School-to-School International (STS) reviewed the assessments, provided feedback, and suggested changes based on our extensive experience fielding similar assessments in developing countries.

   KEIP tested the student instrument as part of the enumerator training pilot exercise at an urban school in South Tarawa. Based on the results from the pilot test and anecdotal feedback from the enumerators, NORC and STS suggested that KEIP remove the letter name identification and number discrimination subtasks; most students scored high on these subtasks, so there was little chance for the program (or any intervention) to have an impact on scores. The STS/NORC team also recommended that KEIP remove one of the oral reading fluency subtasks and the accompanying reading comprehension questions, because it was commonly used in Year 4

classrooms and, therefore, had a high risk of contamination. The KEIP team agreed with these suggestions and the subtasks were cut from the assessment prior to the start of fieldwork. Table 6 below details the name, description, duration, and number of items included in each subtask contained in the assessment that was fielded.

Table 6. Student assessment overview

Subtask Name	Subtask Description	Duration	Number of Items
	Math		
Missing Number	Identify the missing number in a number sequence	Untimed	10
Addition 1	Solve 20 addition problems	60 seconds	20
Subtraction 1	Solve 20 subtraction problems	60 seconds	20
Multiplication 1	Solve 20 multiplication problems	60 seconds	20
Measurement	Perform five math functions found in daily life, including reading an analog clock face, summing money, and measuring with ruler	Untimed	5
Addition 2	Solve five more addition problems (more difficult than those in addition 1)	Untimed	5
Subtraction 2	Solve five more subtraction problems (more difficult than those in subtraction 1)	Untimed	5
Word Problems	Read and solve five math word problems	Untimed	6
	English		
Familiar Word	Read a list of 50 words	60 seconds	50
Made-Up Words	Read a list of 50 made-up words	60 seconds	50
Oral reading fluency	Read a short 70-word passage out loud	60 seconds	70
Reading Comprehension	Answer five comprehension questions on the same 70-word passage (The number of questions varies by child depending on the amount of the text they are able to read. A zero value is automatically assigned to students who are unable to read anything.)	Untimed	5
Listening Comprehension	Answer five comprehension questions about a text read by the interviewer	Untimed	5

b) *Student survey:* NORC developed the student survey to record basic sociodemographic characteristics of the students, including age, gender, parental literacy, reading habits, household size and asset ownership, as well as previous exposure to devices similar to the tablets (smart phones, laptops, etc.). At endline, a tablet-related module was included to record data on how much time students spent using the tablets.

c) *Teacher English Proficiency assessment and teacher survey:* The teacher instrument included an English proficiency test and survey questions. The English proficiency test was developed by

KEIP and included listening, reading, writing, and speaking assessments. The design and marking of the assessment are aligned with the Common European Framework Reference (CEFR), a widely adopted framework to describe in broad terms a learner's language ability. Once the assessments were designed, NORC and STS provided feedback and suggested changes based on our prior experience with similar assessments. Table 7 below provides a description of the final teacher assessment subtasks.

Table 7. Teacher assessment overview

Subtask	Description
Listening	Examines an individual's capacity of listening for gist, listening for detail and selective listening. Consists of 20 multiple-choice questions, five true or false statements, and five sentences in total. Types of question include two-way conversation, teacher monologue, and radio interview.
Reading	Assesses the candidate's understanding of various kinds of texts, including reading for gist, comparing, evaluating and inferring, distinguishing fact from opinion, and careful reading for specific information. Reading materials include a policy document, a cloze text from a report, and five short texts on health issues in Kiribati.
Writing	Asked to complete three different tasks: responding to reading (150 words), a formal letter to the MOE (150 words), and correcting student writing (120 words of comments and suggestions).
Speaking	Includes role play with the interviewer, and monologue/presentation.

Each subtask of the teacher assessment is scored on a scale of 0 to 100, and the total score is the average of the four subtask scores. The teacher sociodemographic survey, developed by NORC, included questions on education level and certifications, teaching experience, as well as previous exposure to devices similar to the tablets.

The teacher survey includes questions on education level and certifications, teaching experience, as well as previous exposure to devices similar to the tablets. At endline, we included a tablet-related module in the teacher survey to record how teachers use the tablets in the classroom.

Qualitative Data

a) *Classroom observations*: The classroom observation rubric was developed from an instrument designed by KEIP for internal purposes. The rubric measured classroom tablet use for two lessons, one English and one math, in each treatment classroom. The rubric tracks teacher and student activities in 6-minute intervals over a period of one hour for each lesson; the rubric is similar to others like the Stallings instrument (World Bank 2015). The observation rubric also captures information about the condition of tablet trial assets in each school.

NORC developed KII and FGD scripts by breaking down the research questions that could be fully or partially answered by qualitative data. NORC drafted the interview guides in English and shared them with STS and KEIP for feedback. Once the instruments were finalized, the instruments were translated to Kiribati, the language that they were ultimately fielded in. The content of each instrument is described as follows:

b) *Key Informant Interviews with principals*: In these interviews, we recorded principals' views on different program aspects, including the design of the tablet instructional materials, interactions with teachers in the context of the program, management of program resources, and minimum requirements in schools for pre-loaded tablets to be effective.

c) *Key Informant Interviews with teachers*: In these interviews, we recorded teachers' views on the same constructs covered in the principals' KII. In addition, we collected data on teachers' experiences using the tablets in the classroom, how they think tablets had impacted student achievement and their own level of English proficiency, and their views on the quality of the tablet training.

d) *Key Informant Interviews with community stakeholders*: KIIs were conducted with School Improvement Committee representatives. In these sessions, we discussed program implementation challenges, interactions with schools, and perceived benefits of the program on student achievement.

e) *Focus Group Discussions with students*: In these sessions, we asked students about their opinions on the tablet content, how adequate they found the tablets in general, how often tablets are used in classroom, and how the program has affected their learning.

Baseline Data Collection

Baseline data collection took place at the beginning of the 2018 school year in Kiribati. KEIP conducted all data collection activities, working closely with NORC and STS.

Training

Prior to the launch of data collection, STS team member Aimee Reeves traveled to Kiribati to observe both the student and teacher enumerator training. Student enumerator training took place February 13-16, 2018 in South Tarawa, Kiribati. Trainees were recruited by KEIP from a pool of assessors who had experience conducting EGRA and EGMA data collection. The training was led by Phil Geeves, a KEIP

team member. 11 trainees, eight females and three males, participated in the training. The four days of training consisted of a brief overview of subtasks followed by practice time in pairs on Tangerine. On day 3 of the training, enumerators practiced administering the assessment to Year 4 and Year 5 students in a nearby urban school on South Tarawa. Assessor accuracy exercises were conducted on three different occasions. Enumerators generally performed well on the accuracy exercises with nearly all scoring above 90% assessor agreement across the three exercises. During the third assessor accuracy exercise on day 4, agreement was lower than desired on several subtasks; however, enumerators noted they had problems hearing the enumerator and student pair. During the training, the STS observer also noted that the English proficiency of the trainees was not as high as expected. While there were concerns about the trainees' English proficiency, the pilot test and inter-rater reliability scores proved this was not a hindrance to the data collection. STS and KEIP discussed in detail reading comprehension questions on which there was lower agreement and recommended that enumerators practice reading the stimuli together on familiar word reading and made-up word reading subtasks. At the end of the training, STS assessed the quality of the training as satisfactory and was especially enhanced by the enumerators' previous experience administering similar assessments in Tangerine.

On Wednesday, February 21, 2018, KEIP led the teacher English proficiency assessment, with NORC and STS in attendance. The training was attended by the four intended assessors, all of whom had experience or are experts in assessing teachers of English as a foreign language (TOEFL). The training objectives were for the lead designer of the assessment, who was also one of the assessors, to present the final piloted assessment and provide explicit guidance to the other assessors on protocols, timing, and operational data collection. The training did not meet these objectives: there was not a final assessment ready to be presented, and no assessor guidance was documented. It also became clear during the training that portions of the assessment hadn't been formally piloted, but only reviewed by students at the teacher training college.

In a one-on-one meeting, STS and NORC discussed their concerns about the teacher English proficiency assessment and possible improvements. NORC noted that the changes the KEIP team made based on STS feedback had improved the quality of the assessment and was satisfied with the KEIP team's decision to develop an assessor handbook prior to operational data collection. On February 22, Jacquie Widin, one of the assessors that had attended the training session, shared an Assessors Handbook with the rest of the team.

Fieldwork and quality control

Fieldwork for the student data collection was launched on February 19, 2018 and concluded on March 2, 2018. As part of the evaluation, STS observed the first day of student data collection in Maiana. During the observation, the team of four enumerators visited a control school with two Year 4 classes. The enumerators followed good practices during their visit, including introducing themselves to the head teacher, explaining the reason for their visit, and building rapport with the staff and students. The experience of the enumerators was clear, as they were efficient and intentional in their procedures. Through the entire baseline data collection, a total of 261 students were assessed. Teacher data was collected between February 27 and March 16, 2018. A total of 14 teachers were surveyed.

KEIP delivered the student data to NORC on March 13. After receiving the data, NORC performed a data quality review (DQR) in order to ensure that the delivered data meets NORC's standards. The three main criteria that guide our data quality assessments are that the data need to be complete, accurate and internally consistent. We used multi-stage data cleaning plans ensuring all data values are within allowable range and reserve codes are used appropriately. Through the DQR NORC noted that over 90% of students were not able to provide information on their caregiver's educational level. NORC made a note of this and included a question in the endline student survey to try and capture the data in a different format.

Endline Data Collection

Qualitative data

Qualitative data focused on treatment schools and included classroom observations, FGDs with Year 4 students and parents, and KIIs with Year 4 teacher, school principals, and local leaders. The table below details qualitative data collection in treatment schools.

Table 8. Qualitative data collection by school

Primary school	Classroom Observations	Principal KII	Teacher KII	Student FGD	Parent FGD	Leader KII
Abitabu	2	1	2	1	-	-
Nein Tebwara	1	1	1	-	-	-
Nikierere	2	1	2	1	1	1
St Paul	1	1	1	1	1	1
Taiti	2	1	2	-	-	-

Classroom observations

The classroom observation training was held October 25-26 in South Tarawa with two participants, one female and one male. NORC recruited two of the baseline enumerators to participate in the two-day training and conduct the classroom observations. Day 1 of the training provided the participants with background on the tablet trial and presented the classroom observation tool. During the second day of training, the team traveled to Nein Tebwara, a treatment school on North Tarawa, to conduct a classroom observation together. For this exercise, each observer and the NORC trainer observed an English and math lesson and marked the observation rubric accordingly. The results of the classroom observation were then compared between the fieldworkers and the NORC trainer to ensure the observation tools were being implemented properly. At the conclusion of the test observation at Nein Tebwara, the NORC trainer concluded that the instrument was being correctly implemented.

The remaining classroom observations were fielded October 29, 2018 through November 2, 2018, with the exception of St. Paul, where the observation took place on November 15 due to logistical delays. During this fieldwork, NORC staff observed 1 classroom observation in each school on Marakei for quality assurance. At the end of each day of classroom observations, NORC staff reviewed the data to ensure it was complete, accurate, and internally consistent.

FGDs and KIIs

Training for FGDs and KIIs was held November 12 -14, 2018 in South Tarawa with four participants, two female and two male. This training included 4 fieldworkers (2 moderators and 2 note-takers), who were interviewed and recruited by NORC staff, along with the NORC trainer. During this training, the participants were again given a description of the project and then presented with each of the individual KII and FGD protocols. For the second day of training, the team traveled to Nein Tebwara in South Tarawa to pilot test the student FGD protocol. During this pilot, the two moderators took turns asking a group of students the FGD questions while the NORC trainer observed. During the second day the fieldworkers also conducted the KIIs with the Nein Tebwara teacher and principal. On the third day of training the team reconvened in South Tawara and debriefed on the pilot FGD and two KIIs conducted the previous days. Based on the pilot feedback, NORC finalized the KII and FGD protocols and prepared them for fieldwork.

Data collection was launched on November 15, 2018 and concluded on November 21, 2018. At the conclusion of fieldwork, the field team transcribed interview notes and audio recordings and delivered the Te-Kiribati transcripts to NORC in November 2018. NORC sent the transcripts for translation and

received the final translated transcripts that were used for analysis on February 10, 2019. During analysis, NORC loaded the transcripts in Nvivo to review, code, and analyze them.

Quantitative training

The student assessment enumerator training took place from Thursday, November 8 to Friday, November 9, 2018. The training was again led by KEIP team member Phil Geeves, who led baseline student assessment enumerator training. The training was attended by seven assessors, five females and two males (all of whom participated in the baseline assessment and had previous experience administering EGRA and EGMA prior to their work with KEIP), and the NORC and STS representatives. The two days of training consisted of a brief overview of the subtasks and protocols followed by practice time on Tangerine. Because the in-school practice was scheduled on day 1 of the training due to a school activity day on day 2 of the training, the first day of training focused only on EGRA and the student questionnaire. EGMA subtasks were refreshed on day 2 of the training. Additionally, NORC provided a detailed review of the student questionnaire questions to ensure adequate training.

In the afternoon on day 1 of the training, assessors practiced administering the survey's introduction and consent, the English assessment, and student questionnaire portions of the assessment to Year 4 students in a nearby urban school on South Tarawa. Each assessor practiced with four students. STS led a short debriefing session following the in-school practice to provide feedback to the assessors and highlight areas for improvement.

Assessor accuracy exercises were conducted on three different occasions on day 2. Overall, assessors performed very well on the accuracy exercises with all scoring above 90 percent accuracy across the three exercises. Specific areas for improvement were highlighted with assessors during a post-assessor accuracy debrief session on day 2.

Quantitative fieldwork and quality control

Fieldwork for the student data collection was launched on November 15, 2018 and concluded on November 21, 2018. As part of the evaluation, STS observed the first three days of student data collection in Marakei. During STS's observation, the team of assessors visited one treatment school with two year 4 classes and one control school with one year 4 class. The assessors followed good practices during their visit, including introducing themselves to the head teacher, explaining the reason for their visit, and building rapport with the staff and students. Out of the 261 students surveyed at baseline, 202 were

surveyed at endline, for an attrition rate of 22.6 percent. These 202 students constitute the panel sample, which we use for the impact evaluation analysis. In addition, 36 new students were surveyed at endline.

Fourteen teachers were surveyed, 12 of which had also been surveyed at baseline.

Teacher assessment data was collected in November 21-28, 2018. 14 teachers were surveyed, eight treatment and six controls. Of these 14 teachers, 12 were also surveyed at baseline. Table 9 summarizes the quantitative data collection results.

Table 9. Surveyed samples of students and teachers

Respondent	Only baseline	Only endline	Panel (both baseline and endline)
Students	59	36	202
Teachers	2	2	12

KEIP delivered the student data to NORC on November 26. After receiving the data, NORC performed a DQR in order to ensure that the delivered data meets NORC's standards.

Ethics clearance

The tablet trials was approved by Kiribati's Ministry of Education. NORC's IRB exempted the study from a review.

Annex IV. Student Sociodemographic Characteristics

In this section we review the results of the baseline data collection. The two main purposes of this section are to (1) describe the general sociodemographic makeup of the target population of students, and (2) discuss differences between treatment and control groups across their sociodemographic characteristics.

Table 10 shows student demographic characteristics at baseline by treatment group. A little over half the surveyed students were girls. On average, students are nine years old and most of them had breakfast the day they were surveyed. On average, it takes students 10 minutes to get to school. The week before data collection, students attended school 4.7 days, on average. Households have, on average, just over seven members. To approximate household wealth, we asked children whether their households had a clock, a radio, a television, a computer, a bicycle, a motorcycle, and a car. We extracted the first principal component of these items to construct an asset index. The index is centered at 0 which is why households can have a negative index. The results indicate that households in the control group have more of these items than their counterparts in the treatment group, although the differences are not statistically significant. Finally, most parents can read in both the treatment and control group according to the surveyed students, this is relatively higher than the literacy rates for people older than 19 years old in Kiribati, which according to the 2015 census is 78 percent.

We did not find statistically significant differences between the treatment and control groups for any of these characteristics, which is to be expected given that treatment was assigned randomly.

Table 10. Baseline characteristics

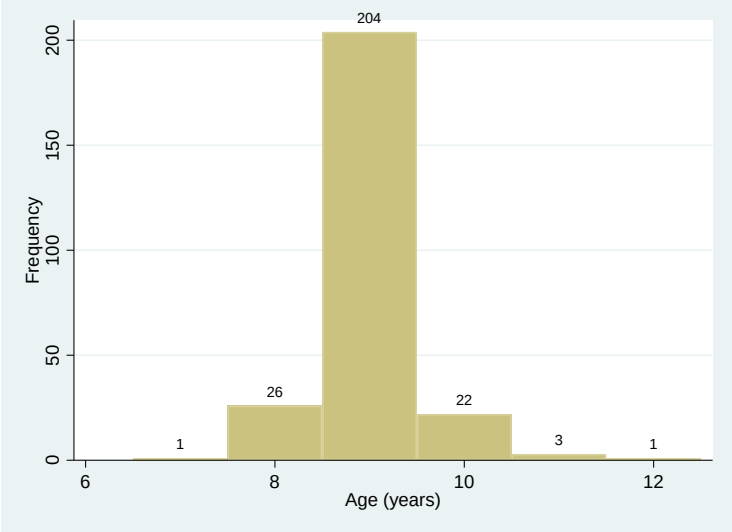
	Control (N=124) (1)	Treatment (N=137) (2)	Diff (3)=(1)-(2)
Female Student	0.55	0.57	-0.02
Student Age	9.01	9.01	-0.01
Travel Time to School	11.66	9.43	2.23*
Had Breakfast	0.91	0.88	0.03
Days Attended Last Week	4.70	4.70	0.00
Household Size	7.15	7.23	-0.08
Asset index	0.26	-0.23	0.49**
Mother can Read	0.95	0.97	-0.01
Father can Read	0.94	0.94	-0.00

Note: The asset index is the first principal component of the household ownership of the following assets: clock, a radio, a television, a computer, a bicycle, a motorcycle, and a car.
Source: Baseline student survey.

* p<0.1 ** p<0.05 *** p<0.01

As Figure 18 shows, most students are 9 years old, and only a handful are outside the range of 8-10 years old. This indicates a low prevalence of overaged children in these schools.

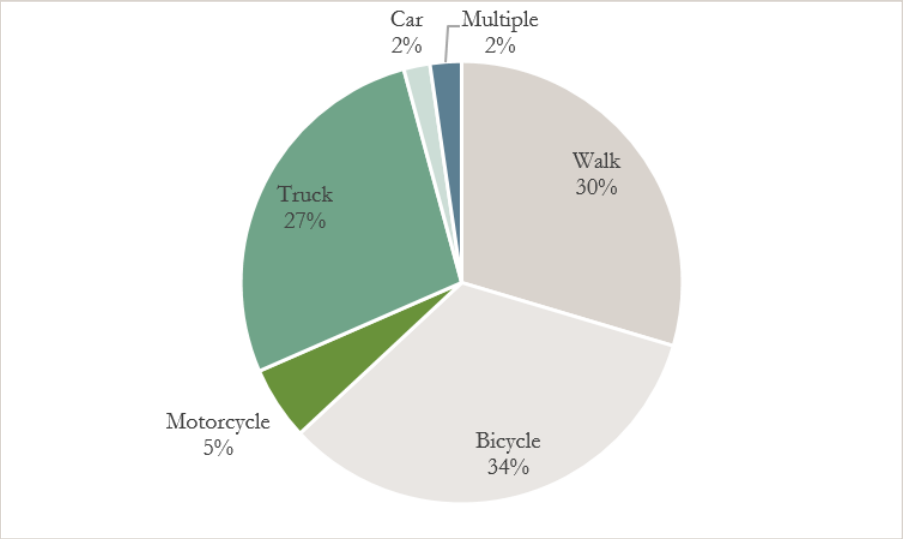
Figure 18. Age distribution



Source: Baseline student survey.

Figure 19 shows that the three most common ways children go to school are bicycling (34 percent), walking (30 percent) and by truck (27 percent).

Figure 19. How students get to school



Source: Baseline student survey.

Data on exposure to computers, summarized in Table 11, shows the fractions of students who said they had used smart phones, tablets, laptops and desktops, by treatment status. About two thirds of children had used smart phones, both in treatment and control groups. For tablets, there is a substantial difference by treatment status, as 47 percent of treatment students said they have used tablets while the fraction for the control group is only 23 percent. This is because, as was mentioned before, the program had started a couple of weeks before baseline data was collected. Very few students had used a laptop and almost none a desktop computer.

Table 11. Previous exposure to computers

	Control (N=124) (1)	Treatment (N=137) (2)	Diff (3)=(1)-(2)
Smart phone	0.69	0.68	0.01
Tablet	0.23	0.47	-0.24***
Laptop	0.12	0.09	0.03
Desktop	0.00	0.01	-0.01

* p<0.1 ** p<0.05 *** p<0.01
Source: Baseline student survey.

Annex V. Controlling for Attrition

One limitation that needs to be considered when using longitudinal data is attrition. Attrition can have three different impacts on the evaluation results. First, the fewer observations, the lower the power to identify treatment effects. Second, unless attrition is random, it could reduce the external validity of the results. For example, if more disadvantaged students tend to be more likely to drop out from the sample, the estimated impacts will be valid for relatively less disadvantaged students only, not the entire population targeted at baseline. Third, and perhaps more importantly, attrition can bias the evaluation results if different types of students tend to abandon the sample contingent on their treatment status. For example, suppose that low performing students in the control group tend to dropout from the sample more than low performing students in the treatment group. This will introduce an imbalance because, in the absence of treatment, students in the control group would outperform students in the treatment group; therefore, the estimated treatment effects would underestimate the true impact of the project.

Attrition was relatively high. Out of the 261 students surveyed at baseline 202 were surveyed at endline, for an attrition rate of 22.6 percent. To control for attrition we use inverse probability weighting (IPW). This is a common technique used to address attrition and other missing data problems. In this method, we first estimate the probability that each of the students surveyed at baseline would abandon the sample. To do this, we first estimate a logit regression where the outcome of interest is an indicator variable for students that are observed at baseline but not observed at endline, and the covariates are a set of baseline characteristics. Table 12 shows the results of this exercise. Attrition is negatively correlated with treatment status; in other words, students in the terminate group were more likely to remain in the sample than students in the control group. Attrition is also negatively correlated with mother literacy and having ate breakfast the day of the survey. On the other hand, the correlation with the English index is positive.

Table 12. Determinants of attrition

	Coef	se
Treatment	-0.759**	(0.325)
Math index	-0.364	(0.244)
English index	0.541***	(0.140)
Female	-0.427	(0.263)
Age	0.0992	(0.318)
Asset index	-0.0219	(0.202)
Ate breakfast	-0.739**	(0.318)
Mother literate	-1.109**	(0.472)
Father literate	-0.000645	(0.648)
N	258	

Standard errors are clustered at the school level
* p<.1 ** p<.05 *** p<.01

We use this model to estimate $\hat{p}$, the probability that students observed at baseline are not observed at endline, and use these estimated probabilities to construct IPW weights following:

$$w^{IPW} = 1/(1 - \hat{p})$$

These weights overweight students that are more likely to drop from the sample, and underweight students that are more likely to be surveyed both at baseline and endline, so the weighted panel sample resembles the characteristics of the original sample.

All regressions in the main body of the report use these weights, as a way to control for attrition in the evaluation results. The key assumption of this method is that attrition is driven by observable characteristics.

Annex VI. Regression Models Showing Results for all Included Covariates

Table 13. Regressions results for English – full list of covariates

	Word reading	Made-up words	Oral Reading Fluency	Reading comp	Listening comp
Treatment effect	1.30	4.76***	5.12*	0.53**	0.35
	(1.73)	(1.39)	(2.39)	(0.23)	(0.23)
Baseline value of the outcome of interest	1.35***	1.18***	1.34***	0.80***	0.48***
	(0.06)	(0.12)	(0.08)	(0.11)	(0.07)
Child's gender (female)	-0.63	0.53	1.01	0.36*	0.06
	(1.32)	(1.57)	(2.47)	(0.16)	(0.11)
Age	-0.62	0.48	-0.52	-0.17	0.05
	(1.20)	(0.76)	(3.76)	(0.19)	(0.09)
Asset index	0.47	1.44*	0.36	0.25**	0.08
	(0.75)	(0.74)	(1.70)	(0.08)	(0.13)
Had Breakfast	0.09	2.25	3.64	-0.06	0.01
	(2.47)	(1.61)	(3.74)	(0.19)	(0.23)
Mother can Read	1.37	3.84	12.44***	0.53*	0.31
	(3.01)	(4.36)	(3.05)	(0.27)	(0.50)
Father can Read	5.90**	1.49	9.05**	0.17	-0.41
	(2.42)	(3.63)	(3.94)	(0.47)	(0.31)
	199	198	199	199	199

Standard errors clustered at the school level in parentheses. * p<0.1 ** p<0.05 *** p<0.01

Table 14. Regressions results for math – full list of covariates

	Word problems	Missing number	Addition 1	Subtraction 1	Multiplication	Measurement	Addition 2	Subtraction 2
Treatment effect	0.09	0.28	3.10	0.91	2.41*	-0.16	-0.04	0.09
	(0.21)	(0.60)	(1.82)	(0.65)	(1.17)	(0.21)	(0.36)	(0.30)
Baseline value of the outcome of interest	0.44***	0.53***	1.15***	0.86***	0.83***	0.26***	0.49***	0.65***
	(0.05)	(0.09)	(0.17)	(0.07)	(0.23)	(0.08)	(0.07)	(0.08)
Child's gender (female)	0.28	-0.11	1.03	-0.23	0.69	-0.03	0.27*	0.31
	(0.22)	(0.23)	(0.83)	(0.48)	(0.59)	(0.09)	(0.14)	(0.20)
Age	-0.21	-0.13	-0.45	-0.24	0.55	-0.23	-0.33**	-0.05
	(0.22)	(0.29)	(1.07)	(0.74)	(0.42)	(0.13)	(0.13)	(0.14)
Asset index	0.13	0.19	-0.12	0.08	0.79*	0.06	-0.09	0.08
	(0.15)	(0.15)	(0.48)	(0.30)	(0.40)	(0.05)	(0.13)	(0.13)
Had Breakfast	0.03	0.89*	1.39	-1.27*	0.72	0.25	0.18	-0.47
	(0.25)	(0.42)	(1.53)	(0.59)	(0.74)	(0.16)	(0.22)	(0.40)
Mother can Read	-0.84	-0.67	-6.24	-1.34	0.25	0.31	0.03	-0.08
	(0.86)	(0.66)	(5.45)	(1.86)	(1.75)	(0.34)	(0.41)	(0.74)
Father can Read	0.45	-1.71***	-0.87	-2.72	0.45	-0.26	-0.00	-0.61
	(0.43)	(0.45)	(2.68)	(1.86)	(1.12)	(0.28)	(0.30)	(0.62)
N	199	199	199	199	199	199	199	199

Standard errors clustered at the school level in parentheses. * p<0.1 ** p<0.05 *** p<0.01