THE IMPACT OF USING TABLETS IN UNDEREQUIPPED SCHOOLS: EXPERIMENTAL EVIDENCE FROM KIRIBATI

Alejandro Ome

ome-alejandro@norc.org

NORC at the University of Chicago

55 E Monroe, Ste 30, Chicago, IL 60603

Carlos Fierros III

fierros-carlos@norc.org

NORC at the University of Chicago

55 E Monroe, Ste 30, Chicago, IL 60603

THE IMPACT OF USING TABLETS IN UNDEREQUIPPED SCHOOLS: EXPERIMENTAL EVIDENCE FROM KIRIBATI

Abstract

This paper documents the impact of an intervention that distributed tablets preloaded with curriculum aligned learning materials in a group of schools in Kiribati. We use an experimental design where five schools were assigned to treatment and five to control groups, and evaluate the impact of the program on English and math skills, after one year of program implementation. We find that the program had a positive impact on English skills of about 28 percent of a standard deviation. For math, we find some evidence of positive effects, but the results are sensitive to changes in the specifications.

Article Information

Word count: 5243

Keywords:
Early Childhood Education
Literacy
ICT Based
Impact Evaluation

THE IMPACT OF USING TABLETS IN UNDEREQUIPPED SCHOOLS: EXPERIMENTAL EVIDENCE FROM KIRIBATI

1. Introduction

Many low-income countries struggle to provide textbooks and other reading materials in schools. In these contexts, using tablets, or computers in general, as a vehicle to distribute learning content, constitutes a promising policy alternative. We document the results of the Kiribati tablet trial (KTT), a small-scale intervention where pre-loaded tablets were given to students for the 2018 school year. We use a randomized design to analyze the impact of this program on students' English and math skills. We find that the program had a positive impact on English skills of 28.2 percent of a standard deviation. For math, we find some evidence of positive effects, but the results are sensitive to changes in the specifications.

The positive results of this study contrast with most of the evidence on the impact of computers in schools (Beuermann et al., 2015; Barrera and Linden, 2009; Berlinski & Busso, 2017; Piper, Zuilkowski, Kwayumba, & Strigel, 2016). Literature reviews on the use of computers also present mixed evidence on the impact of this type of interventions (McEwan, 2015; Bulman & Fairlie, 2016), although more recent studies show promising results on using technology to personalize instruction (Muralidharan, Singh, & Ganimian, 2019). The variance in impact evaluation results is hardly surprising given that interventions can vary greatly across key dimensions (Bando, Gallego, Gertler, & Fonseca, 2017; Berlinski & Busso, 2017). These dimensions include the content available in the computers, the extent to which teachers incorporate the new resources in class, and what alternative resources are available for students, among other factors.

We argue that KTT's characteristics across these dimensions could explain why the program was successful in improving academic achievement, while other similar interventions were not. First, content was curated so it followed the national curriculum; previous research has shown the importance of aligning computer content to school curricula and children's needs (Beuermann, Cristia, Cueto, Malamud, & Cruz-Aguayo, 2015). Second, teachers did incorporate the use of tablets in the classroom; in effect, classroom observations indicate that student tablets were used about 38.8 percent of the time; as Barrera-Osorio & Linden (2009) show, for computers to have an impact they need to be integrated in the class. And third, most primary school students in Kiribati do not have textbooks, so the comparison is made between tablets and essentially no learning tools at the student level (teachers rely on big books, posters, and pictures to teach class). Other studies have analyzed the impact of computers as complements of printed materials (Piper, Zuilkowski, Kwayumba, & Strigel, 2016; Habyarimana & Sabarwal, 2018), and found that computers add very little when students have access to textbooks.

This research contributes to the literature on the use of information and communications technology (ICT) in schools by analyzing an intervention that addressed two factors that previous work had found to be key for ICT programs' success: that content is aligned to the curriculum and that teachers incorporate the use of tablets in class. The documented positive results confirm the importance of these factors for program design.

The main limitation of this study is the small number of schools that are part of the evaluation, as there are only 5 schools in each treatment arm. This is a consequence of how dispersed and hard-to-reach the country is, which makes it difficult to conduct large scale evaluations.[1] This also explains why impact evaluations are rare in Kiribati (to our knowledge,

[1] Kiribati is composed of more than 32 atolls with a total land area of 811 km^2, that are distributed over 3.5 million km^2 (about the size of India), making it one of the most sparsely populated and hard-to-reach countries in the world.

this is the first school-level randomized control trial conducted in the country). Clustering the standard errors at the school level (the usual approach in evaluations like this) tends to over-reject when the number of clusters is small. To avoid this we estimate p-values from a wild clustered bootstrap (Cameron, Gelbach, & Miller, 2008).

The next section of this paper describes the program. Section 3 discusses the data. Section 4 describes program take-up. Section 5 sketches the methods used and shows the evaluation results. Section 6 concludes.

2. Context and program background

Kiribati is an island country in the Pacific Ocean. Its population in 2017 was 114,158, with a per capita GDP of 2,180 current USD (PPP) (World Bank, 2019). The education system consists of six years of primary education followed by three years of junior secondary school. These first nine years of basic education are free and compulsory. At the end of junior secondary school, students sit for an exam to determine their eligibility for senior secondary school, which lasts four years. As is the case in many developing countries, while most children are enrolled in school (according to UNESCO Institute for Statistics (2019), the out of school rate for primary school age children was 3.9 in 2017), students' performance in standardized assessments is relatively poor. Results from the 2017 Standardized Test of Achievement in Kiribati (STAKi), a national assessment conducted in Years 4, 6 and 8 covering literacy and numeracy, show that only 44 percent of Year 4 students meet the standards for te-Kiribati (the local language), and for numeracy this figure is 64 percent (Smith & McNaughton, 2018).

Since 2012, the official education policy requires teachers to use English 50 percent of class time in Year 4 (Ministry of Education, 2017). While education systems in many low income countries struggle with provision of textbooks and other reading materials, the current situation in Kiribati is particularly critical given this relatively recent shift in language of instruction, and subsequent increase in demand for reading resources in English.

In effect, the scarcity of printed materials in English was one of the main reasons why the Kiribati Education Improvement Program (KEIP), a Kiribati-Australia partnership, funded by the Australian Government and aimed at improving Kiribati's education, fielded the KTT. The program provided students in five schools with an ICT package that included 40 tablets for Year 4 students, a docking station to house and charge the tablets, one teacher tablet, and a wireless projector that allowed teachers to show content from their tablets. KTT also constructed solar charging stations at each treatment schools, so schools could charge the equipment.

Student tablets were pre-loaded with resources aligned with the national Year 4 curriculum standards and content.[2] The student tablet content included both (1) lesson materials following the curriculum of each term, and (2) resource materials to reinforce the lesson materials. The lesson materials included resources for English, math, and science instruction, and were comprised of Year 4 weekly lessons throughout the school year. The resource materials included readers, learning activities and games to practice English and math, supplementary reading materials, and audio and video media aimed at strengthening oral English pronunciation and confidence. Tablets were intended to be used everyday day for these lessons.

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. Moreover, the students were only able to access their tablets in the classroom (within a 20 m radius from the docking station) and with the permission of the Year 4 teacher. The teachers' tablets were more versatile than students' tablets in that the

[2] The materials also had to meet the following criteria: (1) culturally appropriate, (2) free and unencumbered with copy right, and (3) able to be downloaded on to the kio-kits (i.e. the tablet kits).

teachers were able to take their tablets home to continue using them, and they were able to connect to the internet for a wider breadth of resources. In addition to all the students' tablet content, the teacher tablet included (1) additional content for English language strengthening, (2) resource materials, and (3) resources for the teachers' recreational use, especially in relation to their mastery of English language.

The program was implemented in four islands near South Tarawa, Kiribati's most populated island. All 10 mono-grade schools on these four islands were part of the trial evaluation.[3] These 10 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. Five schools were assigned to the treatment group and five to the control group.

KEIP provided three teacher training sessions over the course of the intervention. The first one prior to the start of the school year, the second one between the first and second quarter, and the third one at the end of the second quarter. KEIP's reports and teacher survey data indicate that most teachers attended all three training sessions.

3. Data

Baseline data was collected at the beginning of the school year, in February 2018. A total of 261 students, 137 treatment and 124 control, were surveyed. At endline 238 students were surveyed, out of whom 202 were also surveyed at baseline. As Table 1 shows, the attrition rate for the treatment group was 17.5 percent, while for the control group was 28.2 percent. This difference in the dropout rates between the study groups casts doubts on the comparability of the treatment

[3] 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.

and control groups at endline. We use inverse probability weights (IPW) to try to control for this missing data problem. We also estimate Lee bounds (Lee, 2009) of the treatment effect.

Table 2 shows summary statistics at baseline by treatment status for the panel sample (i.e. the students that were surveyed both at baseline and endline). Panel A shows students' characteristics. A little over half the surveyed students are 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. According to students' self-reports, the week before data collection they attended school 4.7 days, on average.

To assess students' English and math skills, we fielded instruments similar to the widely used Early Grade Reading Assessment (RTI International 2009) and Early Grade Math Assessment (RTI International 2014). For English the instrument assessed five reading subtasks, namely reading familiar words, reading made-up words, reading of a paragraph, reading comprehension, and listening comprehension. For math the assessment measured missing numbers identification, multiplication, measurement, word problems, two sets of addition problems and two sets of subtraction problems, for a total of eight subtasks.4 To summarize these subtasks, we constructed one index for English and another for math. For English we use the first principal component of the five evaluated subtasks, standardized using the control mean and standard deviation (Kerwin & Thornton, 2018). We followed a similar procedure to construct the math score. Panel B shows achievement scores for math and English. Scores are standardized using the wave-specific control mean and standard deviation. This included all (control) students assessed at each wave of data collection, regardless of whether they are part of

[4] Instructions for both the English and math assessments were administered in the local language (te-Kiribati).

the panel or surveyed only once, which is why the control means in Table 2 are not zero. Although the differences are not significant, for English they are relatively large.

Panel C shows households' characteristics. 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 results indicate that households in the control group have more of these items than their counterparts in the treatment group. 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.

Overall, although there are significant differences for one of the 11 outcomes analyzed, the two experimental groups are well-balanced. However, the difference between treatment and control groups for the English scores is relatively high, which could compromise the validity of the experimental design. As we explain in section 5, to offset differences at baseline we include the baseline scores as control variables in our estimates of the treatment effect.

Data collection at both baseline and endline also included a teacher English proficiency assessment, and a teacher survey. At endline, we also conducted: 16 classroom observations; five KIIs with school principals; eight KII with teachers; three FGDs with students; two FGDs with parents; and two FGDs with local leaders, all in the treatment group.

4. Program take-up

Classroom observations reveal a relatively high use of program resources. The 16 classroom observations were conducted in a total of 8 classrooms (more classrooms than treatment schools were observed because there were two 4th grade classrooms in three of the treatment schools). In each classroom one English and one math lesson were observed. Note that these observations were conducted at endline, so teachers had had the whole year to get used to incorporating program resources in class. They were also aware of when the observation was going to be conducted.

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 1 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, and 38.8 percent of the time conducting activities related to the students' tablets.

Figure 2 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 83.1 percent of the time.

Students' surveys also included items to record tablet use. As Figure 3 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; 49 percent of students used the tablets less than half the day, 40 percent used them half the day and 11 percent used them all day or almost all day.

During the student FGDs, students reported that their favorite uses of the tablets were reading, listening to songs, playing math games, and watching 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, 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. This suggests that the program helped teachers organize class content and structure without having children use the tablets excessively. Notably, tablet usage data from students' survey (shown in Figure 3) 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. Finally, both quantitative and qualitative data indicate that teachers and students found it easy and useful to use the tablets.

5. Empirical method and results

To estimate the impact of the program, we analyze students' achievement adjusting for baseline characteristics. Specifically, 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. To control for attrition, we estimate models using IPW.[5] Standard errors are clustered at the school level. Given the small number of clusters in the data

[5] To construct the weights we estimate the propensity to attrite as a function of treatment status, baseline scores for English and math, children's age and gender, the asset index, binary variables for parental literacy (one for the mother's and another for the father's) and for whether they child had breakfast that day. The result of this regression are available upon request.

(10 schools), we estimate p-values from a wild clustered bootstrap, as recommended by Cameron, Gelbach, & Miller (2008).

Regressions results are shown in Table 3. Panel A shows results for English. For the treatment effect we show the point estimate, the standard errors clustered at the school level in parentheses, and the p-values from the wild clustered bootstrap in square brackets. In the first column, only the raw comparison between treatment and control at endline is shown. For English the raw difference between the two groups is 30.5 percent of a standard deviation, but is not statistically significant. In the second column, control variables are included, namely the value of the outcome of interest at baseline, children's age and gender, the asset index, dummy variables for parental literacy (one for the mother's and another for the father's) and for whether the child had breakfast that day, and island fixed effects. The parameter is also positive, about four percentage points smaller than the one in the first column, but statically significant (at 10 percent of confidence). The third column corresponds to the same results of column two except that we use IPW to control for attrition. The result for the IPW specification indicate an impact of 28.2 percent of a standard deviation.6 In general, program impacts between 20 and 40 percent of a standard deviation are considered large. The What Works Clearinghouse, a US based repository of high-quality impact evaluations, characterizes impacts greater than 25 percent of a standard deviation as "substantively important" (What Works Clearinghouse, 2014 p.23).

Panel B shows results for math. While the raw difference, shown in the first column, is only 11.9 percent of a standard deviation, once controls are included the coefficient more than doubles and is (marginally) significant. The key driver of this change is the introduction of island

[6] Given the small sample, it is perhaps surprising that an effect of this size is estimated with standard precision. It is worth highlighting that the intra-school correlation for both the endline reading and math scores are quite low, namely 0.07 and 0.11, respectively. This partly explains why standard errors are small enough so significant impacts can be estimated.

fixed effects. In one particular island where the treatment group was overrepresented, the math results declined across the board, so including fixed effects isolates this dynamic. Results in the third column show that once IPW are included, the intervention had a positive impact of 28.7 percent of a standard deviation for math.

Student attrition was high. While IPW helps controlling for attrition, the underlying assumption on selection on observables is debatable. A less restrictive approach to analyze the implications of attrition is using Lee bounds. This method trims the treatment group from above and below the outcome distribution, in order to restrict the comparison to those students who are always observed. The key underlying assumption is that treatment affects attrition in one direction only (monotonicity). Given the substantial difference in attrition rates between treatments groups, which in turn implies that a substantial fraction of the treatment sample needs to be trimmed to calculate the bounds, it is not surprising that the methods produce non-informative bounds for both English and math, as shown in columns (1) and (2) of Table 4. In columns (3) and (4) we use island indicators to narrow the bounds, but the lower bound is significant for math only. It is worth mentioning that the fact that attrition was much higher for the control than for the treatment group suggests that the program may have had a positive impact on school attendance. Students could only use the tablets while in school, so maybe the novelty of it incentivized treated students to attend school more, or parents in the treatment schools saw more value in attending school. In effect, during focus groups discussions, parents reported that students in the treatment group were more eager to attend schools because of the tablets.

We also analyze differential treatment effects by gender. Table 5 shows regression results including an interaction term between treatment and a female dummy. The interaction terms are very small for both English and math and not statistically significant, suggesting that there were not differential impacts by gender.

6. Discussion and conclusion

This study contributes to the literature on computers and learning. More specifically, we document the impact of using tablets in schools in a context where no other student-level learning resources are available. We found that the program had an impact of 28.2 percent of a standard deviation on students' English skills. We also found some evidence of positive effects for math, but the results are sensitive to changes in the specifications. Specifically, while the raw difference in math scores between treatment and control is relatively small and not significant, including island fixed effects substantially increases the treatment effect. This is because in one particular island where the treatment group was overrepresented, the math results declined across the board. Once we control for island fixed effects and other covariates, and include IPW, we find that the program had an impact of 28.7 percent of a standard deviation. We did not find evidence that the program had differential impacts by students' gender.

Survey attrition was non-negligible. Not only was the overall attrition rate high, but the control group's (28.2 percent), was much higher than the treatment's (17.5 percent). Although we control for attrition using IPW (and the results are not too sensitive to this approach), Lee bounds produce non-informative bounds, casting doubts on the robustness of the documented results. It is possible that the small sample prevents us to find lower bounds that are statiscally significant. However, it is also important to highlight that differential attrition could be reflecting a treatment impact on school attendance. Qualitative data supports this hypothesis, as parents in the treatment group said their children were excited about going to school thanks to the tablets; however whether the program did in fact impact retention remains an open question.

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. 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. 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, for which attendance was high, most likely played an important role.

Acknowledgments

This study was funded by the Australian Department of Foreign Affairs and Trade through the Kiribati Education Improvement Program. The authors’ views expressed here are their own and not necessarily those of the Australian Government or the Government of Kiribati. The authors would like to thank Helen Stannard, Erin Blake, Alicia Menendez, Ofer Malamud, Adrienne Lucas, staff at Coffey International, and participants at seminars at NORC at the University of Chicago and the North East Universities Development Consortium Conference, who provided valuable comments and suggestions.

References

Bando, R., Gallego, F., Gertler, P., & Fonseca, D. R. (2017). Books or laptops? The effect of shifting from printed to digital delivery of educational content on learning. *Economics of Education Review*, *61*, 162–173. https://doi.org/10.1016/j.econedurev.2017.07.005

Barrera-Osorio, F., & Linden, L. L. (2009). The Use and Misuse of Computers in Education: Evidence from a Randomized Experiment in Colombia. Impact Evaluation Series No. 29. Policy Research Working Paper 4836. *Human Development Network Education*, (February). Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED509325&site=ehost-live%5Cnhttp://go.worldbank.org/46RX8ZK7U0

Berlinski, S., & Busso, M. (2017). Challenges in educational reform: An experiment on active learning in mathematics. *Economics Letters*, *156*, 172–175. https://doi.org/10.1016/j.econlet.2017.05.007

Beuermann, D. W., Cristia, J., Cueto, S., Malamud, O., & Cruz-Aguayo, Y. (2015). One laptop per child at home: Short-term impacts from a randomized experiment in Peru. *American Economic Journal: Applied Economics*, *7*(2), 53–80. https://doi.org/10.1257/app.20130267

Bulman, G., & Fairlie, R. W. (2016). Technology and Education: Computers, Software, and the Internet. In *Handbook of the Economics of Education* (1st ed., Vol. 5). https://doi.org/10.1016/B978-0-444-63459-7.00005-1

Cameron, C. A., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, *90*(3), 414–427. https://doi.org/10.1162/rest.90.3.414

Habyarimana, J., & Sabarwal, S. (2018). Re-Kindling Learning: eReaders in Lagos. *Policy Research Working Paper*, *8665*(December).

Kerwin, J., & Thornton, R. L. (2018). Making the Grade: The Sensitivity of Education Program Effectiveness to Input Choices and Outcome Measures. *Working Paper*. https://doi.org/10.2139/ssrn.3002723

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies*, *76*(3), 1071–1102. https://doi.org/10.1111/j.1467-937X.2009.00536.x

McEwan, P. J. (2015). Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments. *Review of Educational Research*, *85*(3), 353–394. https://doi.org/10.3102/0034654314553127

Ministry of Education. (2017). *National Curriculum and Assessment Framework*. Bikenibeu, Republic of Kiribati.

Muralidharan, K., Singh, A., & Ganimian, A.J. (2019). Disrupting Education? Experimental

Evidence on Technology-Aided Instruction in India. *American Economic Review*, 109(4): 1426–1460. https://doi.org/10.1257/aer.20171112

Piper, B., Zuilkowski, S. S., Kwayumba, D., & Strigel, C. (2016). Does technology improve reading outcomes? Comparing the effectiveness and cost-effectiveness of ICT interventions for early grade reading in Kenya. *International Journal of Educational Development*, *49*, 204–214. https://doi.org/10.1016/j.ijedudev.2016.03.006

Roodman, D., MacKinnon, J. G., Nielsen, M. Ø., & Webb, M. D. (2019). Fast and wild: Bootstrap inference in Stata using boottest. *Stata Journal*, *19*(1), 4–60. https://doi.org/10.1177/1536867X19830877

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.

Smith, M., & McNaughton, B. (2018). *Final Report Kiribati Education Improvement Program (KEIP)*.

Tauchmann, H. (2014). Lee (2009) treatment-effect bounds for nonrandom sample selection. *Stata Journal*, *14*(4), 884–894. https://doi.org/10.1177/1536867X1401400411

UNESCO Institute for Statistics. 2019. Data From: Rate of our-of-school children of primary school age.

What Works Clearinghouse. (2014). *Procedures and Standards Handbook 3.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, What Works Clearinghouse.

World Bank. 2019. World Bank Open Data. World Bank, Washington, DC. © World Bank. Retrieved from https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=KI

Tables

Table 1. Sample

	Baseline	Endline	Panel	Attrition
Control	124	108	89	28.2%
Treatment	137	130	113	17.5%
Total	261	238	202	22.6%

Table 2. Baseline characteristics

	Control (N=89)	Treatment (N=113)	Diff. / p-value
	(1)	(2)	(3)=(1)-(2)
A. Student characteristics			
Female Student	0.58	0.56	0.03 / 0.699
	(0.02)	(0.07)	
Student Age	9.01	9.02	-0.01 / 0.947
	(0.08)	(0.06)	
Travel Time to School	11.29	9.63	1.66 / 0.279
	(1.26)	(0.87)	
Had Breakfast	0.93	0.88	0.05 / 0.148
	(0.02)	(0.03)	
Days Attended Last Week	4.82	4.67	0.14 / 0.136
	(0.04)	(0.08)	
B. Student achievement			
English	-0.03	0.10	-0.14 / 0.422
	(0.13)	(0.12)	
Math	0.08	0.06	0.02 / 0.929
	(0.17)	(0.12)	
C. Household characteristics			
Household Size	7.34	7.29	0.05 / 0.894
	(0.24)	(0.25)	
Asset index	0.32	-0.29	0.61 / 0.001
	(0.08)	(0.12)	
Mother can Read	0.99	0.96	0.03 / 0.128
	(0.01)	(0.01)	
Father can Read	0.96	0.93	0.03 / 0.258
	(0.02)	(0.02)	

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. Sample sizes are smaller for some variables due to item-specific missing data. Standard errors clustered at the school level. Standard errors clustered at the school-level in parentheses.
[(a)] English and/or te-Kiribati.
* p<.1 ** p<.05 *** p<.01

Table 3. Regressions results

	Endline only	Controls	IPW
A. English			
Treatment effect	0.305	0.261	0.282
	(0.194)	(0.106)	(0.104)
	[0.152]	[0.099]	[0.083]
Baseline score		0.880	0.848
		(0.046)	(0.047)
N	201	201	198
B. Math			
Treatment effect	0.119	0.270	0.287
	(0.242)	(0.113)	(0.106)
	[0.692]	[0.101]	[0.087]
Baseline score		0.894	0.903
		(0.042)	(0.041)
N	202	202	199

Note: All models in the second and third column include the value of the outcome of interest at baseline, island fixed effects, children's age and gender, the asset index, and binary variables for parental literacy (one for the mother's and another for the father's) and for whether they child had breakfast that day. Standard errors clustered at the school level in parentheses. We also report p-values from wild bootstrap clustering in square brackets, using the Stata command boottest (Roodman, MacKinnon, Nielsen, & Webb, 2019).

Table 4. Lee bounds for treatment effects

	English	Math	English	Math
	(1)	(2)	(3)	(4)
Lower	0.045	-0.033	0.134	0.224*
	(0.089)	(0.107)	(0.112)	(0.109)
Upper	0.313***	0.268*	0.310***	0.335***
	(0.088)	(0.112)	(0.091)	(0.100)
Bounds tighten using island indicators			X	X

Note: The dependent variable is the difference between endline and baseline scores, to adjust for baseline differences. We use the Stata command leebounds (Tauchmann, 2014).
* p<.1 ** p<.05 *** p<.01

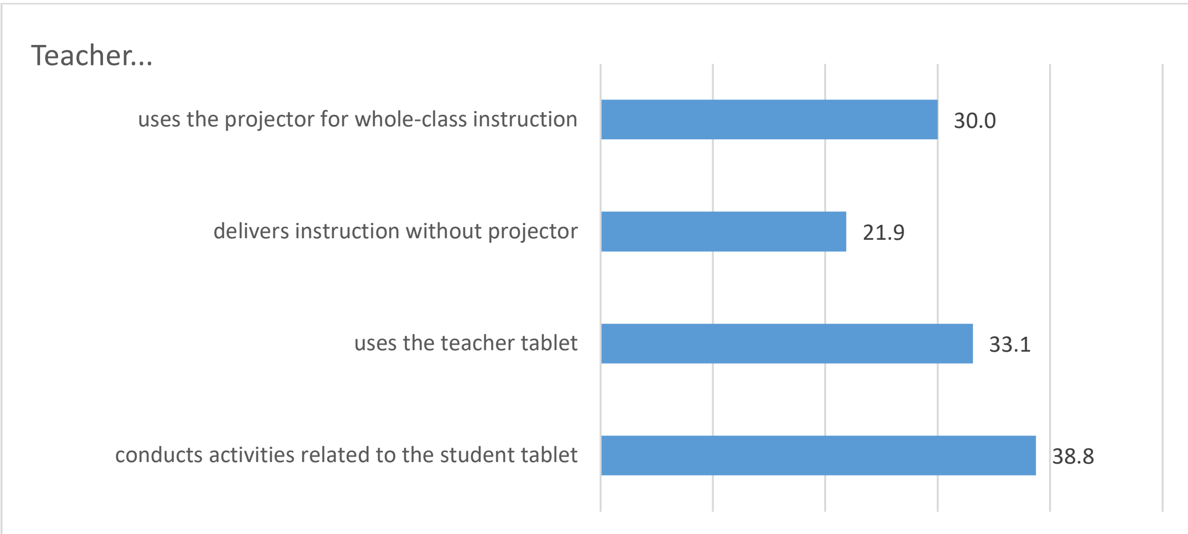
Table 5. Regressions results by student gender

	English	Math
Treatment	0.285	0.271
	(0.149)	(0.085)
	[0.150]	[0.016]
Girl/treatment interaction	-0.005	0.028
	(0.157)	(0.102)
	[0.962]	[0.762]
N	198	199

Note: All models include the value of the outcome of interest at baseline, island fixed effects, children's age and gender, the asset index, and binary variables for parental literacy (one for the mother's and another for the father's) and for whether they child had breakfast that day. Standard errors clustered at the school level in parentheses. We also report p-values from wild bootstrap clustering in square brackets, using the Stata command boottest (Roodman et al., 2019).
* p<.1 ** p<.05 *** p<.01

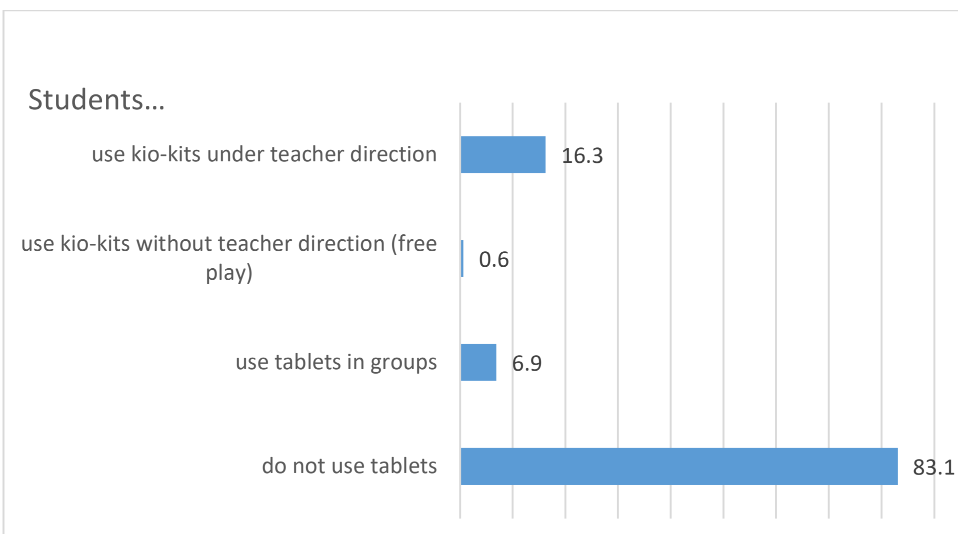
Figures

Figure 1. Teacher activities during class (%)



Note: More than one activity can be selected/taking place at the same time.
Source: Endline classroom observations.

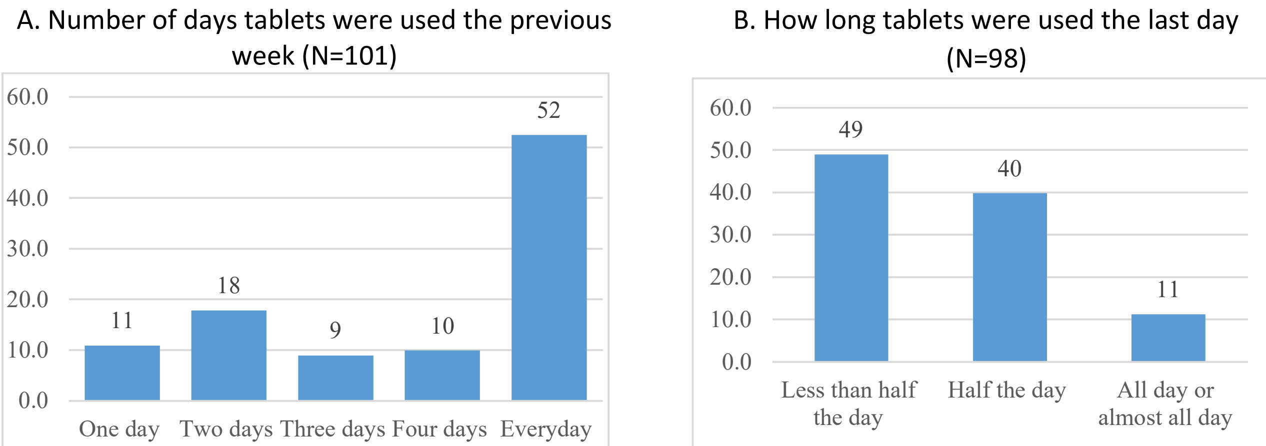
Figure 2. Students' activities during class (%)



Note: More than one activity can be selected/taking place at the same time (except for "do not use tablets").

Source: Endline classroom observations.

Figure 3. Use of tablets as reported by students (% of treatment group)



Source: Endline student survey. Includes panel students only.