Abstract Title Page.

Title: Conditions for the Effectiveness of a Tablet-Based Algebra Program

Authors and Affiliations:

Andrew P. Jaciw
Megan Toby
Boya Ma

Empirical Education Inc.

Abstract Body

Background / Context:
Tablets such as the iPad represent the next iteration of technologies that hold promise to facilitate learning, particularly in mathematics. In the case of algebra, tablets have the potential to bring the curriculum to life by easily linking to supporting materials and they allow an interactive experience whereby manipulation of one type of representation maps onto changes in another. Such applications are difficult to achieve using print editions.

Although it is easy to imagine how technologies might facilitate learning, there is also the possibility that with poor implementation they do more harm then good. Evidence of the effectiveness of computer-based technologies in the classroom is mixed. Electronic tablets may be especially effective if used under the right conditions, including in situations where teachers combine successfully their pedagogical, content and technological expertise to produce successful implementation (Hofer and Swan, 2006).

This study examines the effectiveness of a specific tablet-based algebra program that contains the complete content of the print edition of the program, as well as a series of interactive tools including an equation graphing function, a feature that allows exploration of quadratics and linear relationships, and an activity involving algebra tiles. These tools allow students to manipulate variables and see the results. The technology includes other interactive features as well, such as scratch pads, and prompts for review and practice. The counterfactual is the print edition of the program.

This work assesses both overall effectiveness, as well as impact under conditions of strong implementation. We take advantage of the block randomized design to assess the impact for blocks where implementation is singled out as exceptional.

This paper links to the conference theme, *Effective Partnerships: Linking Practice and Research*, because observations of practitioners and their knowledge of local contexts, independent of the initial research design, led to a reanalysis with a focus on a specific site, yielding a very different conclusion than if we had proceeded without this information. Close collaboration with the practitioner mattered to the study focus and conclusions.

Purpose / Objective / Research Question / Focus of Study:
The original purpose of the study was to assess the overall impact of the tablet-based algebra program on middle school students' achievement in algebra and levels of motivation for learning, and to assess the differential impact by levels of the pretest score and English proficiency. Before the outcome data were analyzed, the program developer identified the participating school from one of the districts, Riverside Unified, as an environment that was especially conducive to successful implementation of new technologies. For example, it had been an early adopter of a student laptop program and was among the first schools to install interactive whiteboards. This naturally led to the question of whether an impact would be observed for this particular block or mini-experiment, given its unique context. Therefore in addition to assessing impacts for the study sample overall, as an exploratory step we assess impact for Riverside and discuss the implications of differences in results.

Setting:
The participating districts were: (1) Long Beach Unified, (2) Riverside Unified, (3) Fresno Unified, and (4) San Francisco Unified. The experiment was conducted over the 2011-2012 school year.

Population / Participants / Subjects:
Across the four districts six schools and 11 teachers participated. In the control group there were 23 sections of Algebra 1 and 664 students on the fall roster. In the treatment group there were 11 sections of Algebra 1 (one per teacher) and 334 students on the fall roster. Riverside had two teachers with seven control sections with 197 students and two treatment sections with 64 students.

Intervention / Program / Practice:
The tablet-based algebra program is an application for the Apple iPad that contains the complete content of the print edition of the program. In addition, the application provides interactive lessons, explanations, quizzes, and problem solving. It comes preloaded with the 300+ videos that are available online to students using the print version of the text. The application contains a variety of interactive tools that allow students to manipulate variables and see the results. The note-taking feature allows students to type in notes, color code for organization, and leave themselves recorded voice messages. By touching a vocabulary word within a lesson, students are brought to the glossary where they are provided with a definition for the term. Preprogrammed quizzes test specific skills before a student begins a chapter and are accompanied by a scratchpad to help with calculations or writing notes. Students are prompted to review, practice, and retest skills. Icons on the sidebar provide tips and links for support, such as view-in-motion explanations or videos that address concepts from a different approach. The application also contains a Search function.
Control: The control materials consist of the print edition of the program and online access to videos. Each lesson within the textbook includes levels of skill development for differentiated instruction. The print text also provides quizzes. Participating classrooms received new sets of the textbook at the beginning of the school year.

Research Design:
The study is a group randomized trial, with sections of the Algebra 1 course randomized within each of 11 teachers within each of four school districts in California. One section in each teacher was randomized to treatment. There were 11 treatment sections and 23 control sections. We assessed impacts on (1) the California Standards Test in math, (2) the Holt McDougal Algebra 1 End of Course Assessment, and (3) the combined score from the Motivated Strategies for Learning Questionnaire, which consists of five subscales: self-efficacy, intrinsic value, test anxiety, cognitive strategy use and self-regulation. To assess implementation, we surveyed teachers on nine occasions. We included questions about time spent using the tablet in instruction, and the time spent by students using the tablet. Because individual teachers were implementing the treatment and control curricula, we also asked about possible contamination. Because randomization was blocked by teacher, the two teachers and nine sections in Riverside constituted a very small, yet independent RCT.

Statistical, Measurement, or Econometric Model / Analysis:
Our impact models included a dummy variable to indicate assignment status. The covariates included the pretest, and dummy variables indicating English learner status, disability status, gender, and ethnicity. A random effect was included at the section level, which is the unit of randomization. To reflect the randomized block design we included dummy variables for teachers. Given the multilevel nature of the experiment we used HLM (Raudenbush and Bryk, 2002) to estimate impact estimates. Missing values for the covariates were addressed using the dummy variable method. With this approach, for each of the covariates that is included in the model, a dummy variable was created. This variable was assigned a value of one if the value of the variable was missing for any student, and zero otherwise (Puma, Olsen, Bell, & Price, 2009). When student achievement outcomes (posttests) were missing, we used listwise deletion and simply dropped the observation from the analysis. Models for assessing differential impact were like those used to assess overall impact, but included a term for the interaction between treatment and the potential moderator. Cases with missing values for the moderator were eliminated. The moderating effects were examined one at a time (i.e., we did not use a single model with all interactions included.)

Findings / Results:
Attrition and Equivalence: Thirty-four sections were randomized. None were lost from the analysis of impact on CST, or impact on the student attitude questionnaire. Three out of 11 treatment sections and seven out of 23 control sections were unavailable for analysis of impact on the end of course assessment. Most of this loss is attributable to two teachers not providing data for their students in both conditions. The covariates were individually balanced across conditions for the overall analytic samples and for the Riverside subsample.

Implementation:
Table 1 (Appendix B) shows the rank ordering of time in minutes per week spent using the tablet in instruction for the 11 sections. The median level is 22.50 minutes per week. The two sections from Riverside spent the most time, with averages of 73.75 and 85.63 minutes. Table 2 (Appendix B) shows the rank ordering of minutes spent per week by students using the tablet in the class. The median level is 41.88. The two sections from Riverside were ranked second and third highest, with 97.50 and 104.38 minutes. A non-Riverside section reported the highest amount of time spent by students using the tablet, with a weekly average of 167.50 minutes.

Next we consider results of the impact analyses for the overall sample, and then separately for Riverside and the sample excluding Riverside. For the overall sample we ran a wider range of analyses. Analysis of Riverside and the sample excluding Riverside was limited to impact on CST.

Overall sample: There was no impact on performance on CST (p=.52) or on the end of course assessment (p=.90). We have some confidence of a positive impact on the strategies for learning questionnaire (p=.07). There was no differential impact on CST by level of pretest (p=.99), or on the end of course assessment by level of pretest (p=.97), or on attitude towards math based on the pretest (p=.36). Impact on CST was not moderated by English proficiency (p=.38). We have some confidence in a differential impact on end of course achievement with a less favorable

effect for English non-proficient (p=.09) as well as a differential effect on student attitude, with a less favorable result for English non-proficient (p=.09).
Riverside: The impact of the treatment on CST performance in Riverside was 11.95 scale score units (*p* = .01)
Non-Riverside: The impact for the sample excluding Riverside was -0.28 scale score units (*p* = .95).
Plausible rival explanations for the Riverside effect: The result for Riverside is based on a small number of blocks (teachers) therefore it is easy to get imbalance between conditions at the section level on observed or unobserved factors that influence performance. Also, with only two teachers in the analysis, characteristics unique to those teachers may be interacting with treatment to produce the effects – they may be exceptional in their use of the program. However, we emphasize that the result at Riverside is experimental, with randomization of sections within teachers within the site; therefore, if the two teachers are exceptional, this benefit would be conferred to their sections in both conditions. The exceptionality would therefore have to be in their use of the program, and this becomes a point of fundamental interest: what is it about these teachers and/or their school context that allows them to make exceptional use of the program that translates into a positive impact? We looked for imbalance between Riverside and the rest of the sample on background characteristics to see if it may explain the difference that we observed in the impact. There was a difference in the proportion of English speakers, ethnicity, and average pretest. However, for the overall sample, none of these variables moderated the treatment effect, therefore they do not explain the differential impact observed. There remains the possibility that students in Riverside differ from the rest on unobserved factors that interact with treatment thereby accounting for the differential impact observed. We conducted several sensitivity checks to examine whether the benchmark result for Riverside holds up to small variations in the analytic approach. All results that were based on alternative multilevel models were consistent with the main result.

Summary and Conclusions:
The randomized trial serves as a case where a gross impact estimate, though unbiased as a consequence of randomization, obscures what may be the more interesting and productive finding. Consequences of the context, though recognized post-randomization, allowed for adaptation of the analysis plan to accommodate a separate analysis for a setting where above average implementation was conjectured to produce above-average impact. The conjecture was borne out. This serves as a caution that the finding of no impact for the study sample as a whole would be misleading and is not generalizeable to particular situations. Still, because the study was not initially designed to assess the variation in impact that is associated with stronger implementation, and because the subsample of strong implementers was very small, the result must be considered exploratory. Further work should invest in articulating theory concerning the conditions for maximizing impact given what we know about how the program works. For example, we observe that the impact for Riverside coincides with above average use of the tablet by teachers and students – is it just the difference in time spent that is making the difference for the outcome, or are there also qualitative differences in implementation? Theory can address this point. Seeing whether impact varies as predicted on the basis of this theory would allow for more circumscribed but also more accurate inferences. This should assist with making program improvements by informing us about the factors that make a difference for program success.

Appendices.

Appendix A. References.

Bebell, D., and O'Dwyer, L.M. (2010). Educational outcomes and research from 1:1 computing settings. *Journal of Technology, Learning, and Assessment, 9,* n1 Jan 2010 (EJ873675)

Hofer, M., & Swan, K. O. (2006). Technological pedagogical content knowledge in action: A case study of a middle school digital documentary project. *Journal of Research on Technology in Education, 41*, 179-200.

Puma, M. J., Olsen, R. B., Bell, S. H., & Price, C. (2009). *What to do when data are missing in group (cluster) randomized controlled trials.* (NCEE 2009-0049). Washington, DC: U.S. Department of Education.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods, Second Edition*. Newbury Park, CA: Sage.

Appendix B. Tables and Figures.

Table 1. Rank Ordering of Time in Minutes per Week Spent Using the Tablet in Instruction

Teacher ID	District	Average time	Weeks with valid data
7	Non-Riverside	0	7
5	Non-Riverside	5.63	8
6	Non-Riverside	6.25	8
2	Non-Riverside	8.57	7
8	Non-Riverside	15.00	1
1	Non-Riverside	22.50	8
3	Non-Riverside	27.50	4
4	Non-Riverside	41.67	6
11	Non-Riverside	61.88	8
10	Riverside	73.75	8
9	Riverside	85.63	8

Table 2. Rank Ordering of Minutes Spent per Week by Their Students Using the Tablet in the Class

Teacher ID	District	Average time	Weeks with valid data
2	Non-Riverside	8.57	7
7	Non-Riverside	18.57	7
1	Non-Riverside	30.00	8
5	Non-Riverside	32.50	8
8	Non-Riverside	40.00	1
6	Non-Riverside	41.88	8
4	Non-Riverside	50.00	6
3	Non-Riverside	67.50	4
10	Riverside	97.50	8
9	Riverside	104.38	8
11	Non-Riverside	167.50	8