Pedagogical Change with Complementary Use of Technology: Evidence from a Clustered Randomized Control Trial[1]

Samuel Berlinski

Inter-American Development Bank

Matias Busso

Inter-American Development Bank

February 2013

Abstract

In this paper we report the results of an experiment with seventh grade Costa Rican children designed to improve their ability to think, reason, argument and communicate using mathematics. We created a structured pedagogical intervention that allowed students the opportunity for a more active role in the classroom. Our multi-treatment experiment also explores a gradient of technologies that are complements to this pedagogical approach and vary widely in cost. We find that the control group learned significantly more than any of our treatment arms. Moreover, technology had a rather negative impact at achieving our objective. We provide evidence that the experiment was internally valid and implemented with high fidelity.

Preliminary and incomplete –Please, do not cite or circulate

JEL classification: C93, I21, I28, O32

Keywords: Education, Technology, Curricular Reform, Teacher Training, Laptops, Interactive Whiteboards, Computer Labs, Mathematics, Field Experiments.

[1] Acknowledgements: This project is the result of a collaborative effort involving many people. In particular, we would like to thank Horacio Alvarez Marinelli, Floria Arias, Moritz Bilagher, Maria Eugenia Bujanda, Elsie Campos, Marco Cordero, Ulises Cordero, Julian Cristia, Maria Antonieta Diaz, Alvaro Gamboa, Torie Gorges, Mauricio Holtz, Teresa Lara-Meloy, Richard Mayer, Jeremy Roschelle, and Magaly Zuñiga. We also thank seminar participants at EDePo (IFS), and Fundación Omar Dengo for helpful comments. Juliana Chen Perazza and Rosa Vidarte provided excellent research assistance. We gratefully acknowledge financial help from the Inter-American Development Bank and Fundación Costa Rica USA. The views expressed herein are those of the authors and should not be attributed to the Inter-American Development Bank.

1. Introduction

In this paper we report the results of an experiment devised to affect the way mathematics is taught and learned in Costa Rican secondary schools. The objective was to create a scalable intervention that will allow students to achieve mathematical competence. This is to say, the student ability to think, reason, argument and communicate using mathematics. This concept is prevalent in the design of PISA examinations (OECD, 2009) and in curricular reforms in many countries (including Costa Rica and the US).

Many practitioners believe that in order to achieve these skills we have to switch from the traditional lecture-style approach (i.e., one where the teacher is the center of the class and the student is a passive receiver) to one where the students take a more active role in the generation of knowledge.[2]

We created a pedagogical intervention designed to give students a more active role in the classroom. A key aspect of this change relies in providing them with guided opportunities to explore and discover. In mathematics, a potentially important lever in this process is the use of technology. Our experiment explores a gradient of technologies that are complements to this pedagogical approach and vary widely in cost.

In economics there is a growing interest in studying what happens in the classroom and to identify successful pedagogical approaches. For example, Dobbie and Fryer (2012) peep into the black-box of 39 charter schools and correlate data on school practices with credible estimates of school's effectiveness. Fryer (2012) looks at the effect of injecting successful charter school strategies into traditional public schools. Machin and McNally (2008) evaluate the reading and overall English attainment of a national pedagogical strategy designed to raise standards of literacy in primary schools by improving the quality of teaching through more focused literacy work.

There are also a number of rigorous evaluations in economics that measure the effects on student learning of providing classroom resources to schools such as flipcharts (Glewwe et al., 2004), textbooks (Glewwe et al., 2009), libraries (Borkum et al., 2009) and even laptop computers (Cristia et al., 2012). The impacts of these interventions are in the best case scenario modest. A careful reading of this literature suggests that the failure of these resources to leverage student learning it is either because they do not address curricular objectives, student needs or teachers are provided with limited training on how to use these additional inputs effectively.

[2]There is surprisingly little empirical evidence on this issue. A recent report of the National Mathematics Advisory Panel on instructional practices in mathematics concludes: "For none of the areas examined did the Task Group find sufficiently strong and comprehensive bodies of research to support all-inclusive policy recommendations of any of the practices addressed." (Gersten (2008), page 6-189). Among the practices evaluated the panel looked at the use of teacher center versus student center approaches and the use of technology in the classroom.

We designed and implemented an experiment with seventh grade students in Costa Rica that blends a modern curricular approach with technology for teaching geometry (one of three units of the seventh grade program or about three months of teaching). We randomly assigned the 85 participant schools in this experiment to one of five conditions: (1) Status-quo (i.e., control); (2) New curriculum design; (3) New curriculum design and an interactive whiteboard; (4) New curriculum design and a computer lab; (5) New curriculum design and a laptop for every child in the classroom. All students (18,000) and teachers (190) in the seventh grade of these schools participated in the experiment.

The technologies we chose for this experiment represent a set of relevant options for Costa Rica (as well as to other middle and high income countries) and vary widely on their costs. For example, in our case the cost of having one laptop for every student in the mathematics classroom is more than three times larger than equipping a classroom with an interactive-board. Buying laptops for every student in the primary school in Costa Rica will be equivalent to a 20 percent permanent increase in the per-student educational expenditure (Berlinski et al., 2011).

Lampert (1990) summarizes the recommendations of the US National Councils 1989 mathematics reform documents as follow[3]: "Mathematics students should be making conjectures, abstracting mathematical properties, explaining their reasoning, validating their assertions, and discussing and questioning their own thinking and the thinking of others" (p. 33). Concrete manipulatives are a natural pedagogical device to facilitate exploration and argumentation in the mathematics classroom. In fact, their use has a long tradition and some evidence of success (National Research Council, 2001). Technology can play an important role in helping students to manipulate figures in order to make conjectures, explain their reasoning and interact with others (Clements, 1999).

We commissioned the design of pedagogical material for this project to local experts supervised by a team from a leading international education academic organization. The team decided that given the objective of achieving mathematical competence, traditional classroom routines had to be changed. In particular, the idea was to include a guided exploration of the topics before a more directed formulation of concepts, rather than the more typical lecture on the concepts, followed by an opportunity to practice on typical problems. In the guided exploration students were asked to informally conjecture about relationships, definitions and concepts. Exactly what this part looked like varied by condition.

In order to support teachers and students in this transition and (with an eye directed at improving fidelity of implementation as well) we created a teachers' manual and a students' workbook (one for every modality of intervention but none for the control). Technology was introduced through a set of applets created in GeoGebra (an open source dynamic mathematics software for learning and teaching) that will help student and teachers explore the key concepts of the unit.

[3] Cited in Gersten et al. (2008) page 6-15.

In the one-to-one and computer lab conditions, students directly manipulated the Geogebra files. In the interactive whiteboard condition, students shared a dynamic geometry experience on the whiteboard, and the teacher determined whether students had direct access to the technology. In the non-technology condition, students used the guided exploration portion of the class to conjecture with hands-on manipulatives.

In coordination with our local partners, we ensured that all the technology was in place at the time of the implementation and suitable technical support was provided to guarantee that it was functional during the experiment. Teachers received 40 hours of on-site and distance training with virtual support achieving 95 participation rate. All the teachers in treatment arms received a laptop computer and a teacher workbook. All the students in the seventh grade (with the exception of the control group) received a student workbook with their name tag.

Collaborating with the local and international experts, we designed a psychometrically valid test of geometry to measure the impact of this intervention. The objective of the test was not only to measure the content knowledge of the students but also their mastery of higher-order geometric practices that require, for example, that students pick, compare, justify or refute conjectures and propositions. Before the start of the experiment, we tested all the students in their general knowledge of sixth grade mathematics using a standardized international test prepared and administered under the supervision of the Latin American Laboratory for Assessment of the Quality of Education (LLECE) of UNESCO.

There are a number recent authoritative meta-analyses (e.g., Gersten et al., (2008) and Cheung and Slavin (2011)) summarizing high-quality studies of educational interventions that involve the use technology that help putting our study in context. The quality of a study in these reviews is related to the rule assigning units to treatment/control status, likeness of treatment and control groups, attrition, outcome of interest, duration, number of participants units and scalability.[4]

Cheung and Slavin (2011) identified 75 studies for K-12 published or unpublished in English from 1980 to 2010. Of those, only 25 were RCTs. The latter had an average of 9 schools (only 2 studies have more than 30 schools) and 780 students (the largest study has 3,136 students) participating from these experiments. None of the studies reported were performed in a developing country and no information on costs was reported. There are no high-quality studies reported either in Gersten et al. (2008) or Cheung and Slavin (2011) pertaining to the use of interactive whiteboards. The latter study almost exclusively finds papers that look at the use of

[4] For example, Cheung and Slavin (2011) include only studies where: (1) Assignment to treatment units is random or there is matching with appropriate adjustments for any pretest differences; (2) Pretest data is available or there are at least 30 units assigned randomly to treatment and controls with no indication of initial inequalities in basic demographics; (3) The outcome of interest is a quantitative measure of mathematics knowledge. Measures made by the experimenter are accepted if they are comprehensive measures of mathematics which would be fair to the control group; (4) The study last for at least 12 weeks; (5) Studies had to have at least two teachers in each treatment group; (6) Programs had to be replicable in realistic school settings.

technology as a supplement to teaching in the form of drills and practice. It reports modest gains of around 10 percent of a standard deviation of this use of technology.

We find that the control group learned significantly more than any of the four intervention groups. The students using the new curriculum without technology learned about 20 percent less than the status-quo. Learning was around 40 percent of a standard deviation lower in one laptop per student schools in comparison to control establishments. In the race between the three technologies (i.e., keeping the pedagogical approach constant) the interactive-board is the one that fairs better. In fact, there seems to be a negative technology gradient.

Randomization yielded groups with similar observable characteristics. The experiment was implemented with high fidelity. Materials and equipment were distributed where and when expected. They remain functional during the experiment. Teachers and students made use of their respective manuals. Indeed, there were significant changes in class dynamics with more participation from students. To a large extent teachers in the treatment arms seem open to the innovations we introduced in the classroom.

So what happened? The results of the paper can be understood by thinking about the effect of technological change on productivity. The introduction of new technologies such as the micro-computer and the dynamo (see David, 1990) did not lead to an automatic increase in productivity. In fact, there was much surprise with the sluggish growth of productivity in the western world.

Helpman and Rangel (1999) provide an interesting explanation for this phenomenon. If productivity with a given technology increases with experience, then the switch to a new technology may lead to falls in output in the short run if that experience is not completely transferable to the new technology even when for the same level of experience (accumulated with the new technology) productivity is higher. Put it in other words, the switch to a new pedagogical approach and the use of technology may lead to a short to medium run fall in achievement as teachers' lose specific human capital in the transitional phase.

We proceed as follows: in Section 2 we present the distinctive features of our research design, in Section 3 we explain the data collection process, in section 4 we discuss the empirical strategy, in Section 5 we present our research sample and discuss the internal validity of the experiment, in Section 6 we show our results on fidelity of implementation, test scores and class dynamics as well as some robustness checks. Section 7 discusses the possible mechanisms behind our results. In Section 8 we conclude.

2. Research Design

2.1 Context

Costa Rica is a relatively small middle-income developing country. In 2011, 4,726,575 people lived in Costa Rica, GDP per capita (2005 PPP $) was $ 10,085 and the United Nations Human Development Index ranked it in the 69th place. The country boasts a long tradition of publicly provided education. Public education in Costa Rica is free and has been compulsory since 1870 with an adult literacy rate of 96 percent.

The educational system is divided in four levels: preschool education (ages 4-6); primary education (6 years); secondary education which has also two cycles, middle school (seventh, eighth and ninth grade) and high school (10th and 11th grade). Currently, education is free and compulsory from the last year of preschool to the end of middle school.

During 2012, the school year in middle school was 196 days long. In the seventh grade, students have six[5] mathematics lessons a week of 40 minutes each (many of those arranged in contiguous blocks). The school year is divided in three terms and the mathematics curriculum in the seventh grade covers: Integers, Geometry, and Rational Numbers. Unlike the mathematics lessons in primary schools, mathematics is taught by a specialized teacher. In their annual teaching plan, teachers assign one term to each topic.

The teaching of mathematics in the Costa Rican context is not very different from those of other secondary schools in low and middle income countries around the world. The mathematics class is characterized by lecture–style teaching where the teacher writes down a definition or procedure in the blackboard using a particular example. Students take notes, ask questions and practice what the teacher explained. To support the teacher in this endeavor he/she follows a commercial textbook of his/her choice and provides the students a long list of examples to practice.

The theorems and procedures are taken as given truths and the objective is to practice until the students achieve mastery in their use. The problems and procedures are not necessarily associated neither to the solution of a hypothetical, real life or scientific problem.

The interaction between students and teacher can be characterizes as follows: The teacher is responsible for explaining and clarifying doubts or difficulties the students have, he/she controls the beginning of the class, the pertinence and formalization of the students interventions, validates their reasoning, explanations or interventions. Students participate when they don't understand, and they contribute with fragmented phrases to statements made by their teachers, they listen to their teachers interventions and ignore those of other classmates.

[5] One of these lessons is assigned for revision, recovery of lost time and to support students that are lagging behind.

Costa Rica has a long tradition of introducing technology in schools (Zuñiga, 2003). The Educational Informatics Program created in 1988 (a joint effort between the Ministry of Education and Fundación Omar Dengo) is a national informatics program that attend students in preschool, primary school and middle school (since 2001). Technology is introduced in the school through computer laboratories[6] with the objective (among others) of promoting the development of logic thinking by using computers to solve problems and work in teams.

2.2 Experimental design

We started with a sample of 100 secondary schools with a minimum of 2 classes and a maximum of 12 classes of seventh grade math in 2011. In order to minimize implementation costs, we choose schools that were located in urban and semi-rural areas of Costa Rica and were easily accessible by roads. We invited schools to participate in the experiment during November 2011 and only the 85 schools that sign agreements to participate were included in the experiment.[7]

In order to assign schools to their experimental status we performed a blocked randomization. First, we ordered schools according to seventh grade enrollment[8] in 2011 and constructed (at random) ten bins of five schools and five bins of seven schools. In every bin schools were assigned to either one of the following 5 conditions: Control (20 schools), Curriculum (20 schools), Interactive-board (15 schools), Computer lab (15 schools) and One-to-one (15 schools). All seventh grade teachers and students in the school were assigned to their corresponding experimental status.

We notify the government and the schools of the lottery results on February 2 (at the beginning of the school year). We organized meetings with the schools to inform them how the experiment will be implemented. During these meeting head-teachers received a manual with the following information: description of the project, main objectives, grade and topic to be covered, teacher training and equipment to be assigned, responsibility over the equipment during and after the project, verification of school infrastructure, logistics for the distribution of equipment, data collection, relevant information for parents and teachers, a calendar for the experiment implementation and contact details.

[6] Although some small rural schools receive equipment for classroom use.
[7] The Superior Council of Education of Costa Rica approved the experiment by resolution CSE-SG-168-2012.We obtained an IRB from Fundacion Omar Dengo.
[8] Ordering schools by enrollment was useful to reduce the cost of buying technology.

2.3 The Deployment of equipment

In schools assigned to the one-to-one status, classrooms[9] were equipped with 30 Classmates Magalhães laptops[10], two laptop carts to store and charge the laptops (while not in use), one LCD projector, one desktop computer, and one router. The students have to sign in to the desktop server to login into the laptop. We installed Ubuntu (a Linux based open software) in all computers as their operating system and GeoGeobra[11] as their main mathematics package.

In schools assigned to the interactive-board status, classrooms[12] were equipped with one (IQ-Board PS) inter-active whiteboard, one desktop and one router. The interactive-board uses pressure sensing technology, this means that you can use your finger or any other writing object to move, click and operate the computer.

Schools assigned to the computer lab status either use their existing computer laboratories run by Fundacion Omar Dengo and if they have no laboratory installed we created a mobile facility providing: 15 Classmates Magalhães laptops[13], one laptop cart, one LCD projector, one desktop computer, and one router.

The schools were inspected by an engineer after the lottery results were announced. In coordination with the head-teacher, classrooms that fulfilled security and structural condition were chosen for the installation of the equipment. Minor adjustments (e.g., wires and sockets were changed, walls were fortified to support the whiteboard) were required in some schools and applied.

In the four treatment categories all teachers received a laptop computer that should be returned to the school at the end of the school year.

The costs of deploying technology in any of the three forms are quite significant and vary considerably by technology. Table 1 reports three measures: the initial investment cost per student, the recurrent cost per student and the total cost of ownership (TCO) per student. Costs are calculated on the assumption that the program is scaled over the five years of secondary education.

The initial investments are capital costs for acquisitions and installation of items at the outset of the program. They are typically incurred in lump sum form. On the other hand, recurrent

[9]In schools with 6 sections of mathematics or less we equipped one classroom and for schools with seven or more sections we equipped two classrooms.
[10]The computers are particularly suitable for school environments: they are resistant to falls, water spills and do not have sharp edges. They have battery life of 4 hours, 1GB RAM, and 160 Gb HDD.
[11]GeoGebra (http://www.geogebra.org/) is a free and open source multi-platform dynamic mathematics software for all levels of education that joins geometry, algebra, tables, graphing, statistics and calculus in an easy-to-use package.
[12]In schools with 6 sections of mathematics or less we equipped them with one interactive-board and for schools with seven or more sections we equipped them with two.
[13]We also equipped 4 schools that already had labs with an additional lab because they were capacity constraints.

expenditures are monthly ongoing costs incurred over the lifetime of the project to effectively deploy the program. The total cost of one-to-one is three times higher than the interactive-board and the computer laboratory is twice as high.

2.4 The design and production of the pedagogical material[14]

We commissioned the design of pedagogical material for this intervention to local experts from Fundacion Omar Dengo and Universidad de Costa Rica advised by a team of international experts from the Center for Technology in Learning (CTL) at SRI-International.

The objective of the material is for students to achieve mathematical competence as defined by PISA (OECD, 2009) and the mathematics curricula of many countries including Costa Rica. Mathematical competence is understood by the student ability to think, reason, argument and communicate using mathematics. This requires that students can pose and solve mathematical problems and model mathematical situations using appropriate representations, symbols, tools and technology.

A lecture-teaching style where students are passive receivers of information does not encourage students' activities that are important for the construction of mathematical knowledge such as: exploration, verification, and communication of mathematical facts. These activities are important to develop the type of inductive and deductive reasoning that allows them to solve both scientific and everyday situations.

As part of the design of the experiment, we commissioned the production of a teacher manual as a way of accompanying teachers and improving the fidelity of implementation of the experiment. The manual was elaborated with a structure that would not demand significant time to study by the teachers. The material was gathered in 19 thematic sessions with themes that will be easily recognizable by teachers. [15] The sessions covered all the materials in the seventh grade Geometry curriculum of Costa Rica.

Each session had the same structure to help teachers with the use and understanding of the material. The sessions were classified in seven sections: (1) Topic Description, (2) Specific Objectives, (3) Materials, (4) Session length, (5) Presentation, (6) Body of the session and (7) End of session.[16]

[14] This sections draws from a report prepared by Arias and Zuñiga (2012a).

[15]The teacher manual for the technology interventions included three more sessions at the beginning to introduce teacher and students to the use of technology in the classroom.

[16]Taking into consideration the usual disruptions to the school calendar (e.g., strikes, school functions, and teacher absences), the duration of each session assumed that the real teaching time will be four lessons per week rather than six.

The body of the session was subdivided in three activities: construction, formulation and application of concepts. The first of these activities is the largest departure from the traditional classroom model. It relies on a predict-check-explain cycle to build a geometric concept. Ultimately, the understanding and explanation demand the use of a specific property. In contrast to the traditional lecture style, the students have a very active role in this process and the teachers a less controlling one.

In the formulation part the teacher is responsible for institutionalizing the knowledge established in the construction section. However, this does not mean that we expect the teacher to recite or copy mathematical results on the whiteboard. In fact, we introduced a clarify-formalize-validate/verify cycle with the students actively participating from this task.

Finally, the application part is the more akin to the usual geometry class. It includes some practice of the concepts that were studied and the conclusion of the session. Unlike in the typical class the manual does not offer a long list of exercises, the idea is that the mathematical work started with the construction part and this just provides an opportunity to consolidate what has been learned.

In order to standardized the work of the students and to provide support to the teachers in implementing this pedagogical approach we decided to create a student workbook as well. The student workbook has hands-on paper based activities and is identical in knowledge content to the teacher manual. The main difference is that the teacher manual has advice on how to proceed or motivate the students at different points in the class and this is not included in the student manual.

Even though the experiment introduces four different interventions (curriculum, one-to-one, computer lab, and interactive whiteboard), the class is structured around a single pedagogical design independent of technology. This is to say, learning is driven by the mathematical actions that are required and by the student-teacher role in producing these activities rather than by the technology. This is a key concept in our research design because by providing a common pedagogical approach for the experiment we are able to disentangle the role of the pedagogical setting from the contribution of technology.[17]

Of course, all this begs the question of how we planned the use of technology in this experiment. The use of technology in the mathematics classroom (like other manipulatives) contributes to learning because it allows time to be devoted to activities that are harder to do using only a blackboard and a textbook such as: grouping and classifying objects, establishing relations, visualizing generalizations and discovering properties.

[17] It also simplifies considerably the production of large amount of the pedagogical material.

However, the introduction of technological resources in the classroom can be disruptive. It changes the classroom routines for student and teachers which can demand the establishment of new rules of engagement. It may also requires substantive new knowledge from the teachers.

Keeping these hurdles in mind, the study considered a relatively simple approach to the use of technology. First, we choose GeoGebra, a software that teachers in Costa Rica were already familiar with. Second, rather than requiring teachers and students to program in GeoGebra, we developed a set of applets in which the students have a series of elements (or buttons) that can be used to manipulate geometrical objects. These manipulations are planned so as to put the students in the best possible position to visualize, explore, conjecture, and construct mathematical arguments.

In the teacher/student manual, technology is introduced in the construction and formulation phases of the lesson. In the construction phase, characterized by the predict-check-explain cycle, the GeoGebra applets are key to the check stage where students may explore with technology more freely. In the formulation phase, characterized by the clarify-formalize-validate cycle, the applets are used to systematically guide students through important cases or variations.

The GeoGeobra applets are the same in the three technology options what varies is the time of exposure that the children will naturally have and their ability to drive the exploration. Obviously, when students have individual laptops, they can use these to check predictions individually. In an interactive whiteboard setting, the teacher can perform the check phase.

The manual for the curriculum intervention proposes activities that use manipulatives such as images, paper, etc. rather than technology.

Beyond the validation received by experienced mathematic teachers in Costa Rica, the support from international experts of CLT at SRI, the material was also reviewed by those in charge of training the teachers and ultimately received the feedback from the teachers that participated from the training (about 45 days before starting the experiment). One manual was printed for each of the four conditions.[18] All students received a student workbook with their name tag before the beginning of the experiment.[19]

2.5 Teacher training[20]

The development of mathematical competence requires a learning environment where the student plays a commanding role. In comparison to lecture-style teaching, the emphasis of

[18] The manual for the computer lab condition was a combination of the one-to-one and the curriculum manual as we could not predict which sessions will occur in the lab and which will occur in the classroom.
[19] The student workbook had 150 pages, 146 pages, 246 pages and 183 pages for the interactive whiteboard, one-to-one, computer lab, and curriculum conditions respectively.
[20] This sections draws from a report prepared by Arias and Zuñiga (2012b).

classroom activity switches from transmitting information and performing numerous drills to creating opportunities for exploration, construction and mathematical argumentation.

Teaching to achieve mathematical competence requires a change in the traditional role of the Costa Rican mathematics teacher. Therefore, teacher training for this experiment aimed at affecting teachers' beliefs about how mathematics is taught and learned. In particular, we emphasized the importance of devoting class time for student exploration and the role of the teacher as a guide/mediator for the students in this process.

In particular, the training we offered to teachers in the experiment focused on:

- Giving teachers a first pass at the approach we use to develop mathematical competence.
- Familiarizing teachers with how to use the teachers' manual, the students' workbook and the GeoGebra applets. We placed particular emphasis in practicing didactic strategies and how to involve the use of applets in these activities.
- For those teachers randomized into a technology group, we put the teachers in contact with the corresponding technology.

In line with the supporting material we designed for the experiment, we assigned in training a leading role to the pedagogical use of technology rather than to the mastery of the technological instrument per se. Indeed, in a representative survey of teachers in Costa Rica teachers choose as their first priorities for professional development in relation to technology the methodology and didactic use of technology in the classroom rather than the use of internet, software and multimedia (FOD-MEP 2010).[21]

A total of 130 teachers participated from the full training program. From those that participated 115 received a certificate that provided professional points in the Civil Service Career system. Successful completion of the training was based on attendance, classwork and passing a final assignment. There were 16 teachers that did not attend the full training session at any point in time. These teachers were offered a recovery training session; 9 of them attended and 7 were absent.

The full-training session had 40 hours[22], divided in 4 weeks[23] with 10 hours on each week. From these 10 hours, 5 hours were allocated to on-site training and 5 hours to distance work supported

[21] We performed a diagnostic survey with 88 teachers in schools that receive technology in the planning phases of training. From them, 66.7% have received some training on the use of computers in teaching, 67% on the use of GeoGebra and 49.2% on the elaboration of applications for GeoGebra. The results of the diagnostics were important to reinforce the idea that given that many teachers had basic knowledge in the use of technology, training should emphasize the pedagogical approach.

[22] The results of the diagnostic survey highlighted the need of training teachers in general aspects of the use of UBUNTU and the use of interactive-whiteboards, given that 72.4% reply that they didn't know the operating system and 86.4% have never worked with an interactive-whiteboard before. We organized two additional sessions of 4 hours each to familiarize teachers with the operating system and the use of the interactive-whiteboards (for those randomized to that condition).

[23] Training occurred between the 9th of April and the 4th of May.

by a virtual classroom.[24] Training was organized by modality (i.e., curriculum, one-to-one, interactive-whiteboard, and computer lab) and delivered in two regional sites.

At the beginning of training, every teacher received their laptop, the corresponding teacher manual and a CD with the presentations for each training session, the GeoGebra applets, a manual for the use of the LCD projector and other complementary material.

2.6 The design of the assessment[25]

The research team developed an assessment to measure learning and to potentially distinguish the gains from the different conditions. The process started at the conclusion of the curriculum development phase. We followed the model of assessment development described in Shechtman et al. (2010) which is based on the following four activities: domain analysis, conceptual assessment framework, assessment task development, and assessment assembly.

In domain analysis, we determine which content and skills should be included in the assessment. Following an in depth analyses of the Costa Rican curriculum, the group of experts advising us determined that the focal content knowledge areas to be covered in the assessment. In addition to content, we outlined the two types of conceptual skills and practices that we expect students to develop during the Geometry unit.

Foundational skills and abilities (M1), typically covered by seventh grade curricula require that students use parts of definitions, classify figures according to given properties, locate parts of figures (points, segments, angles), and make simple calculations to find a missing side or angle. Higher-order geometric practices (M2) required that students pick, compare, justify or refute conjectures and propositions; deduce a third observation from two givens; formulate or justify a geometric argument, and generalize from one or more cases. Table 2, details the differences between these skills and practices.

We develop the actual assessment using the following procedures: (1) we created a pool of items for both foundational and higher-order content areas; (2) we rely on local and international experts to review the items; (3) we performed cognitive think-alouds of the higher-order items with students; (4) we pilot the test and use Item Response Theory to select items for the final assessment; and (5) a panel of experts review the final assessment form.

In preparation to develop assessment items, we built a blue print listing how the M1 and M2 practices would be evaluated. For grading simplicity and expedience, we decided that all items had to be multiple-choice. M1 items were drawn from existing Costa Rica tests and teacher

[24] Previous to the start of training, a virtual classroom was designed and installed to support the distance work of teachers. This is a moddle based site which contains forums and spaces to share resources, questions, and issues with the individual and group assignments.

[25] This sections draws from a report prepared by Lara-Meloy, et al (2012c).

materials. We were able to find a few items in the item pool to satisfy the M2 requirements and we also generated a pool of M2 items during an item camp.[26]

Two test forms of thirty-five items each were created from seventy candidate items; the two forms shared no items in common. Then, we classified questions by difficulty level: low, medium, and high. We created bins of 5 questions for each form and randomly assigned equal proportion of low, medium and high difficulty questions to each bin. We randomized all the choices for the multiple choice questions.

We piloted the test in eighth grade classrooms at two schools not participating in the project. We took the eighth grade rather than the seventh because most concepts in the exam are not taught until the seventh grade. Within each class, students were randomly assigned to one or the other form. There were a total of 400 students assigned to the test, and 200 taking one or the other form. On the day of the exam 184 and 181 took each exam respectively.

For the purposes of selecting items for the final instrument, we were interested in discarding items that 1) appeared to be too easy or too difficult for the target population, 2) did not correlate well with the total test scores, and/or 3) could be solved by "intelligent guessing" or other knowledge not related to the M1/M2 constructs. To provide a visual aid for examining the first two criteria (difficulty and item-test correlation) we subjected the test forms to Item Response Theory (IRT) analysis[27], and generated plots for each item's response curve (the probability of a correct response given a particular student proficiency).

In Figure 1, we plot the item characteristic curve for question 3 in exam 1 and question 34 in exam 2, the student scores are split into deciles representing the 10th, 20th, etc. percentiles of ability scores (θ), and the proportion of students answering the item correctly in each decile is plotted with a red dot. Ideally, these dots will track the item characteristic curve relatively well, indicating a good model fit.

The shape of the curve is determined by two parameters that roughly correspond to its center on the horizontal axis and the "steepness" of the curve. The difficulty parameter represents the location of the curve on the horizontal scale – technically it is the θ-coordinate where the curve crosses the line $Y = 0.5$. It represents the place on the ability axis where a student would have a 50% probability of answer this item correctly. Items with difficulty parameters in the high positive range would correspond to curves shifted to the right and therefore more difficult items (since a student would have to be located relatively high on the ability axis in order to have a 50% chance of answering the item).

[26] We developed four templates that were used in the item camp. We conjectured that using Geogebra might help students visualize geometric justifications, so we also wanted items with four geometric diagrams as response options.

[27] Technically, we modeled each item using a two-parameter logistic curve, conditional on a single overall student proficiency score.

A second parameter controls the "steepness" of the curve (one might think of it as the slope of the tangent line where the curve crosses Y = 0.5). Items with relatively steep curves are considered discriminating (and thus the parameter is called the discrimination parameter). Ideally, students with abilities slightly lower than that indicated by the item's difficulty would be highly unlikely to answer the item correctly, while those with abilities slightly higher than the item's difficulty would be very likely to answer the item correctly. The item is useful for discriminating between higher and lower ability students at that particular point on the ability spectrum.

Consider an item with a very shallow discrimination parameter, as question 34 in exam 2 in Figure 1. A student at the very low end of the ability distribution (θ=-3) has slightly more than a 30% chance of answering the item correctly, while a student at the high end of the distribution (θ=3) has just under a 40% chance of answering the item correctly. In other words, knowing whether a student answered this item correctly or not gives us very little information regarding their likely location on the ability scale.

After pilot testing, a summary document was prepared of each item's statistical characteristics, as well as the relevant sections from the think-aloud protocols. Each item was examined in turn, with the goal of selecting a final set of items out of the pool of seventy that were piloted. In many cases poorly discriminating items were discarded; in others it was determined that minor modifications could correct the obvious flaws in the item based on information from the think-aloud protocol. The final exam had 32 items in total.

4. Data

To assess the effect of the intervention on students learning and class dynamics we collected information on students, teachers, classrooms and schools at different times of the school year 2012 and also gathered administrative school information for 2011 and 2012.

The intervention affected nearly 18,000 students, 190 teachers in 85 schools. We collected data on all teachers and all schools. However, because of cost considerations it was unfeasible to collect data on all students. On average, each teacher was in charge of 3 sections (classrooms). Therefore, we decided to randomly select one section per teacher and collect data on that section only.

We administered a student survey before the beginning of the geometry unit to compare the distribution of characteristics of students in different treatment arms and determine whether the randomization has created comparable groups. The student survey was collected in the classroom and contained information on students' characteristics, their family and socio-economic background, and experience with computer use.

At baseline we administered a standardized achievement test[28] used in a regional study of educational attainment in Latin America in 2008. The test was prepared and administered under the supervision of the Latin American Laboratory for Assessment of the Quality of Education (LLECE) of UNESCO.

At the end of the geometry unit, we administered a second student survey and a geometry test. The objectives of this second survey were threefold. First, to determine whether or not there was differential attrition of students. Second, to measure treatment compliance and fidelity of implementation. Third, to build a set of scales that will allow us to measure the effect of the intervention on class dynamics and student's attitudes towards mathematics. We also administered a geometry test, discussed in Section 2.6, specifically designed to measure the main outcome of interest, that is, learning of seventh grade geometry.

Before and after the intervention took place we also collected teachers' surveys. At baseline we asked for information regarding background characteristics of the teachers and general information about their mathematics class. This information is used to verify comparability of treatment and control groups. In a second survey, after the intervention, we incorporated a set of questions to measure fidelity of implementation and changes in class dynamics and teaching practices.

Finally, we collected a rich dataset of administrative information about the school. This information included data on school location; computer equipment and infrastructure; size in terms of students, classrooms and teachers; repetition rate, etc. The information of 2011 was used mainly to stratify the sample when we randomized treatment. The information of 2012 was used to assess balance of characteristics and as a sampling universe to build our sample of students to be surveyed and tested.

5. Empirical strategy

We want to measure the causal effect of assignment to one of the four possible treatment conditions. Random assignment of schools to treatment/control status allows us to identify the average treatment on the treated by simply comparing the means of each of the four treatment groups with respect to the control group.

We estimate by ordinary least squares a set of models of the following form:

$$Y_{isj} = \alpha_0 + \sum_{k=1}^{4} \alpha_k T_{sj}^k + \delta_{sj} + \varepsilon_{isj}. \qquad (1)$$

[28] The Second Regional Comparative and Explanatory Study (SERCE).

In model (1), Y_{isj} is an outcome of interest (e.g., student performance on the geometry test or teachers' openness to innovation) of individual i, in strata s and in school j. T_{sj}^{k} is a dummy variable equal to one if the school j was assigned to treatment k={1,2,3,4}={new pedagogical approach, interactive whiteboard, computer lab, one-to-one}, δ_{sj} is a set of strata fixed effect, and ε_{isj} is a random specification error. The parameter of interest is $\alpha_k$the average treatment effect[29] (e.g., the average effect on student performance in the geometry test of being in a one-to-one school versus the status quo).

Of course, identification of the causal parameters of interest with randomization is an asymptotic property and our samples are finite. Thus, we also present results that control on observable characteristics (we also use equation (1) to show that these characteristics are balanced at the start of the experiment) with the added bonus that this might lead to smaller standard errors (i.e., more precise estimates). In other words, we estimate by ordinary least squares a set of models of the following form:

$$Y_{isj} = \alpha_0 + \sum_{k=1}^{4} \alpha_k T_{sj}^{k} + \beta X_{isj} + \delta_{sj} + \varepsilon_{isj}, \qquad (2)$$

where X_{isj} is a vector of student, teacher and school control variables.

In most cases, we estimate two specifications; one with and one without controls. Student controls include pre-treatment SAT score, dummies of gender, dummies of age, mother's education and number of books at home. Teacher control variables include gender, age, and years of experience. School control variables include number of students in seventh grade math, number of classrooms in seventh grade math, dummies of province, and a dummy variable that is equal to one if the school has a computer lab already available before treatment.

In all of models, we take into account the potential correlation between students' and teachers' performance and behavior by clustering the standard errors at the level of the school which is the unit of randomization. However, the standard error estimates are typically not sensitive to the level of clustering.

6. Research sample and internal validity

Our research sample covers all 85 schools and all 190 teachers that participated in the study. Table 3 shows its sample size. On average schools had 2.2 to 2.4 seventh grade math classes per school; except in the computer lab treatment arm schools were on average there were 1.8 classes in each school. For each teacher one complete class of eligible students was surveyed twice. All students were eligible except for those with some kind of known physical or cognitive disability.

[29] There is perfect compliance of the schools with the randomization status. Therefore, there is no practical distinction between treatment and intention to treat.

On average, each class had 25 students and about 1 percent of the students presented some disability.

Non-response rates were low. The first column of Table 4 presents the average and the standard deviation of several quality measures of our sample. The student post-treatment survey and geometry test had a non-response rate of 9.1 percent. Columns 2-5 show the regression coefficients and standard errors of a model described by equation (1) where the dependent variable is a dummy equal to one if the student was missing on the geometry test day and zero otherwise. The base category is the control group. There are no significant differences between treatment arms in the non-response to the geometry test. Column 6 presents the p-value of a joint Wald test of the null that all coefficients are equal. We cannot reject that null hypothesis.

The geometry test was administered on average about 14 days after the end of the second term, with very little differences on the dates schools of each treatment arm were visited. In each visit, the survey team would provide the teacher with numbered copies of the exam so that she would administer the test to the absent students within the following week. About 5.6 percent of the tests were administered by the teacher rather than by the survey team with no statistically significant difference between students in different treatment status.

The pre-treatment SAT test had a non-response rate of 21.1 percent with a larger non-response rate in the schools that received a computer lab. Nevertheless, we cannot reject the null that all treatment arms had similar non-response rates at baseline. Since the SAT is a control variable in our main econometric models, in order to avoid losing too many observations we imputed the missing SAT using mean class characteristics and in our models we add a dummy if the student had an imputed SAT. Our main results are robust to dropping observations with an imputed SAT in the sense that the point estimates do not change. However, the standard errors are a bit larger.

The non-response rate of teachers was 0.5 percent at baseline (only one teacher did not complete the survey) and about 3.2 percent (or six teachers) did not complete it after the intervention had ended. Again, there are no significant differences by treatment status.

The main advantage of a randomize control trial is that it guarantees that the only difference between units in treatment and control groups is the treatment they received. Treatment assignment is orthogonal to any individual characteristics that affect the main outcomes of interests (test scores and class dynamics). This has two implications. First, that the distributions of those characteristics before treatment ought to be the same between treatment groups and, second, that any difference in outcomes is caused by the treatment. With a large enough number of schools one would expect that the characteristics of students, teachers and schools in each treatment arms were identical. On the other hand, with a finite number of schools, any particular randomization could in principle render some difference in the distributions. To check the comparability of the five research groups, we directly compare the average students, teachers and schools in the treatment and control groups using pre-treatment characteristics.

Table 5 shows pre-treatment sample means and differences in those means across treatment groups. Overall these differences are small and usually not statistically significant. Half of the students in the research sample are female and on average they are approximately 13 years old. Two out of the sixty-eight coefficient in Table 5 are statistical significant. The students in the interactive whiteboard are younger and more likely to be female than those in the control group. However, these differences are very small. In the control group 50.4 percent are male and the average age is 12.96 while in the interactive whiteboard is 46.7 percent are male and the average age is 12.86.

The rest of the student's background characteristics are similar across treatment groups. Approximately 43 percent of the students have mothers with primary education and 40 percent with secondary education. On average, they report to have 3 books at home and 72 percent have access to computers at home which suggests familiarity with technology. For all these variables, we cannot reject that they are statistically the same across treatment groups.

The pre-treatment SAT score is perhaps the most important variable since it provides an indication of math knowledge acquired by these students before starting seventh grade. On average these students responded correctly 46.1 percent of the questions in the exam[30]. The differences between treatment arms are negligible.

The characteristics of teachers and schools are also very similar across treatment groups. On average, half of the teachers are female and they have about 11 years of experience. Schools have on average 220 students. The majority of them have some kind of computer lab and internet access. Repetition rate is 8.7 percent and 44.7 percent are suburban schools. We cannot reject the null that all of these variables are the same for all five treatment groups.

7. Results

We present in this section the core results of the paper. We proceed in three steps. First, we show evidence of fidelity of implementation of the intervention. Second, we estimate the average effects of the program on tests scores. We assess robustness of the results and also investigate treatment heterogeneity on tests scores. Finally, we study the effect of the treatment on a set of students and teachers scales that intend to measure changes in class dynamics and attitudes towards math.

At the beginning of the school year we announced to schools their treatment status. In the first panel of Table 6 we show that by the time of the announcement 83.6 percent of the teachers had already been assigned to their classes with no apparent differences in the different intervention groups.

[30] In the nationally representative sample of sixth grade students in 2008 the response rate was 49 percent.

We asked all teachers of the schools participating in the experiment to teach seventh grade geometry during the second term of the school year (the term suggested by the Ministry of Education). Only 3 classes (in the control group) out of 190 did not comply with this request and about 12.6 percent taught the introductory 4 units of geometry during the first term and then stopped at our request. Again, the majority of these classes are in the control group. If knowledge depreciates over time then this deviation should bias the results against the control group schools.

We have administrative documentation that attest that the experiment materials and equipment were received, installed and function at the time the intervention started. Table 6 corroborates that these materials were used and available in the classroom according to our design. In the second panel, we show that teachers in schools that received a computer lab and one-to-one intervention reported having computers in their mathematics class for their students and for themselves. Similarly, only teachers in the interactive whiteboards group reported having an interactive board available in their math class. Additionally, in all four treatment groups teachers reported having used the workbooks and manuals in the classroom. The third panel reports information from the student survey administered after treatment. The same facts observed in the teachers' surveys are confirmed by the students of these schools.

Table 7 present the main results of the paper. Each column shows the OLS estimation of equations (1) and (2). The dependent variable is the score in the geometry test (computed using the IRT parameters of the control group[31]) and then standardized using the mean and standard deviation of the control group.[32] Therefore the coefficients can be interpreted as the treatment effects in terms of that standard deviation. Column 1 controls only for the strata dummies. In consecutive columns we add: students' controls, teachers' controls, and schools' controls.

All treatment groups that changed the pedagogy to a student-center approach learned less geometry than the control group. The average treatment effect of the new curricular approach is a reduction in test scores of 17.1-20.2 standard deviations. The complementary effect of technology is ambiguous. On the one hand, the usage of interactive boards ameliorates the negative impact of the change in pedagogy but it does not offset it completely. Students in this group learned between 10 and 15.5 standard deviations less than those in the control group. On the other hand, the results for computer lab are very similar to the results of the new curriculum suggesting that it does not help much and the usage of computers on a one-to-one ratio worsens the results. Students in this group learned between 35.5 and 41.2 standard deviations less than students in the control group.

[31]The test has sound psychometric properties. For example, the scale reliability coefficient (Cronbach alpha) is 0.71. It also shows a correlation of 52% across schools with the SAT.

[32] Results are basically the same if instead of constructing a test score using IRT we use the percent of correct answers as a dependent variable.

In the bottom panel of the table we show p-values of pair-wise comparisons between different treatments. In each case, the null hypothesis is that the treatment effects are equal and the alternative is that one treatment effect is smaller than another. Basically, the one-to-one treatment effect is smaller than any of the other treatment effects at standard levels of significance. However, we cannot reject that interactive whiteboards, computer lab and new curriculum without technology have the same treatment effect.

Panel 1 of Figure 2 show the results for the M1 and the M2 test items. Recall that the M1 items are designed to measure foundational geometry abilities or basic concepts whereas the M2 items are designed to measure higher-order geometric practices; which a priori are easier to acquire using the new curricular approach. We find no differences in results when comparing the performance on items M1 or M2 to the overall performance. We speculate that in order for the students in the treatment group to outperform students in the control group on the M2 items they should have done at least similarly on the M1 items. But they did not.

We next provide evidence that the results are robust. The geometry unit is divided into five sections: introduction, measurement and classification of angles, relations between angles, triangles and quadrilaterals. In Panel 2 we remove, one at a time, all the items that belong to each of the five sections. If a given treatment group found the material on a section particularly difficult then we would find some reversal in the relative rank of the treatment effects when that section is not considered in the score. That is not the case. The relative performance of each group is maintained when each section is removed.

The items on the test also vary by difficulty. In Panel 3 we classify the thirty-two test items in eight groups of four items each according to the number of percent correct answers in the sample. We remove one group of items at a time and re-standardize the score. The relative performance on the test is the same when any given difficulty group is discarded.

We also check, in Panel 4, whether the results are driven by particular schools in the sample. To do this and still preserve the validity of the experiment we make use of the stratification of our research sample. We remove one strata at a time. Results are very stable which suggest that no individual school or strata drives the treatment effect and that there is not much heterogeneity on the treatment effect with school size.

Is there heterogeneity on other dimensions? In Table 8 we compute the treatment effects on tests score using different samples. In columns (1) and (2) we show that female students in the treatment group performed slightly worse than the male students. Columns (3) and (4) show that high ability students did significantly worse than low ability students. A possible explanation is that the traditional lecture-teaching style was geared towards the more able students. The intervention changed class dynamics assign relatively more emphasis to tasks that benefit lower achieving student.

Columns (5) and (6) divide the sample according to the experience of the teacher. Classes led by teachers with experience below the median tended to performed significantly worse than teachers with more experience. Columns (7) and (8) divide teachers according to their preferences for a student-centered or a teacher-centered approach to teaching.[33] We found no significant differences on student tests scores in this respect.

The experience of the teacher in the classroom is likely to be correlated with the ability to better mediate the new pedagogical approach. This better mediation might be because of a deeper knowledge of mathematics, a better handling of class dynamics, or the ability to better prepare classes. The results in Table 8 suggest that it is likely this mediation ability is a complementary factor of the new pedagogical approach and of the usage of technology in class. In other words, more experienced teachers are better mediators of the new pedagogical approach and thus their students obtained relatively better grades on the test.

Now, if the way the teachers mediated the new approach changed because of the intervention we should also observe changes in the class dynamics. To measure these changes we borrowed some well-established measurements from the psychology and education literature.[34] Throughout the questionnaires we asked students and teachers a series of questions that would later form several scales. Tables A2 and A3 list the questions used to build each scale. The questions pertaining to each scale were randomly mixed in the survey instrument. Students and teachers could give categorical answers of the type "strongly agree", "agree", etc. We aggregated those answers into scales using a maximum likelihood principal components estimator with one latent factor.[35] Scales were then standardized using mean and standard deviation of the control group.

Table 9 shows the results. In this table we aggregated the three technology groups into one group to gain power. In other words we use ordinary least squares to estimate a model similar to the

[33] We provided teachers with two scenarios. The first scenario read like this: "Ms. Gomez was leading her class in an animated way, asking questions that the students could answer quickly; based on the reading they had done the day before. After this review, Ms. Gomez taught the class new material, again using simple questions to keep students attentive and listening to what she said." The second scenario read like this: "Ms. Perez's class was also having a discussion, but many of the questions came from the students themselves. Instead of answering the questions directly, Ms. Perez guided the discussion so that the students themselves found the answers." We then asked four questions: (i) Which type of class discussion are you more comfortable having in class? (ii) Which type of discussion do you think most students prefer to have?; (iii) From which type of class discussion do you think students gain more knowledge?; From which type of class discussion do you think students gain more useful skills? The possible answers were: Definitely Ms. Gomez; I tend to like Ms. Gomez approach more; I tend to like Ms. Perez approach more; Definitely Ms. Perez. We aggregated the answers in a scale constructed using maximum likelihood principal components analysis. A lower value of the scale means that the teacher agrees more with Ms. Gomez (teacher-centered) approach and a higher value of the scale means that the teacher agrees more with Ms. Perez (student-centered) approach.

[34] We used questions and scales from the University of Chicago Consortium on School Research 2011 teacher and students surveys; the Patterns of Adaptive Learning Scales (PALS) from the University of Michigan (2000) and surveys questionnaires developed by SRI-International.

[35] Results were almost identical when building the scales via a polytomous IRT model.

one described in equations (1) and (2) but we let , T_j^k be a dummy variable equal to one if the school j received treatment k={1,2}={new pedagogical approach, technology}.

In the classes using the new curriculum with no technology and in the classes using the new curriculum with technology there was more classroom activity in the form of individual or group discussion as opposed to a lecture given by the teacher. This translated into more active learning, especially in the group of schools that received technology. Students in this latter group report doing more exercises in class for others to solve, having to explain concepts to the class and discuss possible solutions or arguments with other students. Interestingly, in both groups it seems that students like mathematics less than in the typical classroom. The treatment effects on other student scales are small and insignificant. Students are not more engaged with the class and do not feel that the teacher press them to exert more effort to learn mathematics.

Teachers also report changes in behavior. First, there is more dialogue among the teachers of the schools in treatment groups about the curriculum, teaching strategies, shared goals, and class mediation. Second, treated teachers were given access to new ideas through teacher training and professional development. Finally, teachers exposed to the new curriculum feel less in control of the class and of the learning experience of their students.

8. Discussion and conclusion

In a future version of the paper we will present further results from teacher logs and class observations to better understand the mechanisms behind the results and to shed light on where the complementarities failed.

References

Arias, F., and M. Zuñiga, 2012a, Desarrollo de curriculum y materiales para el aula del proyecto GeoMate.

Arias, F., and M. Zuñiga, 2012b, La capacitación de profesores en el proyecto GeoMate.

Borkum, E., F. He, and L. Linden, 2012, School Libraries and Language Skills in Indian Primary Schools: A Randomized Evaluation of the Akshara Library Program, NBER Working Papers, 18183.

Berlinski, S., M. Busso, J. Cristia, and E. Severin, 2011, Computers in Schools: Why Governments Should Do their Homework, in *Development Connections: Unveiling the impact of New Information Technologies*, Alberto Chong (ed), New York: Palgrave Macmillan.

Cheung, A. and R. Slavin, 2011, The Effectiveness of Educational Technology Applications for Enhancing Mathematics Achievement in K-12 Classrooms: A Meta-analysis, Report of the Center for Research and Reform in Education, Johns Hopkins University.

Clements, D., 1999, 'Concrete' Manipulatives, Concrete Ideas, Contemporary Issues in Early Childhood 1, pp. 45-60.

Cristia, J., P. Ibarraran, S. Cueto, S. Ana, E. Severín, 2012, Technology and child development: Evidence from the One Laptop per Child Program, IDB Working Paper Series, 304.

David, P., 1990, The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox, American Economic Review, 80 (2), pp. 355-361

Dobbie, W. and R. Fryer, 2012, Getting Beneath the Veil of Effective Schools: Evidence from New York City, forthcoming American Economic Journal: Applied Economics.

Fundación Omar Dengo (FOD) - Ministerio de Educación Pública (MEP), 2010, Diagnóstico Nacional de los niveles de Acceso, Uso y Apropiación de las tecnologías digitales en los docentes del sector público y subvencionado.

Fryer, R., 2012, Injecting Successful Charter School Strategies into Traditional Public Schools: Early Results from an Experiment in Houston, mimeo Harvard University.

Gersten, R., J. Ferrini-Mundy, C. Benbow, D. Clements, T. Loveless, V. Williams, I. Arispe, M. Banfield, 2008, Report of the Task Group on Instructional Practices, in: Foundations for Success: Report of the National Mathematics Advisory Panel. Available at: http://www2.ed.gov/about/bdscomm/list/mathpanel/report/instructional-practices.pdf

Glewwe, P., M. Kremer, and S. Moulin, 2009, Many Children Left Behind? Textbooks and Test Scores in Kenya, American Economic Journal: Applied Economics, 1 (1), pp. 112-135, 2009.

Glewwe, P., M. Kremer, S. Moulin, and E. Zitzewitz, 2004, Retrospective vs. prospective analyses of school inputs: the case of flip charts in Kenya, Journal of Development Economics, 74 (1) pp. 251-268.

Helpman, E., and A. Rangel, 1999, Adjusting to a New Technology: Experience and Training, Journal of Economic Growth, 4, pp. 359–383.

Lampert, M., 1990, When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching, American Educational Research Journal, 35, pp. 281-310.

Lara-Meloy Arias, T., S. Berlinski, M. Busso, L. Gallagher, J. Roschelle, and M. Zuñiga, 2012, Design and development of the student assessment instruments for GeoMate.

Machin, S., and S. McNally, 2008, The Literacy Hour, Journal of Public Economics, 9 (5–6), pp. 1441–1462.

National Research Council, 2001, Adding it up: Helping children learn mathematics. J. Kilpatrick, J. Swafford, and B. Findell (Eds.). Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

OECD, 2009, Learning Mathematics for Life: A Perspective from PISA. Available at: http://www.oecd.org/pisa/pisaproducts/pisa2003/44203966.pdf

Zuñiga, M., 2003, Aprendizaje mediado por tecnologías digitales: La experiencia de Costa Rica, in UNESCO (2003), Educación y Nuevas Tecnologías: Experiencias en América Latina, pp. 99-114.

Tables

Table 1: Annualized Unit Costs (USD)

	Interactive Board	Computer Lab	One-to-One
Equipment Deployment			
Initial Investment	34.9	73.4	111.3
Recurrent Costs	6.0	14.9	19.1
Total	40.8	88.3	130.4

Note: Equipment deployment includes hardware and set up costs. For one-to-one and laboratory hardware includes laptops, desktops, carts, routers, projectors, and lockers. For smart-boards hardware includes smart-boards, laptops for teachers, desktops, routers, projector, and lockers. Infrastructure adjustments necessary for set up are included in the set up costs. The assumed lifespan of the hardware (laptop, desktop, smartboard, etc) is 5 years. The recurrent repair costs are assumed to be 5% of the initial costs.

Table 2: Description of Test Items

M1: Basic geometry skills and abilities	M2: Higher-order geometric practices
Recognize geometric figures, given measurements, angles (e.g., recognize whether three lengths make up a triangle).	Identify parts of a geometric figure
Calculate an unknown given numbers or explicit information (big topic)	Classify geometric figures given properties in (angles, lengths)
Draw geometric figures given certain properties	Visualize (geometrically) theorems, problems, solutions
Construct figures to exemplify geometric situations	Justify theorems, using geometric figures, definitions or theorems
Recognize formal demonstrations of theorems.	Deduce patterns / formulate conjectures given examples
Write definitions	Read and use conventions and formal vocabulary

Table 3: Sample Size

	Control	Curriculum	Interactive-board	Computer Lab	One-to-One	Total
Schools	20	20	15	15	15	85
Classes 7th Grade	44	46	36	28	36	190
Teachers in 7th Grade Math	44	46	36	28	36	190
Student Data	1084	1182	965	738	861	4830

Table 4: Non-Response Rates

	Average and S.D. All	Difference w.r.t. Control (coeff and s.e.)				p-value	Sample Size
		Curriculum	Interactive-board	Computer Lab	One-to-One		
	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Student Level Variables							
Missing on Geo test day	0.091	-0.017	-0.008	0.029	0.009	0.201	4625
	[0.288]	[0.024]	[0.021]	[0.022]	[0.026]		
Geo test date (# days after end of geo unit)	14	1.813	-0.963	2.367	-1.675	0.094	4157
	[6.489]	[1.971]	[1.956]	[1.902]	[2.419]		
Geo Test admin by Teacher	0.056	0.032	-0.000	-0.014	-0.004	0.399	4157
	[0.230]	[0.022]	[0.017]	[0.027]	[0.020]		
Non-eligible Geo test (disability)	0.011	-0.010	-0.018	-0.020	-0.014	0.585	4881
	[0.103]	[0.012]	[0.012]	[0.012]*	[0.011]		
Missing SAT (among eligible)	0.211	-0.027	-0.084	-0.138	-0.083	0.572	4157
	[0.408]	[0.091]	[0.070]	[0.084]*	[0.079]		
Teacher Level Variables							
Missing Teacher Survey (pre-treatment)	0.005	-0.025	-0.020	-0.021	-0.020	0.902	190
	[0.073]	[0.019]	[0.015]	[0.016]	[0.015]		
Missing Teacher Survey (post-treatment)	0.032	0.003	0.016	-0.046	-0.013	0.304	190
	[0.175]	[0.035]	[0.038]	[0.038]	[0.033]		

Note: standard errors in brackets are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Differences in Pre-Treatment Means

	Average and s.d. All	Difference w.r.t. Control (coeff and s.e.)				p-value	Sample Size
		Curriculum	Interactive-board	Computer Lab	One-to-One		
Student-level Variables							
Percent Male	0.500	-0.029	-0.038	0.008	-0.016	0.330	4157
	[0.500]	[0.019]	[0.022]*	[0.026]	[0.025]		
Age	13.033	0.072	-0.093	0.032	0.015	0.004	4127
	[0.920]	[0.061]	[0.043]**	[0.052]	[0.058]		
Mother's Education (Primary)	0.436	0.046	-0.021	0.042	0.019	0.368	4106
	[0.496]	[0.044]	[0.048]	[0.064]	[0.050]		
Mother's Education (Secondary)	0.399	0.003	0.026	-0.045	0.030	0.180	4106
	[0.490]	[0.025]	[0.027]	[0.037]	[0.031]		
Mother's Education (University)	0.165	-0.049	-0.006	0.003	-0.049	0.275	4106
	[0.371]	[0.030]	[0.036]	[0.043]	[0.032]		
Number of Books at home	3.113	-0.085	-0.024	0.038	-0.151	0.541	3560
	[1.590]	[0.083]	[0.100]	[0.124]	[0.128]		
Have a PC/laptop at home	0.720	-0.033	0.020	-0.033	-0.009	0.342	3543
	[0.449]	[0.036]	[0.033]	[0.045]	[0.039]		
SAT (% Correct)	0.461	-0.019	0.003	-0.007	-0.022	0.394	3278
	[0.144]	[0.017]	[0.029]	[0.017]	[0.016]		
Teacher Level Variables							
Percent Female	0.495	0.004	0.179	0.161	0.048	0.342	190
	[0.501]	[0.125]	[0.110]	[0.150]	[0.122]		
Age	36.619	0.888	0.852	-1.066	0.713	0.269	189
	[7.743]	[1.324]	[1.278]	[1.209]	[1.422]		
Experience	11.652	0.500	1.414	0.154	-0.389	0.428	184
	[6.543]	[1.251]	[1.070]	[1.293]	[1.138]		
School-Level Variables							
Students 7th Grade	219.694	-0.650	-2.643	-5.310	4.757	0.944	85
	[114.174]	[14.637]	[15.904]	[15.904]	[15.904]		
Classes 7th Grade	6.847	0.000	-0.306	-0.372	0.094	0.618	85
	[3.053]	[0.367]	[0.399]	[0.399]	[0.399]		
Computer Lab	0.741	-0.000	-0.017	0.050	-0.083	0.877	85
	[0.441]	[0.142]	[0.154]	[0.154]	[0.154]		
Internet in School	0.729	0.150	0.101	-0.165	0.035	0.204	85
	[0.447]	[0.140]	[0.152]	[0.152]	[0.152]		
7th Grade Repetition	0.087	-0.018	-0.008	-0.013	-0.012	0.974	85
	[0.062]	[0.020]	[0.021]	[0.021]	[0.021]		
Not Urban	0.447	-0.050	0.110	-0.157	-0.157	0.335	85
	[0.500]	[0.144]	[0.156]	[0.156]	[0.156]		

Note: standard errors in brackets are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Fidelity of Implementation

	Average and S.D.	Difference w.r.t. Control (coeff and s.e.)				p-value
	All	Curriculum	Interactive-board	Computer Lab	One-to-One	
Reported by Teachers						
Learned teaching assignment before lottery	0.837	-0.106	-0.118	0.035	-0.055	0.285
	[0.370]	[0.079]	[0.084]	[0.082]	[0.079]	
Class learned geometry 1st Term	0.016	-0.020	-0.045	-0.037	-0.041	0.794
	[0.125]	[0.050]	[0.049]	[0.041]	[0.044]	
Class learned 4 geo units in 1st Term	0.126	0.066	-0.034	-0.092	-0.117	0.113
	[0.333]	[0.122]	[0.109]	[0.098]	[0.094]	
Students' Laptops	0.319	-0.093	-0.113	0.859	0.717	0.000
	[0.466]	[0.055]*	[0.058]*	[0.068]**	[0.080]**	
Interactive Whiteboards	0.183	0.004	0.951	0.001	0.006	0.000
	[0.386]	[0.007]	[0.027]**	[0.008]	[0.010]	
Teachers' Laptops	0.604	0.050	0.083	0.464	0.422	0.000
	[0.489]	[0.141]	[0.114]	[0.114]**	[0.113]**	
Class Materials	0.633	0.784	0.898	0.873	0.666	0.007
	[0.482]	[0.069]**	[0.052]**	[0.087]**	[0.062]**	
Reported by Students						
Students' Laptops	0.333	-0.051	-0.048	0.922	0.917	0.000
	[0.471]	[0.050]	[0.049]	[0.060]**	[0.056]**	
Interactive Whiteboards	0.214	0.002	0.950	-0.045	-0.049	0.000
	[0.410]	[0.038]	[0.029]**	[0.027]	[0.029]*	

Note: All models were estimated at the teacher/section level. In all cases, the sample size is 190. standard errors in brackets are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Geometry Test Results

	[1]	[2]	[3]	[4]
Curriculum	-0.202*	-0.162*	-0.171*	-0.171**
	[0.120]	[0.091]	[0.091]	[0.080]
Interactive-board	-0.100	-0.127	-0.120	-0.155*
	[0.129]	[0.104]	[0.098]	[0.093]
Computer Lab	-0.202	-0.196	-0.199	-0.210*
	[0.154]	[0.123]	[0.123]	[0.118]
One-to-One	-0.412***	-0.366***	-0.366***	-0.355***
	[0.128]	[0.097]	[0.094]	[0.091]
Observations	4,157	4,157	4,157	4,157
R-squared	0.057	0.203	0.207	0.219
Controls				
Strata	x	x	x	x
Students		x	x	x
Teachers			x	x
School				x
p-values of one side test H1:				
One2One <= Lab	0.055	0.055	0.060	0.093
One2One <= Curriculum	0.008	0.003	0.004	0.008
One2One <= Interactive board	0.002	0.002	0.002	0.012
Lab <= Interactive board	0.225	0.270	0.243	0.320
Lab <= Curriculum	0.500	0.375	0.397	0.366
Curriculum <= Interactive board	0.119	0.325	0.249	0.411

Note: standard errors in brackets are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. Individual controls include gender, age, mom educ, books, SAT. Teacher controls include gender, age, experience. School controls include # students in 7th grade, # classrooms in 7th grade, Lab in school, region dummies. Dependent variables: score is the % correct score and standardized IRT-score (was produced using the IRT parameteres in the control sample.) Both scores are standardizes using meand and s.d. of the control

Table 8: Treatment Effect Heterogeneity

	Student				Teacher			
Population:	Female	Male	Low Ability	High Ability	Low Experience	High Experience	Preference for Student-centered pedagogy	Preference for Teacher-centered pedagogy
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Curriculum	-0.218	-0.168	-0.015	-0.263*	-0.296	0.007	-0.191	-0.153
	[0.136]	[0.113]	[0.069]	[0.155]	[0.190]	[0.094]	[0.164]	[0.111]
Interactive-board	-0.068	-0.077	0.013	-0.097	-0.012	-0.094	-0.128	-0.001
	[0.149]	[0.125]	[0.081]	[0.157]	[0.251]	[0.124]	[0.188]	[0.139]
Computer Lab	-0.262*	-0.235	-0.143*	-0.245	-0.334*	-0.146	-0.431**	-0.132
	[0.156]	[0.149]	[0.080]	[0.189]	[0.195]	[0.180]	[0.206]	[0.140]
One-to-One	-0.461***	-0.351***	-0.149**	-0.512***	-0.513**	-0.204**	-0.388*	-0.399***
	[0.145]	[0.118]	[0.066]	[0.163]	[0.215]	[0.096]	[0.200]	[0.120]
Observations	2,125	2,032	1,658	2,499	2,182	1,862	1,976	2,068
R-squared	0.076	0.051	0.037	0.081	0.103	0.031	0.090	0.084

Note: standard errors in brackets are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. Models include controls for strata.

Table 9: Effect on Scales

	Models without Controls		Models with Controls		Sample
	Curriculum	Technology	Curriculum	Technology	Size
Students-level Outcomes					
Active learning (SRI)	0.035	0.079	0.029	0.079	4052
	[0.054]	[0.041]*	[0.047]	[0.034]**	
Preference for math (SRI)	-0.100	-0.019	-0.140	-0.055	3970
	[0.079]	[0.070]	[0.076]*	[0.059]	
Academic engagement (Chicago)	0.010	0.056	-0.041	0.015	3973
	[0.083]	[0.078]	[0.075]	[0.066]	
Academic press (Chicago)	0.006	-0.011	-0.011	-0.033	3917
	[0.049]	[0.041]	[0.048]	[0.039]	
Avoid novelty (PALS-UM)	0.050	0.075	0.074	0.085	3943
	[0.053]	[0.051]	[0.053]	[0.048]*	
Bad behavior (PALS-UM)	0.069	0.073	0.091	0.071	4030
	[0.070]	[0.062]	[0.056]	[0.054]	
Classroom activity	0.107	0.162	0.120	0.166	4157
	[0.049]**	[0.043]***	[0.044]***	[0.038]***	
Teacher-level Outcomes					
Innovation (Chicago)	0.239	0.003	0.244	0.077	184
	[0.203]	[0.153]	[0.219]	[0.172]	
Reflective dialogue (Chicago)	0.331	0.372	0.305	0.417	185
	[0.218]	[0.188]**	[0.211]	[0.197]**	
Access to New Ideas (Chicago)	0.219	0.355	0.177	0.384	184
	[0.259]	[0.206]*	[0.258]	[0.199]*	
Teaching efficacy (Chicago)	-0.270	-0.268	-0.195	-0.210	187
	[0.176]	[0.158]*	[0.178]	[0.162]	

Note: standard errors in brackets are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. Individual controls include gender, age, mom educ, books, SAT. Teacher controls include gender, age, experience. School controls include # students in 7th grade, # classrooms in 7th grade, Lab in school, region dummies. Dependent variables are scales that were produced via a polytomous IRT model whose parameters were estimated using the control population only. All scales are standardized using the meand and s.d. of the control population.

Figures

Figure 1

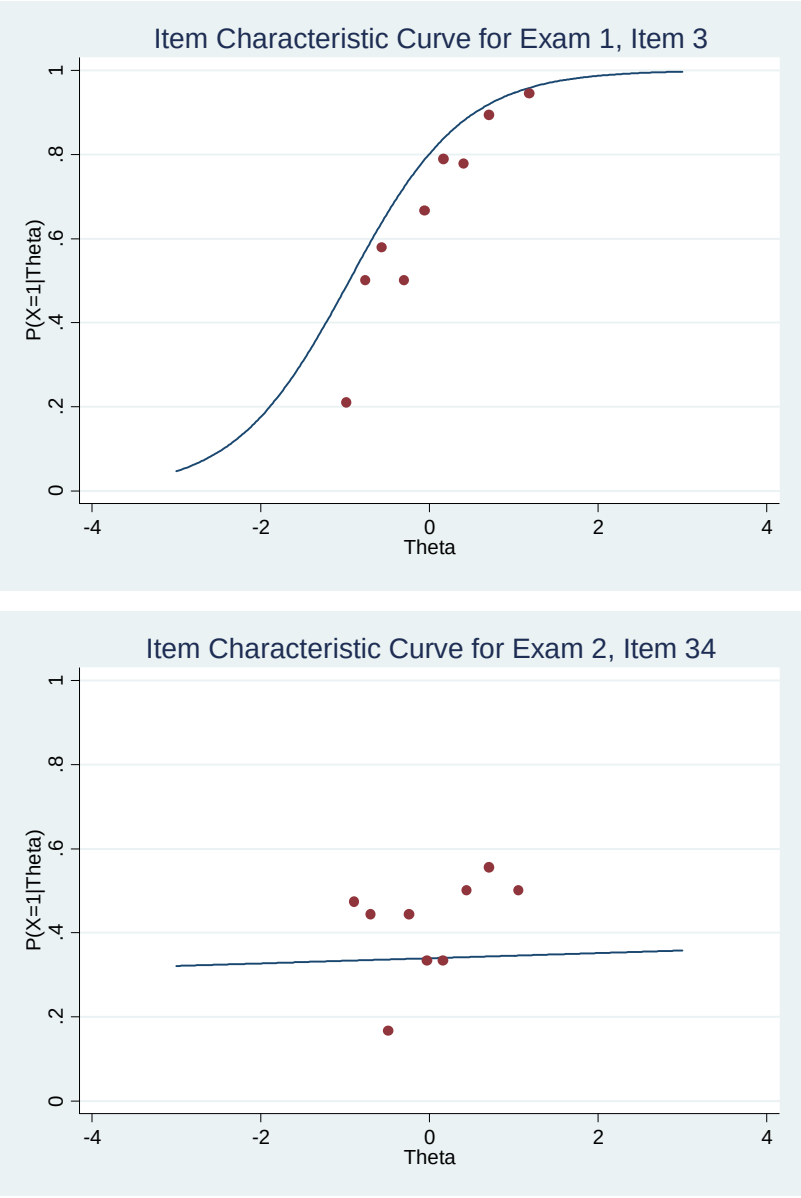
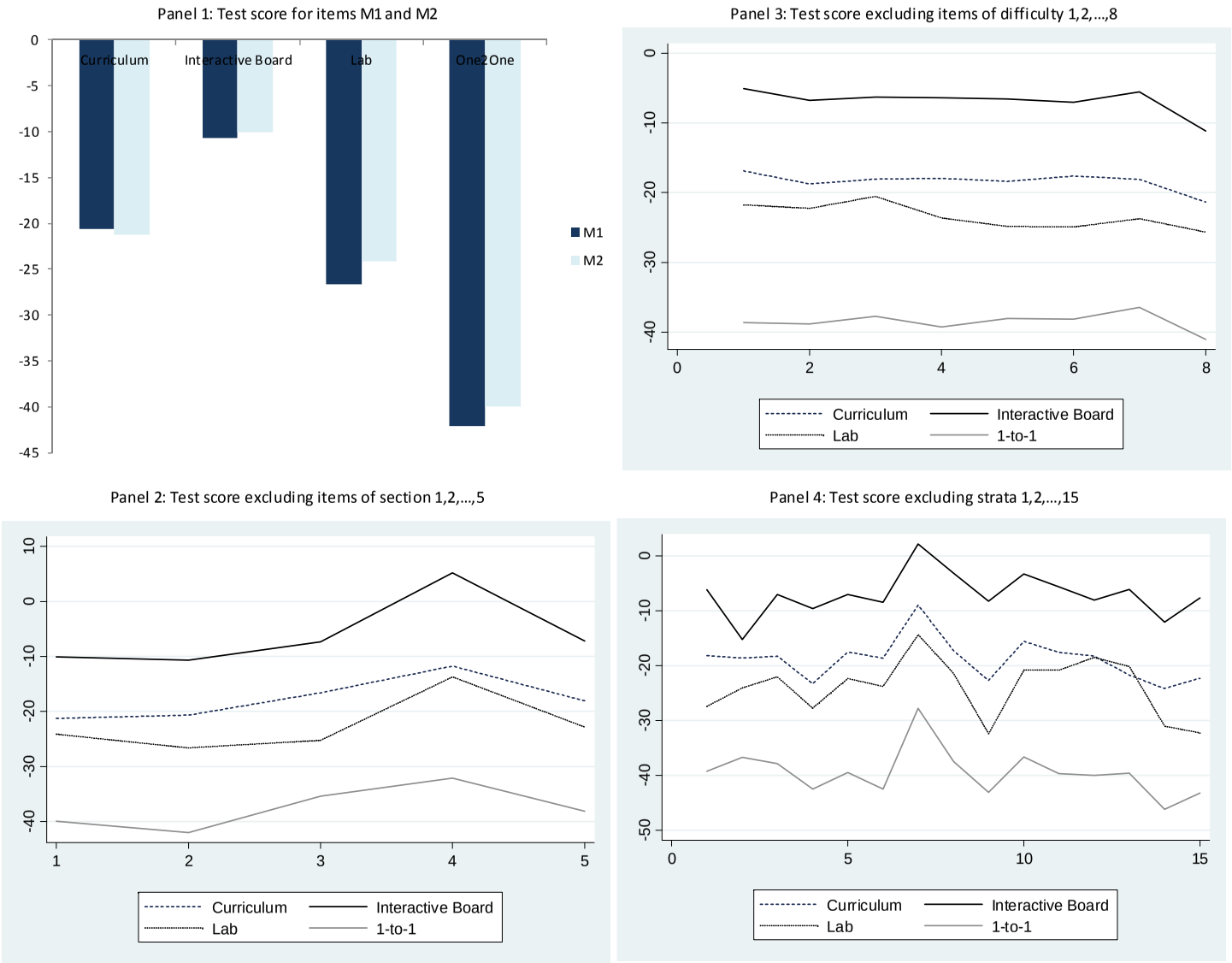


Figure 2: Robustness of results on scores



Appendix tables

Table A1: Geometry Test Results (% correct)

	Geo Test Score			
	[1]	[2]	[3]	[4]
Curriculum	-0.193	-0.153*	-0.161*	-0.158**
	[0.119]	[0.088]	[0.088]	[0.076]
Interactive-board	-0.071	-0.100	-0.097	-0.127
	[0.131]	[0.103]	[0.097]	[0.091]
Computer Lab	-0.245*	-0.244**	-0.251**	-0.248**
	[0.144]	[0.113]	[0.112]	[0.109]
One-to-One	-0.405***	-0.360***	-0.358***	-0.344***
	[0.124]	[0.091]	[0.087]	[0.086]
Observations	4,157	4,157	4,157	4,157
R-squared	0.058	0.202	0.206	0.215
Controls				
Strata	x	x	x	x
Students		x	x	x
Teachers			x	x
School				x
p-values of one side test H1:				
One2One <= Lab	0.087	0.105	0.124	0.171
One2One <= Curriculum	0.004	0.001	0.001	0.004
One2One <= Interactive board	0.001	0.001	0.001	0.005
Lab <= Interactive board	0.085	0.083	0.070	0.136
Lab <= Curriculum	0.326	0.172	0.176	0.193
Curriculum <= Interactive board	0.082	0.237	0.187	0.323

Table A2: Student Scales

Active learning (SRI)	1	In the math class I write math problems for other students to solve
	2	I write a few lines to explain how I solved a math problem
	3	I use math in situations outside the classroom
	4	I solve math problems that involve many steps and take more than 20 minutes
	5	I explain to the class how to solve a math problem
	6	I discuss possible solutions to math problems with other students
Preference for math (SRI)	7	How much do you like mathematics?
	8	Think about the most recent unit in your math class. Think about the activities and the math you learned. How much did you enjoy your math class during this unit?
	9	Think about the most recent unit in your math class. If math classes were always like this, would you be excited to take math classes in the future?
Academic engagement (Chicago)	10	I often count down the minutes until class is over.
	11	What I am learning in class is so interesting, I don't want class to end.
	12	I usually look forward to this class.
	13	I usually get bored with what we are learning in class.
	14	The topics we are studying are interesting and challenging.
	15	I work hard to do my best in this class.
Academic press (Chicago)	16	Nobody wastes time in class.
	17	Usually this is a difficult class.
	18	Usually the teacher asks difficult questions in class.
	19	Usually the teacher asks difficult questions on tests.
	20	Usually this class challenges me.
	21	This class really makes me think.
	22	Generally this class requires me to work hard to do well.
	23	The teacher expects everyone do their best all the time.
	24	The teacher expects everyone to work hard.
Avoid novelty (PALS-UM)	25	During class I prefer to work on tasks that are familiar to me rather than to learn how to do new ones.
	26	I don't like to learn a lot of concepts during class.
	27	I prefer to do my work as usual rather than to try something new.
	28	I like academic concepts that are familiar to me rather than ones I have never heard before.
	29	I would rather chose to work on something I already know how to do rather than something I have never done before.
Bad behavior (PALS-UM)	30	Sometimes I bother my teacher during class.
	31	Soemtimes I get in trouble with my teacher during class.
	32	Soemtimes I behave in a way that upsets my teacher during class.
	33	Sometimes I do not follow my teacher's instructions during class.
	34	Sometimes I cause disorder during class.
Classroom activity	35	Last week, how often during math class did the teacher conduct the following activities?
	36	Lectures / demonstrations.
	37	Class discussions.
	38	Work individually or in pairs.
	39	Work in groups.

Table A3: Teachers Scales

Innovation (Chicago)		The teachers in this school…
	1	Are really trying to improve their teaching
	2	Are willing to take risks to make the school better
	3	Are eager to try new ideas
	4	Have a positive "I can do" attitude
	5	Are continually learning and seeking new ideas
	6	Are encouraged to "grow" professionally
Reflective dialogue (Chicago)		In this school year, have you had conversations with your colleagues more than twice about…
	7	What helps students learn the best
	8	The mathematics curriculum
	9	The goals of this school
	10	Managing classroom behavior
	11	Teaching styles and learning
	12	Teachers in this school discuss instruction in the teachers' lounge, faculty meetings, etc
	13	Teachers in this school share and discuss student work with other teachers
	14	Experienced teachers invite new teachers to observe their class, provide feedback, etc
	15	The teacher body at this school makes new teachers feel welcomed
Access to New Ideas (Chicago)		In general, your training and professional development…
	16	Have been sustained and coherently focused, instead of brief and unrelated
	17	Include sufficient time to think carefully about new ideas and how to evaluate them
	18	Has been very connected with the school's plans for improvement
	19	Include opportunities to work productively with other teachers in this school
	20	Include opportunities to work productively with other teachers in other schools
	21	Have presented themes and provided follow up on them
	22	Have been about the needs of students in this class
	23	Encourage teachers to seek develop for themselves
Teaching efficacy (Chicago)	24	With enough effort I can even make students with the most difficulty understand the subject
	25	Events I can not control have a greater influence on the performance of my students than I do
	26	I am good at helping my students achieve significant improvements
	27	Some students will not make much progress this year, regardless of what I do
	28	I am sure I can make a difference in the lives of my students
	29	There is little I can do to ensure that all my students achieve significant progress this year
	30	I perform well under any learning challenge