# The Efficacy of a Pre-Algebra Cognitive Tutor in Chile and Mexico

Jason Imbrogno, Ignacio Casas, Paul Goodman, and Luis Quintero

November 11, 2013

#### Abstract

A math cognitive tutor (MCT) system widely used throughout the U.S. was adapted for use in Chilean and Mexican public middle schools. The curriculum requires large changes in pedagogy, including the use of computers for individual students to progress through an extended pre-algebra program. The study was conducted over a 6-month time period. Using a hierarchical linear model (HLM), we show that students enrolled in schools which were randomly assigned to adopt the MCT significantly improved their standardized math test scores as compared to control group peers. However, the implementation of the changes in the schools and classrooms was not perfect. Those schools which were better prepared to make changes, especially those with sufficient computers and technical support services, saw their students master more of the software part of the curriculum. Students and teachers generally viewed the MCT positively. The results on math performance and attitudes are promising for further propagation of the MCT curriculum. Knowledge from this study regarding the structure and implementation required for schools to successfully exploit the unique teaching capabilities of the MCT should guide the future diffusion of this specific technology.

# 1 Introduction

# 1.1 Implementing the Math Cognitive Tutor in the Classroom

This paper examines the impact of an educational intervention on the mathematics performance of 7th grade students in public middle schools in Chile and Mexico. A math cognitive tutor (MCT) targeted at the pre-algebra level represents the intervention. Of the total time devoted to mathematics, students are expected to spend around  $\frac{3}{5}$  of it in the classroom and  $\frac{2}{5}$  in the computer lab using the MCT software.

There are significant pedagogical changes that are strongly suggested for teachers to use in the classroom portion of the overall curriculum. For the planning of the traditional classes, the teacher can utilize performance reports provided by the MCT system that are individualized to each student.

A more collaborative "group learning" strategy, where the teacher serves more as a mentor or coach while students practice problems (instead of a pure lecturer), is also suggested for those classes. Teachers required training before beginning their use of the MCT curriculum. More information on the classroom changes and training provided to teachers can be found in Casas, Goodman, & Pelaez (2011) or Casas, Imbrogno, & Vergara (2013). This paper does not include much further discussion of the classroom changes, but instead focuses on the computer-based MCT instruction and its impact on the 7th grade students.

#### 1.2 The ACT-R Theory Behind the MCT

The MCT is based on Anderson's (1993a, b) cognitive theory called adaptive control of thoughtrational, or ACT-R. Cognition is modeled as a system of piecemeal knowledge components. According to the theory, the link between declarative, or factual, knowledge and procedural knowledge (problem-solving skill) is strengthened as a power function of practice. Repeated attempts to solve a problem through the use of a particular skill allows students to perform that skill both more quickly and more accurately. In other words, practice improves the performance measure of time spent or errors made by reducing them over repeated attempts. The models in Nowell & Rosenbloom (1981) and Anderson & Schooler (1991) showing this relationship mathematically can be shown in the simplest form using a performance measure P, learning rate b (less than one), and number of attempts N as:

$$P = b^N \tag{1}$$

As the number of attempts increases, the performance measure (amount of time to solve a problem, probability of making an error) decreases. A graphical representation of this relationship is often referred to as a "learning curve."

Using the MCT software, students can demonstrate proficiency in many knowledge components separately. The MCT combines student actions and a generalized power function to estimate how well the student understands each knowledge component. It uses this estimation to build individualized instruction that focuses on the specific components with which each student struggles. The MCT presents different problems to different students as they progress through the software because its choice of problems to present to each individual is determined by its interpretation of which knowledge components the student has and has not learned. The MCT tailors instruction to the demonstrated ability level of the student and selects problems designed to increase student learning in areas of weakness until a level of mastery is shown. Students gradually build up their more complex problem-solving skills by separate acquisition of a number of these smaller building blocks.

Using the ACT-R theory of individual knowledge components, the MCT is able to break down student misunderstandings at a finer grain level than even individual problems. The MCT itself tracks the knowledge components as separate skills,<sup>1</sup> and an example of the component skills in a given problem should suffice to demonstrate the effectiveness of the ACT-R approach in the MCT. Say a student is asked to identify the greatest common factor (GCF) of 27 and 18. A problem in the MCT proceeds in separate steps. In the first step of the problem, he is asked to list separately the factors of 27 and 18 (skill 1). In the second step, he is asked to identify the greatest common factor of 27 and 18 by appealing to his previous lists (skill 2). Finally, he is asked to identify the greatest common factor of 27 and 18 by referring to his last step (skill 3). See Table 1 for a visual on the problem steps. The three separate skills require a student to (1) factor numbers, (2) choose common numbers between sets, and (3) choose the greatest number from a set. If a student makes a mistake or asks for a hint in the first step, at the conclusion of the problem the MCT is unlikely to produce an immediately subsequent problem asking him to identify the GCF of two numbers again. Instead, it will track back and give the student a problem that focuses on the missing skill (skill 1) either explicitly, as in "List the factors of 36" or in another domain, such as "Please reduce the fraction  $\frac{12}{48}$ ," where the student must begin the problem by demonstrating the same skill 1 as in the GCF problem.

## 1.3 Bridge to Algebra MCT

The main technological system in this study is the Bridge to Algebra MCT produced by Carnegie Learning, Inc. It covers math material commonly referred to in the U.S. as pre-algebra (such as number sense and algebraic thinking, fractions, decimals, linear functions, and number systems). Koedinger & Anderson (1993), Koedinger, Anderson, Hadley, & Mark (1997), Anderson & Schunn (2000), and Anderson (2002) provide descriptions of the the application of ACT-R to the development of the software itself, plus early implementation and design issues. The MCT provides each student a personalized learning environment. A problem generator provides each student with a different set of problems for each skill module. When presented with a problem, the student can ask for hints during all the problem solving processes. Problems are presented in order of complexity. The system keeps track of the number of mistakes and hints used over time. When a skill is completed, the student receives feedback and moves to the next module. The principles underlying the tutor include individualized instruction, opportunities for practice, a scaffolding and hint system that focuses attention on appropriate processes for problem solving, an extensive feedback system that facilitates learning for the student, and an extensive data system that permits diagnosis of student problem solving processes. These principles are designed to enhance learning.

Students do not have to finish the entire software curriculum, and, in fact, very few of them actually do. The prepared Bridge to Algebra curriculum consists of 14 units, 57 sections, and 552 skills.<sup>2</sup> Students progress through the curriculum by mastering skills and sections. According to the MCT developers, these are the best measures of the student learning that has occurred via the software part of the curriculum.

 $<sup>^{1}</sup>$ A knowledge component and a skill are the same thing. "Skill" is the term used by the MCT itself in its presentation to students.

 $<sup>^{2}</sup>$ The entire Bridge to Algebra curriculum offered by Carnegie Learning is longer. A subset was used in this study.

Each section in the MCT contains many skills, and the student must pass all of them to pass a section. The student can advance to the next section without passing all skills once he spends sufficient time and effort on the section, as measured by the number of problems attempted. But the student would still "fail" the section even though he moves on. In order to "master" a section, the student must master all skills within that section. By design then, it is more difficult to master a section than a skill, even leaving aside the fact that there is more material involved. Mastering 9 out of 10 skills in each of four sections, plus 10 of 10 in a fifth section, would result in a skills mastered percentage of 92% (46 skills mastered out of 50), but a sections mastered percentage of just 20% (1 section mastered out of 5). Also by design, the number of "skills mastered" is much greater than the number of "sections mastered," and the two measures are highly correlated. Because the MCT curriculum is personally adaptive for each student, the number of questions required to master a skill varies. If a student correctly answers the steps in a few problems involving a specific skill, he is adjudged to have mastered that skill. But errors and hints, again related to a specific skill, often result in more questions composed of that skill being asked of a student. Therefore, skills and sections mastered are better indicators of demonstrated understanding than simply time spent or problems faced, or even problems answered correctly. Skills mastered and sections mastered are part of the output of the MCT for the use of the teachers and students that reflect the underlying estimation of the student-specific learning curves.

# 1.4 Related Efficacy Work

The main rationale for this research is the following: in most countries in the world, economic development is based on an educated population, well versed in math, science, and other similar disciplines (National Research Council, 2007; Hanushek & Woessmann, 2008). An educational system that creates a supply of people well versed in mathematics and science is likely to be in a better position to improve the country's economic development. In countries where literacy, language, and mathematics understanding are low, the opportunity for economic development will be low. The goal of this study is to demonstrate the success, or lack of success, of a classroom MCT intervention. If successful, broader diffusion of this MCT could be expected.

Second, this research examines the generalizability of the MCT system. The existing assessments of this specific MCT occur in English-speaking settings. One should note that the issue here is not simply translating from English to Spanish. There are additional changes in the problem content and reprogramming to fit the local contexts. For example, word problems related to "starting a lemonade stand must be adapted for international student understanding. This study takes place in public schools in Chile and Mexico. Changes were made in the language, problems, and programming to adopt the MCT to Spanish-speaking environments. There are no systematic studies of this MCT in a Central or South American environment, though other researchers (such as Banerjee, Cole, Duflo, & Linden, 2007) have shown positive treatment effects from similar computer-assisted math learning programs in international settings.

The third purpose is to shed more light on the processes that produce successful, or unsuccessful,

MCT interventions. As suggested in Cook (2003), we seek to peer into the "black box" and describe some of the implementation quality and measurements of intervening processes that lead to effective use of the MCT. Though our measurements are different, we follow a similar thought process as Pane, McCaffrey, Steele, Ikemoto, & Slaughter (2010) in this regard. This MCT technology has been adopted in large school districts in the U.S.<sup>3</sup> for many middle and high school math courses. Papers on the student performance results of this technology enhanced learning system have found mixed evidence. Some studies on MCT (Koedinger & Anderson, 1993; Koedinger et al., 1997; Morgan & Ritter, 2002; Ritter, Anderson, Koedinger, & Corbett, 2007; Arroyo, Woolf, Royer, Tai, & English, 2010; Ritter, 2011) have shown positive treatment effects while others (Dynarski, 2007; Cabalo, Jaciw, & Vu, 2007; Campuzano, 2009; Pane et al., 2010) have shown insignificant or even negative treatment effects. We add to the program evaluation literature by focusing on implementation quality in addition to the more straightforward question of overall effectiveness in an effort to understand why some studies report positive treatment effects and others disagree. If the difference is a matter of proper implementation, we hope to improve upon the process of bringing the MCT into classrooms.

# 1.5 Country Settings

In the past two decades, several technology driven initiatives to improve education in South American countries have been introduced with varying degrees of success (de Ferranti et al., 2003; Scheurmann & Pedro, 2009; Chong, 2011). In general, the focus of these initiatives has been two-fold: providing basic technology infrastructure in the schools, plus computer literacy (for teachers and students) through training. Even though there have been clear advancements in the access to information and to electronic educational materials, the promises of these types of interventions to achieve improved learning have not yet been fully accomplished. Performance in national and international tests such as Program for International Student Assessment (PISA) and Trends in International Mathematics and Science Study (TIMMS) in the region has not increased as expected, and in some cases has even deteriorated (Hanushek & Woessmann, 2009).

Chile was ranked 96th out of 133 countries regarding quality of primary education in a 2009-2010 report by the World Economic Forum (Schwab, 2009). Even though international 2009 PISA tests reported large gains in math and language education in Chile with respect to the 2006 tests, Chile is still far below the average of the Organization for Economic Co-Operation and Development (OECD) countries. From the 2009-2010 World Economic Forum report:

The main area requiring improvement for Chile going forward remains the unsatisfactory quality of its educational system, notwithstanding increasing investment in education and rising educational attainment rates. Despite a slight improvement in both cases, primary and higher education continue to be assessed fairly poorly at 96th and 45th ranks, respectively, pointing to the need for further upgrading if Chile is to catch up with best practice countries and establish an innovation-conducive environment.

<sup>&</sup>lt;sup>3</sup>Including Los Angeles, Chicago, St. Louis, Miami, Baltimore, and Pittsburgh

According to the Chilean Ministry of Education,<sup>4</sup> there is a dramatic gap in quality of education between public and private schools (K-12) in Chile. This situation has produced large inequality and critical social unrest in Chile for the past few years. With very few exceptions, the best results on the national Chilean tests (in English, The System for Measuring the Quality of Education; Spanish acronym SIMCE) are achieved exclusively in the private schools, even though they represent just 8% of the K-12 educational system. Many of the public school math teachers have little or no formal training in math. This MCT implementation was done exclusively in public schools.

The situation is much the same in Mexico. Although their rank of 65th out of 133 in health and primary education in the same World Economic Forum report outstrips Chile, there is still much progress to be made, especially in math and science in the public schools. From the report:

Last but not least, the higher education and training system (74th) does not seem to provide the economy with the necessary pool of skilled labor, notably scientists and engineers (94th), and is not creating an environment conducive to adopting new technologies (71st in the technological readiness pillar) and generating new ones (78th in the innovation pillar). Further action is needed to liberalize markets, upgrade the educational system, and improve public governance in the country.

# 2 Data

# 2.1 Test Scores and Demographics

The first international implementation for the adapted Bridge to Algebra MCT occurred in public schools of one selected district each in Santiago, Chile, and Mexico City, Mexico, in 2011. Within the selected Chilean district, all 24 schools were invited to participate. Of those, 15 expressed initial interest in participating in the study and, after further consultation about the necessary training and pedagogical changes that would be required of teachers from schools that ended up in the treatment group, 12 remained in our study sample. The final design randomized six schools into treatment and six into control.<sup>5</sup> In total, there are 310 Chilean students in six treatment schools and 358 in six control schools that are used in the results for this paper. All of the students are in 7th grade. About half the schools had a single classroom for all students, and all schools with multiple classrooms had only one teacher responsible for covering them all.

In Mexico, the randomization took place at the classrooml level in four public schools of a district

<sup>&</sup>lt;sup>4</sup>Information pulled from the Ministry's website at http://www.mineduc.cl/

 $<sup>{}^{5}</sup>$ There was also an all-boys school in the Santiago district which is significantly better than the rest of the district schools in terms of socioeconomic statuses and achievement levels of students. That school participated in the study in a different way. Due to concerns that they did not have enough computers to allow appropriate access for all students, the school split its student population into treatment and control groups. Because it was so different from the other district schools, it is left out of the analysis that follows. The thrust of the empirical findings does not change when the school is included. If anything, the treatment effect is shown to be stronger when the high quality school is included.

near Mexico City. Each school had the same teacher teaching four different sections of math classes. Within each school, one of the four sections was randomly assigned to use the MCT while the other three served as control groups. The choice of one treatment and three control classes per school stemmed from computer access limitations. In total, there are 156 Mexican students in treatment classrooms and 478 in control classrooms. Again, the students were 7th graders.

The student achievement outcome measure used in this study comes from two comprehensive, grade-level pre-algebra exams given to all students. In both countries, the exams used were outside copies of the national standardized exams from Chile (SIMCE, as referenced earlier). One exam was given in May, near the beginning of the school year and before the MCT implementation. The other was given six months later. Both exams consisted of 44 multiple choice questions. The tests were reviewed and approved by both the Chilean education authorities and the MCT developers prior to their use in the study. The developers agreed that the material in the exams was both grade-level appropriate and covered by the software. The math material focuses on pre-algebra concepts, as does the software used by the treated students.

Unfortunately, there is little student characteristic data from both countries. The only demographic variable is gender, and student pretest scores are the only measure of prior achievement. Gender is binary, and we assign males = 1 and females = 0. Tables 2 and 3 show the average and (standard deviation) of the pretest score and percent male by classroom in both Chile and Mexico. Classroom 4 in school 4 in Mexico consisted exclusively of female students. Table 4 aggregates the same data by type within each country, and displays it similarly.

Achievement gaps are an important consideration for many policymakers and school administrators. We already briefly discussed the differences in Chile and Mexico between public and private schools. New programs and curricula that prove useful only to select students, especially those at the top of the achievement scale, are not as likely to experience widespread adoption because of concerns over widening the achievement gap and leaving slower students behind. In order to investigate the differential treatment effects of the MCT curriculum across the ability distribution, students were separated, by school, into three separate tertiles<sup>6</sup> depending on their pretest scores: high, middle, or low. Next, the students in each tertile were aggregated across schools. We set up the tertile designations separately by school because there are clear differences in initial achievement levels across schools. We are interested in separating out the differential treatment effects on more versus less skilled students. By design, this must consider the tertiles within the same school or else school differences in pretest scores would also be picked up at the stage of tertile assignment; schools with a high average pretest score would have more representation in the high tertile sample, for example. Due to the discrete nature of the exam scores and the fact that border scores were assigned to the lower tertile, the total observations did not end up equal across the tertiles. In Chile (Mexico), the lowest tertile has 254 (236) total observations across treatment and control students, the middle tertile has 220 (202), and the high tertile has just 194 (196) observations.

<sup>&</sup>lt;sup>6</sup>Other splits were examined. The results are largely the same for quartiles and quintiles. The tertile divisions were ultimately chosen to present here because they adequately capture the different ability levels while maintaining sufficient sample sizes across each school-tertile designation.

#### 2.2 MCT Software

The MCT software also logs each student's usage of and progress through the software-based part of the curriculum. The data logs store information regarding the total software usage time, average hints and errors per problem, as well as the total units, sections, skills, and problems encountered. The software also stores the percentage of the skills and sections seen by the student that were mastered. Obviously this data only exists for the treated students since the control group did not use the software. Table 5 shows the average and (standard deviation) of the units completed, total usage hours, skills mastered, and sections mastered for each treatment classroom.

Whether measured by units completed or total hours, the students in Chilean schools 1 and 6 and Mexican school 4 spent the most time, on average, using the software. There are obviously differences across the classrooms and schools in the amount of time their students spent using the software. The MCT curriculum calls for 40% of class time to be devoted to the software. In practice, this would constitute two days of a typical school week, which over a six month time period would mean roughly 25 hours spent in the computer lab in lieu of traditional math lectures. Note that this standard for total MCT software usage hours by students was not met, on average, by any classroom in any school. Students were also encouraged to use the software during their free time and even after-school when the lab was available. Conversations with the school administrators, which will be more fully explained and evaluated later, confirmed that the twice per week standard for computer use was rarely met. Students often only visited the lab once each week during class time, and very few of them accessed the software in their own time. Some of the schools did not have enough computers for every student, further hindering their personalized time with the software when they were forced to share the machines. Units, sections mastered, and skills mastered are highly correlated measures<sup>7</sup> across students (less so for usage hours).

## 2.3 Schools

The differences between schools are another component of the analysis in this paper. Our research team determined values for the treatment schools in four specific areas of interest (basic inputs, infrastructure, implementation, and learning environment) in an attempt to quantify the differences between them. It is valuable to examine how these school-level variables might affect student software usage. The basic inputs variable is an attempt to quantify all of the following: quality of students and teachers, budget issues, socio-economic status of students, absenteeism rates, and discipline rates in the schools. The infrastructure measure was probably the most important one in this context, and likewise the one where with the lowest average score. Infrastructure captures internet connection quality, presence of tech support, and the computer to student ratio. Each treatment school was expected to have at least one computer per student in the computer labs, but that was in fact not the case in some of the schools. Schools which did not provide enough computers were obviously at a severe disadvantage concerning student access to the software part of the MCT curriculum. Students

<sup>&</sup>lt;sup>7</sup>Smallest pairwise correlation between any combination is 0.93.

were forced to rotate and take turns in the computer lab, restricting the hours of access and amount of material covered in the math program. Connectivity to the internet was also necessary to run the software. Implementation is a measure of teacher and principal enthusiasm, motivation, and commitment toward the new curriculum. Training for instructors is a necessary part of the MCT implementation, and training session attendance and participation by teachers was incorporated into the rating. Finally, learning environment is a measure of the frequency of MCT use, peer cooperation among students in both the classroom and lab, and student motivation and behavior in the lab.

Each rating was made by three different, affiliated members of the research team based on conversations, notes, and feedback from the Chilean project manager and the teacherss and principals in the schools. Raters were blind to the school itself in the write-ups. Each of the four criteria was evaluated on a 1-5 scale, with 5 being the best rating. There was a high convergence among raters.<sup>8</sup> Table 6 shows the ratings given to each school for each variable. It also includes a total rating which is simply the sum of the four component pieces. School 2 in Chile and school 4 in Mexico rate particularly poorly on these measures. School 3 in Mexico is the only school with a "perfect" rating.

#### 2.4 Surveys

We administered surveys for nearly all students in the treatment schools.<sup>9</sup> The students were asked questions on a range of topics related to mathematics, computers, and the MCT software. Their survey responses were anonymous and could only be linked to students at the school level. The surveys requested that students indicate their perceptions in the following areas: (1) Ease of use and clarity of the MCT software, (2) Teacher help with the MCT, (3) Computer lab infrastructure, and (4) Effectiveness of the MCT. The anonymous aspect of the surveys reduces concerns about biased reporting by students. The surveys are complementary to the school characteristics discussed in the previous section. The school variables are derived measures from conversations with the researchers and principals involved in the project at the school level, plus overall input and perspective from the Chilean project manager. It is a top down view of the differing school characteristics. In contrast, the survey responses can be thought of as a bottom up view of the schools. They are the beliefs, feelings, and perceptions of the students themselves. The students likely have no idea how their particular school or classroom matches up against other schools in the survey topics, but their input is quite relevant since the curriculum change applies directly to them.

The survey questions associated with each topic area are shown below. All questions use a scale ranging from "very low" to "very high," with the responses agreeing with the survey question statement (i.e. for the first component of Ease of Use, the "very low" answer said "very unsatisfied").

Surveys were also administered to the teachers in the six Chilean schools which adopted the

<sup>&</sup>lt;sup>8</sup>All three researchers rated all ten treatment schools (six in Chile, four in Mexico) on the four characteristics using the 5-point scale. On 31 of the 40 school-characteristic ratings, all three researchers assigned the exact same score. The remaining 9 school-characteristic ratings which showed differences amongst the raters never differed by more than one point by any pairwise set of raters and were resolved through discussion and consensus in order to assign one value for them. Fleiss'  $\kappa$  of 0.79 reveals substantial agreement among raters.

<sup>&</sup>lt;sup>9</sup>Mexican school 4 has no available survey data.

MCT. The survey questions and responses from teachers are shown in the Appendix.

- 1. Ease of Use and Clarity of MCT Software
  - Indicate how satisfied you are with the ease of use of the Cognitive Tutor.
  - Indicate how satisfied you are with the clarity of the instructions that the Cognitive Tutor offers.
- 2. Teacher Help with the MCT
  - Indicate how satisfied you are with the help the teacher gave you to use the Cognitive Tutor.
  - Indicate to what extent your teacher helped you use the Cognitive Tutor in the lab.
  - Indicate to what extent your teacher handled the Cognitive Tutor adequately in the math classes (i.e. he knew the contents, steps, hints, etc)
- 3. Infrastructure
  - Indicate your level of agreement with the following statement about computers: The computers worked adequately in the lab.
  - Indicate your level of agreement with the following statement about computers: The internet connection worked in the lab.

#### 4. Effectiveness of the MCT

- Indicate to what extent your math learning improved with the Cognitive Tutor.
- Indicate your level of agreement with the following statement about the Cognitive Tutor: The Cognitive Tutor is a useful resource to learn math.
- Indicate your level of agreement with the following statement about the Cognitive Tutor: I would like to keep using the Cognitive Tutor in the math classes.
- Indicate your degree of comfort with this new way of learning math.

# 3 Research Questions

The main objective of this study is to estimate the causal impact of the Bridge to Algebra Cognitive Tutor curriculum on Chilean and Mexican students' pre-algebra achievement. The design includes treatment and control groups with before and after measures of math ability. We also investigate the process variables of the MCT itself to determine whether the amount of use is reflected in the achievement measure. Because the MCT curriculum requires significant changes in the classroom structure, school technological infrastructure, and student and teacher behavior, the study incorporates measures of school characteristics that include computer access, student background, discipline issues, teacher fidelity to training, and classroom fidelity to suggested usage of the MCT. Last, students are surveyed in the schools using the MCT to better understand its perceived ease of use and effectiveness.

Using this framework as a guide, we address the following questions:

1. Does the MCT significantly affect math performance compared with students in a controlled condition?

ROR.

- 2. Do the MCT process indicators at the student level affect changes in math performance?
- 3. Do school characteristics affect the process indicators in the MCT?
- 4. What are the student attitudes about the MCT experience?

# 4 Research Model

#### 4.1 Initial Balance Check

It is necessary to check whether the randomization process that determined which schools would implement the MCT curriculum and which would not created balanced treatment and control groups before evaluating results. If the groups are not equal on observable characteristics and pretest scores, it would not only call into the randomization processes employed but also the relevance of the results. Since the randomization was done differently in each country, we evaluate each country separately. The only observable characteristic we have is gender. Balance checks are run on both the proportion of male students and the average pretest score by treatment assignment using two-sided t-tests of equivalence with unequal variance. The comparisons are also done separately by tertile.

# 4.2 Treatment Effects

We follow the hierarchical linear modeling (HLM) approach advanced by Raudenbush & Bryk (2002) and used in a similar design in Pane, et al (2010) to estimate the causal impact of the Bridge to Algebra MCT on student math performance. We have students nested within classrooms within schools. The treatment group is represented by the students who used the MCT, and the control group by the students who did not. The estimation and results for the two countries are kept separate because the unit of randomization was different in the two settings. In Chile, entire schools were randomized into treatment and control groups; in Mexico, randomization occurred within each school such that one classroom was chosen for treatment and the rest were kept as controls.

Let  $Y_{ijk}$  be the posttest score for student *i* in classroom *j* in school *k*. Similarly, let  $y_{ijk}$  represent the pretest score and  $X_{ijk}$  student covariates.<sup>10</sup> At the student level (level 1), I model the posttest

 $<sup>^{10}\</sup>mbox{Commonly},$  these could include race, gender, free/reduced lunch status, etc. In this study, all I have access to is information on gender.

score as a function of the pretest score, covariates, classroom level (level 2) variables ( $\mu_{jk}$  for now), and an error term  $\epsilon_{ijk} \sim N(0, \tau_1^2)$ .

$$Y_{ijk} = \mu_{jk} + \beta_1 y_{ijk} + \beta_2 X_{ijk} + \epsilon_{ijk} \tag{2}$$

The classroom level equation incorporates the treatment assignment. The model will follow the experimental design from Mexico. Let type be denoted as  $T_{jk}$ . The treatment classes have T = 1 and the control classes have T = 0. Let  $\overline{y_{jk}}$  represent the class average of the pretest score. Though we are agnostic about the nature of the effect of the average class pretest score on individual student performance on the posttest in this paper, common arguments for its inclusion revolve around peer effects in the classroom. For  $\eta_{jk} \sim N(0, \tau_2^2)$ ,  $\mu_{jk}$  is modeled as:

$$\mu_{jk} = \gamma_0 + \gamma_1 T_{jk} + \gamma_2 \overline{y_{jk}} + \eta_{jk}$$
(3)

Combining Equations (2) and (3) yields:

$$Y_{ijk} = \gamma_0 + \gamma_1 T_{jk} + \gamma_2 \overline{y_{jk}} + \beta_1 y_{ijk} + \beta_2 X_{ijk} + \epsilon_{ijk} + \eta_{jk}$$

$$\tag{4}$$

Looking at Equation (4), the parameter of interest is  $\gamma_1$ . The estimate of  $\gamma_1$  is the treatment effect, in terms of standardized test scores, for the 7th grade students using the MCT.

In Chile, the randomization occurred at the school level. A similar line of reasoning as shown above helps derive the following empirical model for Chile, where all that changes in the equation itself is the elimination of the j subscripts on  $\overline{y}$ , T, and  $\eta$ . This obviously also changes the estimation slightly.

$$Y_{ijk} = \gamma_0 + \gamma_1 T_k + \gamma_2 \overline{y_k} + \beta_1 y_{ijk} + \beta_2 X_{ijk} + \epsilon_{ijk} + \eta_k$$
(5)

Equations (4) and (5) are estimated across the entire sample within each country, and then separately according to each tertile. We used the *xtmixed* command in Stata to estimate the models. Specifications are run both including and not including the classroom pretest average  $(\overline{y_{jk}})$ .

#### 4.3 MCT Process Data

The HLM models proposed in the last section are also used, with slight modification, to address the second research question. Here we investigate whether or not accomplishing more of the software itself is predictive of better math test performance. The control students do not use the MCT software at all, so they are not included in this analysis. Rather, looking at just the treated students,

we explore the relationship between software usage (in terms of units completed, total usage hours, sections mastered, and skills mastered) and performance (as measured by posttest scores) by controlling for pretest score, gender, and school/classroom. The approach is to estimate Equation (6) below for Mexico, which is very similar to Equation (4) shown previously.<sup>11</sup> The switch to  $\lambda$  from  $\gamma$  is purely for notational convenience to further distinguish the equations. Here, the type  $T_{ijk}$  is dropped and  $MCT_{ijk}$  represents the specific MCT data considered. Separate regressions are run for each of the four MCT process variables. The country's are evaluated separately. The parameter of interest is  $\lambda_1$ . The estimate of  $\lambda_1$  is the amount by which we expect a student's posttest score to increase when he increases the value of the MCT variable under consideration by one.

$$Y_{ijk} = \lambda_0 + \lambda_1 M C T_{ijk} + \beta_1 y_{ijk} + \beta_2 X_{ijk} + \epsilon_{ijk} + \eta_{jk}$$

(6)

## 4.4 School Characteristics

We are interested in explaining the relationship between the school characteristics, MCT process variables, and posttest scores. However, we only have ten schools containing students using the MCT. Regression techniques with such a small sample size are not useful. To investigate the relationship between the school characteristics and other data, our approach is quite simplistic. We average the MCT variables and posttest scores for each school, and then check the pairwise correlations between them and each of the four school characteristics (basic inputs, infrastructure, implementation, and learning environment). Carnegie Learning stresses that school administrators and teachers must be on board with this curriculum change for the MCT to be effective. In addition, there must be sufficient computer access available for students. As discussed earlier, our conversations with the on-location research team and individual school officials revealed varying degrees of proper implementation across schools. We want to know whether implementation and infrastructure differences possibly help explain the large differences we see across students and across schools in terms of completion and mastery of the Bridge to Algebra MCT and test scores.

# 4.5 Surveys

The survey responses are aggregated across classrooms within each school. We set up a combination of questions that closely matches specific areas of interest in the evaluation of the MCT, as explained earlier. Since each interest area (ease of use of tutor, teacher help with tutor, infrastructure, and effectiveness) is composed of multiple responses, the results we report indicate the percentage of people across all questions in an area who answer "very high" or "high." These survey responses are helpful since they capture student perceptions of school infrastructure and, more importantly, the benefits of the MCT software. Surveys were not conducted in the control schools since the questions are focused on the subject of the MCT itself, so there is unfortunately not a comparison group for the results on student attitudes.

<sup>&</sup>lt;sup>11</sup>The change to the estimation in Chile is obvious, and mirrors the change between Equations (4) and (5).

# 5 Results

#### 5.1 Balanced Sample

The results on the balance check are shown in Table 7. The table displays the t-statistic resulting from the two-sided test, with the 95% confidence interval shown in parentheses below it. A negative t-statistic denotes that treatment students, on average, had a higher pretest score or percentage male. Only one of the t-statistics are significantly different from zero. There is a higher percentage of male students in treatment than in control for low tertile group in Mexico. Overall, on both pretest scores and percentage male, the treatment students match the control students. The randomization design effectively split our sample into reliable experimental groups. The treated and control groups only differ by treatment assignment.

#### 5.2 Treatment Effects

The first outcome measure we address is whether or not the MCT is an effectice curriculum for increasing student test scores. Tables 8 and 9 shows the average pre- and posttest scores, plus the difference scores, by school and type in both countries. The difference score is the posttest score minus the pretest score. If a difference score is positive, a student scored better on the posttest than he did on the pretest. As shown by comparing both tables, the average treatment students have greater (positive) difference scores than the average control students. In Chile, every individual treatment school had a positive average difference score but one and every control school had a negative average difference score. These tables support the descriptions of Chilean public primary schools in the World Economic Forum's report - namely, that students are lagging behind and underperforming in math. The average student not using the Cognitive Tutor actually scores *lower* on a similar test after 6 months of pre-algebra instruction. In Mexico, where the randomization was done within each school by classroom, the average difference-in-difference score<sup>12</sup> by school was positive for every school.

In order to quantify this difference in performance, we turn to the results of the estimations of Equations (4) and (5), shown in Table 10. Both pre- and posttest scores are transformed into z-scores (separately by country and by pre- and posttest) with mean 0 and standard deviation 1 to make the empirical results more easily interpretable. The coefficient value on *Type* represents the increase in standard deviations of the posttest score that treated students are predicted to gain over control students. In both Chile and Mexico, there a positive and statistically significant treatment effect. Using the top line specification in both countries, where all students are included in the estimation, students who use the MCT score nearly  $\frac{1}{5}$  of a standard deviation of the posttest score higher on the posttest than their control group peers. This translates to nearly 1.2 extra problems on the 44 question exam at the end of the school year that treatment students answer correctly,

 $<sup>^{12}\</sup>mathrm{In}$  other words, the average difference score of the treated students less the average difference score of the control students.

even though both groups began the year with equivalent scores on the pretest.

Note that the coefficient estimates on *StudentPretest* are all positive and less than one. This makes sense in our context. Scores between the pretest and posttest should be highly correlated (smarter students score high on both, weaker students score low on both), and the coefficient being less than one represents regression to the mean on the posttest. An additional answer correct on the pretest would be expected to raise a student's posttest score, but by less than a full correct answer since the extra correct answer on the pretest could be a random guess unrelated to the underlying student ability that the tests are meant to ascertain.

The results by tertile in Table 10 are also illuminating. The number of observations decreases to the point of making the treatment effect in most specifications statistically insignificant.<sup>13</sup> However, the point estimate on Type is positive for every specification. The personized nature of the MCT allows students of all ability levels to improve their math test scores.

## 5.3 MCT Process Data and Test Scores

As Pane et al. (2010) explain in their paper, it is difficult to disentangle the unobserved effects of individual ability or motivation from the instructional effect of the software through a specification like Equation (6), and hence a significant  $\lambda_1$  simply helps confirm that the skills required for progression through the software are strongly related to the skills measured by the standardized tests. We generally agree with this sentiment but would stress that a positive finding here lends credence to the argument that the software itself is integral to the new curriculum. At first it would seem obvious that mastering more of the software part of the curriculum would be essential for achieving higher test scores. Unfortunately, upon reflection this relationship is not actually that obvious. Being in the treatment group meant an entire change in math teaching and learning for that school year. The incorporation of computer technology in the math courses is just one aspect of the overall shift in teaching strategy seen in the treatment classrooms, albeit the most obvious one. Students completed more group projects and teachers were encouraged and trained to act more as facilitators and coaches than as lecturers in the classroom. The software aspect of the curriculum shift, where students are sent to the computer lab for two out of the five class days each week, is the only part of the MCT treatment program for which we have data. But it is quite conceivable that other aspects of the pedagogical shift caused the treatment effect finding above, and the process of the software is not particularly integral for the treatment students' math achievement. In addition, the exams are constructed by the Chilean education ministry, not the developers of the MCT software. If, for example, the math software heavily frontloads much of the tested material but then covers nonessential topics throughout the remaining units, then students who accomplish more of the software program would likely not see higher exam scores than their peers who got through only the early units. Demonstrating that MCT mastery is aligned with standardized test mastery is a relevant finding for the efficacy of the software on its own.

<sup>&</sup>lt;sup>13</sup>We reach the 10% level of significance in Chile for the high and middle tertiles (with classroom pretest included) and in Mexico for the middle tertile (both specifications).

The z-scores are again used for the test data in these estimations. Table 11 shows the results of Equation (6) and its equivalent specification for Chile for each separate MCT variable. For convenience, only  $\lambda_1$  is shown in the table, though the rest of the estimates can be obtained from the authors. Each column of Table 11 comes from a different estimation, where only the listed MCT variable (units, total hours, sections mastered, skills mastered) is included as MCT in Equation (6). Since we do not include control students in these regressions, we do not separate students by tertile due to concerns over sample size. The results for both countries are basically the same. The total hours of usage is not significant in predicting posttest scores in Mexico (and marginally significant in the *negative* direction in Chile), but the coefficient estimates on number of units completed, skills mastered, and sections mastered are all positive and significant. This aligns with the contention by Carnegie Learning that the "mastered" variables are predictive of performance. Mastered sections and skills encompass both student effort and student ability, while units completed is reflective of effort put forth on the software portion of the curriculum.

The coefficients on sections and skills mastered in the regressions relate the marginal effect on posttest scores of increasing sections or skills mastered by one unit while controlling for school. In terms of magnitude for the software effects, consider the following. Across the 310 students in Chile. the sections mastered (skills mastered, units) variable has an average of 24.3 (252.3, 8.2) and standard deviation of 12.8 (154.2, 3.9). The regression results show that an increase in usage of the MCT software by one standard deviation of sections mastered (skills mastered, units) improves posttest scores by  $0.28 \, (0.31, \, 0.25)$  standard deviations. That is a consequential increase in achievement on the posttest. In Mexico, the sections mastered (skills mastered, units) variable has an average of 21.2 (190.3, 7.1) and standard deviation of 12.9 (144.1, 3.8). Our results there show that an increase in usage of the MCT software by one standard deviation of sections mastered (skills mastered, units) improves posttest scores by 0.40 (0.28, 0.30) standard deviations. The Bridge to Algebra MCT curriculum, taken as a whole, is effective at improving math scores for students of all abilities. Furthermore, the notion that students can expect to do better on the math exams when they have accomplished more units and mastered more of the sections and skills taught in the software is supported. The processes of the software piece of the MCT curriculum are effective at improving math score outcomes.

# 5.4 School Characteristics and Software Usage

We are interested in explaining the relationship between the school characteristics and the MCT process variables. Carnegie Learning stresses that school administrators and teachers must be on board with this curriculum change for the MCT to be effective. In addition, there must be sufficient computer access available for all students. There are large differences across schools in terms of average completion and mastery of the Bridge to Algebra MCT.

Table 12 shows the correlations between each of the four school characteristics (basic inputs, infrastructure, implementation, and learning environment) and the average value for the MCT data (usage hours, units completed, sections mastered, and skills mastered) by school. It is clear from the

table that infrastructure and implementation school ratings are highly correlated with student MCT usage. Though this result is based on just ten observations and is therefore not the strongest in this paper, we believe it has practical importance for policy considerations. Every school was expected to have one computer per student and reliable connectivity to the internet, but in reality this was not seen. Those schools which experienced this ideal infrastructure possessed an environment which allowed their students to excel, while those who adopted the MCT curriculum but did not have the ability to properly use it saw their students lag behind. Committed teachers, principals, and administrators (implementation) are central to realizing the effectiveness of the MCT. This is evidenced most strongly by teacher fidelity to pre-rollout training. Students will also master more skills using the MCT when the frequency of lab use and the conditions in the lab (learning environment) are high. School and governments considering the adoption of the MCT curriculum need to fully commit the time, energy, and resources to the endeavor. Simply sending students to an inadequate computer lab to use the software every now and then will be decidedly less effective than consistently utilizing proper facilities. The schools which most closely followed the recommended implementation activities saw their students complete and master more of the MCT software curriculum. In that light, the results presented in this section speak to the possibility that the treatment effects shown earlier are lower bounds of the true treatment effect. If all schools were able to properly take advantage of the MCT curriculum, we have reason to believe that the treatment students as a whole would have experienced even greater posttest scores.

## 5.5 Student Attitudes about MCT

Table 13 presents student responses on the four dimensions measured by multiple questions per dimension. To simplify presentation, we indicate the percentage of people across questions in a dimension who feel very positive or positive. For example, 83% of the students in Chilean school 1 asked questions about the "Ease of Use of Tutor" said they were very satisfied or satisfied on both questions within this dimension.

Our fourth and final research question addressed student attitudes toward the MCT. In general, students report high levels of satisfaction with the new curriculum and its implementation. Most of the students rate the ease of use of the Tutor, teacher help with the Tutor, and effectiveness of the Tutor highly or very highly. The process of the MCT itself seems to lead to positive dispositions toward the technology. Infrastructure gets the lowest ratings, which is not surprising since the schools are primarily in poorer areas. The lower rating of infrastructure by students when compared to other survey response areas also matches the lower values we saw for infrastructure when compared to the other school characteristics.

The attitude results are a complementary outcome to the positive treatment effects. Students feel the MCT helps them learn mathematics, they enjoy using it, and they find the teacher to be supportive. As shown in the Appendix, teachers also had a positive disposition toward the MCT curriculum.

# 6 Conclusions and Discussion

The results of this paper add to the growing body of work that investigates technology-based math curricula. It is the first to our knowledge that looks at the MCT curriculum in Central and South American schools. All previous MCT studies have focused on U.S. schools. The results presented in this paper are largely supportive of the MCT *Bridge to Algebra* curriculum for Chilean and Mexican middle school students. Schools which expressed interest in adopting the curriculum were randomly assigned to treatment or control groups.

Though they scored the same on the pretest, treatment students outperformed their control group peers on the standardized exam posttest in both Chile and Mexico. The treatment effect of the MCT curriculum is significant and positive. The overall treatment effect shows that treated students answer an additional 1.2 questions correct on the 44 question final exam than their control group peers.

The finding on treatment effects is not driven by students in just one or two treatment schools performing well, or conversely those in a few control schools severely dragging down the overall control group. In addition, the positive treatment effect exists across the student ability distribution. Though there was generally insufficient power due to limited observations, every regression specification that separated students by initial score into three groups (high scorers, middle, and low) resulted in a positive treatment effect. The interactive and personalized structure of the software part of the curriculum and the emphasis on group-based collaborative work on math projects seems to help all students, regardless of initial math ability.

Within treatment schools, we took advantage of the wealth of data provided by the MCT software. This paper is among the first to employ this data set from the MCT in evaluating student outcomes. Even the IES paper for Congress (Campuzano et al., 2009) published for the U.S. Department of Education's What Works Clearinghouse, widely considered the most comprehensive report on the effect of technology use in U.S. classrooms, only incorporated the actual time logged in (the equivalent to our "usage hours" MCT variable) from the various softwares. This is especially relevant since the IES report generally showed a lack of technology usage time mattering for achievement purposes, as measured by autside exams. We would concur with that finding. But in this paper, *accomplishing* more of the program is positively related to higher posttest scores even after controlling for school effects and pretest scores. Simply spending more time logged in to the software (usage hours) is not significantly related to posttest scores, but actually completing more of the 14 units or mastering more of the 57 sections or 552 skills was.

We developed a process for evaluating the degree of efficacy for proper implementation in schools, and a way to rate four different characteristics of schools that matter in the implementation of the MCT. Those schools which were more prepared to handle the demands of this new curriculum saw their students accomplish more of it. Though this finding seems completely intuitive, we believe it is worth particular emphasis. Schools with sufficient computers, reliable internet connections, and committed principals and teachers saw students accomplish more of the software program. In turn, those students could be expected to achieve higher scores on the exams. The adoption of the MCT curriculum requires large investments in time and money. School administrators need to prepare their teaching staff and supply enough technological infrastructure to realize the educational benefits of the investment. With the proper inputs, the processes of the MCT curriculum, including both the software component and the in-class changes, lead to student improvements in math abilities and enjoyment of the coursework. The take away message for schools considering the MCT is best summed up by the admonition that if you are going to do something, do it right.

There are some alternative explanations to the results in this paper, though we have tried to mitigate them to the best of our ability. It is possible there is a large selection effect occurring here, and that these findings are not broadly reflective of the results a random school would find if they adopted the MCT. In other words, the findings are constrained to those schools willing to participate. In large part, we do not disagree with that statement. This is a substantial change from a traditional textbook- and lecture-based curriculum. School administrators who are unwilling to make the effort to ensure the best possible learning environment and computer facilities very well may not see a positive treatment effect. This is supported by the results on school characteristics even when considering schools who wanted to adopt the new curriculum and ended up doing just that. In addition, this concern is often encountered in education policy papers. Due to the randomization within the group of schools willing to participate in treatment, we believe that our results show high internal validity. The required tradeoff in a clean experimental design often necessitates concerns over external validity. The inclusion of both random and fixed effects through HLM (such as in Equation (4)) is the most that can be done in this setting to show broader applicability.

The amount of missing data was not terribly disconcerting in this study. Mexican school 4 did not return our surveys. Otherwise, we matched enrollment figures to test scores to MCT data for a very high percentage of students. For example, in Chilean treatment schools, our original enrollment-based target was 313 students. We ended up with two test scores (pre and post) and MCT data for 310 of those same students. The overall rate of missing information is similar to other education studies that actually report on the issue, if not better. We were able to match almost every treatment student with an initial and final test score to his software data. We observed no students transferring from the treatment condition to the control condition, or vice versa.

The results presented here are robust to other specifications, including just school- or classroomrandom effects models (instead of incorporating HLM, which includes fixed effects). We also estimated the treatment effect model using difference scores as the dependent variable and excluding the independent pretest scores on the right hand side. The results are basically the same, which is not surprising considering that we had initially balanced samples based on pretest scores.

It is possible that these results are driven largely by teacher effects, rather than any effect from the curriculum itself. We unfortunately have no information on the teachers other than their names, so it is impossible to check balance between treatment and control groups on teacher observables (such as years teaching or highest degree attained) or previous teacher-student output (such as last year's student standardized test scores by teacher). However, we have no reason to believe there are large discrepancies here. In fact, in all the Mexican schools and some of the Chilean ones, the same teacher taught all the classes within the same school. If there are some pedagogical learning gains from incorporating the MCT curriculum, this would result in mitigated treatment effects since the control groups would receive some benefit, too.

The attitude data provides an additional way to think about the effectiveness of the MCT. The reported ease of use seems consistent with the actual improvements in math scores. The positive responses in Table 13 suggest that the students would be open to future classroom uses of the software. Reporting positive outcomes for both test scores and attitudes represents a more comprehensive picture of effectiveness.

There were specific research questions posed earlier in this paper. We have shown the following:

- 1. The MCT improves math performance for treated students over control students.
- 2. All process indicators except for usage hours are significantly and positively related to math performance indicators. Accomplishing more of the MCT curriculum is predictive of larger posttest scores for students in treatment.
- 3. Better school characteristics, especially infrastructure and implementation, lead to increases in MCT completion and mastery.
- 4. Students are able to understand and use the MCT on their own and receive help from the teacher when it is needed. The students also believe that the MCT is an effective tool for learning math at the pre-algebra level.

Our initial strategy was to see how the MCT impacted learning of mathematics. Our road map for future research is to examine more closely the effect of the MCT conditional on school, teacher, and student characteristics and to identify how they contribute to improving math performance. We would also like to incorporate more details from the software, such as hint-seeking behavior of the students, in a fashion similar to Equation (6). Finally, we need to learn how performance in a tutorbased class impacts on math performance in subsequent years. The IES reports led by Campuzano (2009) and Dynarski (2007) have shown that many math achievement gains using technology-infused curricula fade after the initial year unless the students continue using similar curricula through future grades.

Table 1:	Demonstration	of MCT	Skills	Breakdown

				~				
What	is	the greatest	$\operatorname{com}$	nmon fa	ctor o	of 27	and	18?
Table	1:	Demonstrati	ion o	of MCT	Skill	ls Bre	eakdo	wn

Step Description	Skill	Skill Description
List factors of 27	1	Factor a number
List factors of 18	1	Factor a number
Identify common factor of 27 and 18	2	Choose common numbers between sets
Identify greatest common factor of 27 and 18	3	Choose the greatest number from a set
The print of the second se		there

					<u>_</u>
	Table 2	2: Student Observatio	ons in C	bile	
School Num	Type	Classroom Num	Obs	Pretest	Gender
1	Treatment	1	30	18.8 (3.5)	0.37(0.49)
		2	31	21.3 (4.9)	0.55(0.51)
		3	32	18.7 (4.3)	0.59 (0.50)
2	Treatment	1	25	21.8(5.0)	0.64(0.49)
		2	19	17.5(4.93)	$0.42 \ (0.51)$
3	Treatment	1	21	19.3(5.5)	0.67(0.48)
4	Treatment	1	34	22.8(6.3)	0.50(0.51)
		2	31	24.0(6.3)	$0.52 \ (0.51)$
5	Treatment	1	21	15.6(3.9)	0.29(0.46)
		2	21	17.6(3.9)	0.81(0.40)
		3	21	16.6 (4.6)	$0.52 \ (0.51)$
6	Treatment	A	24	21.3(4.6)	0.63(0.49)
7	Control	1	28	17.2(4.3)	0.46(0.51)
			29	18.0(4.2)	0.52(0.51)
		3	23	17.1(5.1)	0.30(0.47)
8	Control	1	25	20.9(5.6)	0.64(0.49)
		2	23	17.7 (4.3)	0.48(0.51)
A		3	25	19.0~(6.3)	0.72(0.46)
9	Control	1	32	24.3(5.5)	$0.53\ (0.51)$
× 1		2	31	24.6(5.7)	$0.45\ (0.51)$
		3	29	23.3(5.8)	0.59(0.50)
10	Control	1	18	18.0(5.5)	0.61(0.50)
		2	21	20.8(6.2)	0.67(0.48)
11	Control	1	23	15.5(5.2)	0.57(0.51)
		2	24	18.4(3.5)	0.42(0.50)
12	Control	1	27	18.6 (4.3)	0.37(0.49)

Table 2: Student Observations in Chile

School Num	Type	Classroom Num	$\mathbf{Obs}$	Pretest	Gender		
	Treatment	1	40	23.3(5.9)	0.45(0.50)		
1	Control	2	41	21.3(6.0)	0.59(0.49)		
	Control	3	37	21.7(5.6)	0.65(0.48)		
	Control	4	38	20.8(5.8)	0.68(0.47)		
	Treatment	1	41	25.0 (6.0)	0.41 (0.50)		
2	Control	2	44	25.3(5.4)	0.36(0.49)		
	Control	3	39	26.4(6.9)	0.46(0.51)		
	Control	4	40	24.4(6.1)	0.35(0.48)		
	Treatment	1	36	23.6(5.1)	0.50(0.51)		
3	Control	2	41	22.0(5.2)	0.42(0.50)		
	Control	3	39	24.9(6.0)	$0.46\ (0.51)$		
	Control	4	39	21.5(5.1)	$0.46\ (0.51)$		
	Treatment	1, U	39	18.1(6.9)	0.74(0.44)		
4	Control	2	44	20.5~(6.0)	0.82(0.39)		
	Control	$\sim$	37	20.7~(6.4)	0.78(0.42)		
	Control	4	39	17.8(5.6)	0 (N/A)		

Table 3: Student Observations in Mexico

		Contro			37	20.7(6.4)
		Contro			39	17.8(5.6)
	PT PT	Table 4	: Characteristic	s by Type -	- Agg	gregated
	$\sim$					5 .0
		Country		Treatme	ent	Control
R		Country Chile	Observations	Treatme 310	ent	Control 358
DR.		Country Chile	Observations Pretest	<b>Treatme</b> 310 19.9 (5.5	ent 5)	Control 358 19.8 (5.8)
DR		Country Chile	Observations Pretest Gender	Treatme           310           19.9 (5.5           0.54 (0.50	ent 5) 0)	Control           358           19.8 (5.8)           0.52 (0.50)
DR		Country Chile Mexico	Observations Pretest Gender Observations	<b>Treatme</b> 310 19.9 (5.5 0.54 (0.50 156	ent 5) 0)	Control           358           19.8 (5.8)           0.52 (0.50)           478
OR		Country Chile Mexico	Observations Pretest Gender Observations Pretest	<b>Treatme</b> 310 19.9 (5.5 0.54 (0.50 156 22.5 (6.5	ent (5) (0) (5) (5) (5) (5) (5) (5) (5) (5) (5) (5	Control           358           19.8 (5.8)           0.52 (0.50)           478           22.3 (6.3)

Country	School, Class	Usage Hrs	Units	Sections Mas	Skills Mas
	1, 1	17.8(2.8)	11.1(2.9)	31.4(12.0)	360.5 (141.1)
Chile	1, 2	17.9(3.6)	12.5(2.1)	36.8(10.1)	428.8 (104.3)
	1, 3	17.8(3.0)	9.2(2.8)	25.0(10.9)	275.4(126.3)
	2, 1	8.5(2.4)	5.9(3.5)	18.8(11.6)	174.5 (130.0)
	2, 2	9.4(2.8)	5.3(3.6)	16.8(11.2)	154.1(125.6)
	3, 1	10.3 (1.8)	8.5(3.5)	26.1 (13.4)	259.2 (153.0)
	4, 1	8.7(2.4)	6.6(2.8)	22.1 (10.8)	196.0 (114.4)
	4, 2	8.7(2.6)	6.8(3.9)	22.8 (13.7)	207.3(152.7)
	5, 1	13.2 (3.5)	7.2(2.9)	17.5(7.5)	195.5 (99.8)
	5, 2	11.5(2.2)	6.6(2.8)	19.6(9.3)	189.2 (106.9)
	5, 3	7.9(2.2)	4.9(2.9)	15.0 (10.0)	133.8(102.2)
	6, 1	18.6(2.8)	11.1 (2.9)	31.3(12.5)	357.6(143.5)
	All students	12.8(5.0)	8.2(3.9)	24.3(12.8)	252.3(154.2)
	1, 1	7.8(1.3)	9.0(2.4)	27.5(10.1)	249.5 (115.0)
Mexico	2, 1	5.8 (1.9)	6.3(2.3)	18.2(7.5)	136.1(70.1)
	3, 1	14.3(3.4)	11.3(1.6)	34.0(8.4)	351.8(103.5)
	4, 1	1.3(0.77)	2.4(1.0)	6.2(3.8)	37.6(24.5)
	All students	7.2(5.0)	7.1(3.8)	21.2(12.9)	190.3(144.1)

Table 5: MCT Process Data

Table 6: School (	Characteristics
-------------------	-----------------

	All stud	ents	7.2(5.0)	7.1 (3	3.8) 2	21.2(12.9)	190
	TH	Tab	le 6: Schoo	ol Chara	cteristics		
	Country	School	Basic	Infra	Imple	Learn	Total
	/		Inputs			Env't	
	·	1	3	3	5	4	15
×,	Chile	2	2	1	3	2	8
P.J.		3	3	4	4	3	14
$\mathcal{O}_{\mathcal{F}}$		4	5	3	5	4	17
Y		5	3	3	5	5	16
		6	5	4	5	5	19
		1	5	5	4	5	19
	Mexico	2	4	1	5	5	15
		3	5	5	5	5	20
		4	1	1	3	3	8

Country	Students	Pretest	Percent Male
	All	-0.38	-0.49
		(-1.02, 0.69)	(-0.10, 0.06)
Chile	High Tertile	0.10	1.38
		(-1.22, 1.35)	(-0.04, 0.25)
	Middle Tertile	0.23	-1.91
		(-0.68, 0.86)	(-0.25, 0.00)
	Low Tertile	-1.00	-0.04
		(-1.05, 0.34)	(-0.13, 0.12)
	All	-0.36	-0.51
		(-1.39, 0.96)	(-0.11, 0.07)
Mexico	High Tertile	-0.94	1.01
		(-1.72, 0.61)	$\bigcirc$ (-0.08, 0.25)
	Middle Tertile	-0.99	1.70
		(-1.26, 0.42)	(-0.02, 0.29)
	Low Tertile	0.01	-3.61
		(-1.13, 1.14)	(-0.40, -0.12)

 Table 7: Balance Check

t-statistic shown above

95% confidence interval in parentheses below

C - V					
Fable 8:	Test S	Scores	for	Treatment	Students

	Country	School	Pretest	Posttest	Difference Score
		1	19.57(4.40)	19.58(5.84)	0.01 (5.09)
	Chile	2	$19.91 \ (5.37)$	21.14(6.19)	1.23(5.45)
,		3	19.33(5.46)	22.62(7.24)	3.29(5.52)
	È Z	4	$23.37 \ (6.29)$	24.22(7.62)	0.85(5.27)
	· ·	5	16.60(4.16)	16.97 (4.50)	$0.37 \ (4.54)$
N.		6	21.33(4.56)	21.08(4.22)	-0.25(5.08)
$\mathbf{\nabla}$		All students	19.93(5.47)	20.56(6.51)	0.63(5.13)
		1	23.28(5.93)	28.33(7.04)	5.05(4.38)
	Mexico	2	25.05(5.97)	27.95(6.57)	2.90(4.89)
		3	23.64(5.13)	22.56(6.39)	-1.08(4.76)
		4	18.08(6.88)	19.10(7.56)	1.03(4.36)
		All students	22.53(6.53)	24.60(7.87)	2.06(5.08)

Average (standard deviation)

Country	School	Pretest	Posttest	Difference Score	
	7	17.46(4.45)	17.28 (4.07)	-0.19 (4.94)	<u> </u>
Chile	8	19.23(5.57)	18.45(6.63)	-0.78(5.46)	
	9	24.08(5.63)	23.71 (7.11)	-0.37(5.11)	$\langle \cdot \rangle$
	10	$19.51 \ (6.00)$	17.87(6.36)	-1.64 (6.66)	
	11	17.00(4.60)	15.89(4.13)	-1.11 (4.90)	
	12	18.56(4.30)	18.11(5.57)	-0.44 (5.59)	
	All students	19.77(5.81)	19.11(6.49)	-0.65(5.32)	
	1	21.28(5.75)	23.69(7.44)	2.41(6.07)	
Mexico	2	25.37(6.15)	27.26(6.74)	1.89(4.14)	
	3	22.82(5.59)	22.15(6.16)	-0.66 (4.46)	
	4	19.68 (6.10)	18.48 (6.74)	-1.20 (5.46)	
	All students	22.31 (6.26)	22.92 (7.47)	$0.60\ (5.30)$	

Table 9: Test Scores for Control Students

Average above, standard deviation in parentheses below.

Table 10: Treatment Effects

Country	Students	Class Pretest	Student Pretest	Gender	Type
	All		0.58 (0.03) ***	$0.06 \ (0.05)$	0.18 (0.09) **
Chile	All	0.29 (0.11) ***	$0.56 \ (0.03) \ ***$	$0.06\ (0.05)$	0.17 (0.08) **
	High Tertile	6	0.88 (0.09) ***	0.05~(0.10)	$0.26 \ (0.17)$
	High Tertile	0.35 (0.22)	0.82 (0.11) ***	$0.03 \ (0.10)$	0.23~(0.13) *
	Middle Tertile		0.66 (0.10) ***	$0.09\ (0.07)$	0.17 (0.10)
	Middle Tertile	$0.24 \ (0.23)$	0.49 (0.18) ***	$0.09\ (0.07)$	0.17~(0.10) *
	Low Tertile		0.55 (0.09) ***	$0.09\ (0.07)$	0.14(0.10)
	Low Tertile	$0.20 \ (0.15)$	0.47 (0.11) ***	$0.08\ (0.07)$	0.14(0.10)
X	All		0.67 (0.03) ***	-0.02(0.05)	0.19 (0.09) **
Mexico	All	$0.25 \ (0.16)$	$0.67 \ (0.03) \ ***$	-0.03(0.05)	0.18 (0.10) *
	High Tertile		$0.72 \ (0.09) \ ***$	$0.10 \ (0.09)$	0.19(0.14)
	High Tertile	$0.32 \ (0.27)$	0.67 (0.10) ***	$0.10 \ (0.09)$	0.18(0.14)
	Middle Tertile		$0.84 \ (0.13) \ ^{***}$	-0.15 (0.09) *	0.34~(0.18) *
	Middle Tertile	-0.01 (0.31)	0.84 (0.17) ***	-0.15(0.09)	0.34~(0.18) *
	Low Tertile		0.48 (0.10) ***	-0.06(0.09)	$0.09 \ (0.15)$
	Low Tertile	0.62 (0.23) ***	0.40 (0.10) ***	-0.07(0.09)	0.08~(0.13)

Estimated std errors are reported in parentheses. Significance denoted as \*\*\*1%, \*\*5%, \*10%

Table 11:	MCT	and	Test	Scores
-----------	-----	-----	------	--------

Country	MCT Data						
	Usage Hrs	Units	Skills Mas	Sections Mas			
Chile	-0.016 (0.009) *	0.064 (0.012) ***	0.002 (0.0003) ***	0.022 (0.003) ***			
Mexico	-0.027 (0.020)	0.080 (0.029) ***	0.002 (0.0006) ***	0.031 (0.007) ***			
Estimated std errors are reported in parentheses.							
Significanc	e denoted as $***1^{\circ}_{\prime}$	%, **5%, *10%					
DEDE							
r	Table 12: Correlati	ons of School Chara	cteristics and Softwa	re Usage			

	Basic Inputs	Infrastructure	Implementation	Learning Env't
Usage Hrs	0.22	0.55	0.60	0.33
Units	0.54	0.78	0.61	0.49
Skills Mas	0.45	0.76	0.57	0.39
Sections Mas	0.61	0.80	0.58	0.44

 Table 12: Correlations of School Characteristics and Software Usage

	ASE	20	
~		Table 13: Student S	urveys

	Country	School	Ease of Use	Teach Help	Infrastructure	Effectiveness
		1	83%	92%	22%	90%
	Chile	2	71%	83%	66%	90%
		3	87%	86%	82%	88%
	K.	4	83%	80%	55%	87%
$\mathbf{i}$	)	5	89%	94%	60%	89%
		6	90%	95%	84%	93%
		1	86%	91%	37%	88%
	Mexico	2	48%	34%	48%	52%
		3	64%	85%	13%	85%

Percentage answering high/very high in aggregate

# 7 References

Anderson, J. (1993). Rules of the Mind. Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J. (1993). "Problem Solving and Learning." American Psychologist, 48, 35-44.

Anderson, J. (2002). "Spanning Seven Orders of Magnitude: A Challenge for Cognitive Modeling." Cognitive Science, 85-112.

Anderson, J., & Schooler, L. (1991). "Reflection of the Environment in Memory." *Psychological Science*, 2, 396-408.

Anderson, J. & Schunn, C. (2000). "Implications of the ACT-R Learning Theory: No Magic Bullets." In R. Glaser, *Advances in Instructional Psychology (Vol. 5)*. Mahwah, NJ: Lawrence Erlbaum Associates.

Arroyo, I., Woolf, B., Royer, J., Tai, M., & English, S. (2010). "Improving Math Learning Through Intelligent Tutoring and Basic Skills Training." *Intelligent Tutoring Systems Part II: 10th International Conference, ITS 2010*, 423-432. Pittsburgh, PA: Springer Berlin.

Banerjee, A., Cole, S., Duflo, E., & Linden, L. (2007). "Remedying Education: Evidence from Two Randomized Experiments in India." *Quarterly Journal of Economics*, 1235-64.

Cabalo, J., Jaciw, A., & Vu, M. (2007). Comparative Effectiveness of Carnegie Learning's Congitive Tutor Algebra I Curriculum: A Report of a Randomized Experiment in Maui School District. Palo Alto, CA: Empirical Education.

Campuzano, L., Dynarski, M., Agodini, R., Rall, K., & Pendleton, A. (2009). Effectiveness of Reading and Mathematics Software Products: Findings From Two Student Cohorts. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education

Casas, I., Goodman, P., & Pelaez, E. (2011). "On the Design and Use of a Cognitive Tutoring System in the Math Classroom." *Technology for Education (T4E), 2011 IEEE International Conference*, 9-17. Chennai, India: Curran.

Casas, I., Imbrogno, J., & Vergara, A. (2013). "A Cognitive Tutoring Strategy for Math Teaching and Learning in Latin America." Manuscript submitted for publication.

Chong, A. (2011). Development Connections: Unveiling the Impact of New Information Technologies. New York, NY: Palgrave MacMillan.

Cook, T. (2003). "Why Have Education Evaluators Chosen Not to Do Randomized Experiments?" The ANNALS of the American Academy of Political and Social Sciences, 114-149.

de Ferranti, D., Perry, G., Gill, I., Guasch, L., Malloney, W., Sanchez, C., & Schady, N. (2003). Closing the Gap in Education and Technology. Washington, DC: The World Bank.

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., & Campuzano, L. (2007). *Effectiveness of Reading and Mathematics Software Products: Findings From the First Student Cohort*. Washington, DC: National Center for Educational Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Hanushek, E., & Woessman, L. (2008). "The Role of Cognitive Skills in Economic Development." *Journal of Economic Literature*, 607-688.

Hanushek, E., & Woessman, L. (2009). Schooling, Cognitive Skills, and the Latin American Growth Puzzle. Cambridge, MA: National Bureau of Economic Research.

Koedinger, K., & Anderson, J. (1993). "Effective Use of Intelligent Software in High School Math Classrooms." *Proceedings of the AIED 1993 World Conference on Artificial Intelligence in Education*, 241-248. Charlottesville, VA: Association for the Advancement of Computing in Education.

Koedinger, K., Anderson, J., Hadley, W., & Mark, M. (1997). "Intelligent Tutoring Goes to School in the Big City." *International Journal of Artifical Intelligence in Education*, 30-43.

National Research Council. (2007). *Rising Above the Gathering Storm*. Washington, DC: National Academy Press.

Newell, A., & Rosenbloom, P. (1981)." Mechanisms of Skills Acquisition and the Law of Practice." In J.R. Anderson (Ed.), *Cognitive Skills and Their Acquisition*, 1-55. Hillsdale, NJ: Lawrence Erlbaum Associates.

Pane, J., McCaffrey, D., Steele, J., Ikemoto, G., & Slaughter, M. (2010). "An Experiment to Evaluate the Efficacy of Cognitive Tutor Geometry." *Journal of Research in Educational Effectiveness*, 254-281.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical Linear Models*. Thousand Oaks, CA: Sage Publications.

Ritter, S. (2011). The Research Behind the Carnegie Learning Math Series. Pittsburgh, PA: Carnegie Learning.

Ritter, S., & Morgan, P. (2002). An Experimental Study of the Effects of Cognitive Tutor Algebra I on Student Knowledge and Attitude. Pittsburgh, PA: Carnegie Learning.

Ritter, S., Anderson, J., Koedinger, K., & Corbett, A. (2007). "The Cognitive Tutor: Applied Research in Mathematics Education." *Psychonomics Bulletin and Review*, 249-255.

Scheurman, F., & Pedro, F. (2009). Assessing the Effects of ICT in Education: Indicators, Criteria, and Benchmarks for International Comparisons. Paris, France: European Union.

Schwab, K. (2010). The Global Competitiveness Report: 2009-2010. Bern, Switzerland: World Economic Forum.

# 8 Appendix

#### Survey for Teachers

The survey questions asked to teachers are shown here. The one surprising result is that teachers do not agree with the students about the computers and internet connections. The teachers were quite positive on questions 1.1 and 1.2, while the student responses for the same questions (captured by "infrastructure" in their survey results) was not high. For example, in school 1, only 22% of the students believed that the computers and internet connections worked reliably, while their teacher responded in the affirmative.

Topic 1: Indicate your agreement with the following statements about using computers

- 1. Computers worked adequately in the lab
- 2. Internet connection worked well in the lab
- 3. Computers made the students' work easier in the lab

Topic 2: Indicate your agreement or disagreement with the following statements about the Cognitive Tutor

- 1. The Cognitive Tutor is a very useful resource to teach math
- 2. The Cognitive Tutor software is a useful resource for learn math
- 3. In the future I would like to have the Cognitive Tutor implemented in my math classes
- 4. Textbooks used (teacher's manual) were really helpful as a support for carrying out the classes

Topic 3: Indicate how comfortable you felt using the Cognitive Tutor software in your math classes

Topic 4: Indicate how valuable the combined collaborative work and individual use of the Cognitive Tutor in the lab were

Topic 5: Indicate, from your perspective, the degree of effectiveness of the Cognitive Tutor to show the performance of your students in math

Topic 6: In general, indicate how satisfied you are with the training you received to use the Cognitive Tutor

Topic 7: In general, indicate how useful the training was that you received to use the Cognitive Tutor

Topic 8: Regarding the training, indicate how satisfied you are with

- 1. Clarity of activities and contents
- 2. The level of proficiency obtained to use the Cognitive Tutor
- 3. The level of proficiency obtained to get the Cognitive Tutor running in all its aspects
- 4. Textbooks used (teacher's manual) were really helpful as a support for carrying out the classes

Topic 9: Indicate how valuable the planning of the classes was during your training and Cognitive Tutor support

Topic 10: Which suggestions do you have to get better training and Cognitive Tutor support

Topic 11: Indicate the degree of difficulty you encountered in implementing the Cognitive Tutor in your math classes

Topic 12: From your perspective, how would you grade the implementation of the Cognitive Tutor

Topic 13: Indicate your suggestions for a more successful implementation of the Cognitive Tutor project

Table 20 shows the exact teacher survey responses.<sup>14</sup> School 9 is the only school with multiple teachers. Open-ended responses are again not shown in the table.

 $^{14}$ Topics 1 and 2 are measured on a 6 point scale from strongly disagree (1) to strongly agree (6). Topic 3 is measured on a 5 point scale from not comfortable (1) to very comfortable (5). Topics 4 and 9 are measured on a 5 point scale from not valuable (1) to very valuable (5). Topic 5 is measured on a 5 point scale from ineffective (1) to very effective (5). Topics 6 and 8 are measured on a 6 point scale from very dissatisfied (1) to very satisfied (6). Topic 7 is measured on a 5 point scale from not useful (1) to very useful (5). Topic 9 is measured on a 5 point scale from not valuable (1) to very valuable (5). Topics 10 and 13 are open responses. Topic 11 is measured on a 5 point scale from not difficult (1) to very difficult (5). Topic 12 is measured on a 5 point scale from very bad (1) to very good (5).

								R
Table 14	· Distribution	of	Surv	ov A	net	vore	by '	Teachers
	. Distribution			cy 1		-	by	
		S	cho	ol I	Nui	nbe	r	
	Topic. Ques	1	2	3	4	5	6	
	1.1	5	4	6	5	5	$5^{\prime\prime}$	C.r
	1.2	5	4	6	5	$5 \checkmark$	6	7
	1.3	6	5	6	6	5	6	
	2.1	6	5	5	6	6	6	
	2.2	5	5	5	6	5	6	
	2.3	6	6	5	6	6	6	
	2.4	6	4	5	6	5	6	
	3.1	5	5	5	5	5	5	
	4.1	5	3	4	5	4	5	
	5.1	4	3	4	5	3	5	
	6,1	6	6	5	6	5	6	
	7.1	5	5	5	5	4	5	
S	8.1	6	6	6	6	6	5	
	8.2	6	6	6	6	6	5	
	8.3	6	5	5	6	6	5	
$\langle \langle \gamma \rangle$	9.1	4	-	4	5	4	3	
	10.1		Op	en a	ansv	ver		
	11.1	1	3	-	1	2	2	
	12.1	4	4	3	5	5	5	
Ri	13.1		Or	en a	ansv	ver		
$\langle \rangle^{\gamma}$								