Federal Reserve Bank of Chicago

Technology's Edge: The Educational Benefits of Computer-Aided Instruction

Lisa Barrow, Lisa Markman, and Cecilia Elena Rouse

WP 2007-17

Comments Welcome

Technology's Edge: The Educational Benefits of Computer-Aided Instruction

By

Lisa Barrow
Federal Reserve Bank of Chicago

Lisa Markman
Princeton University

Cecilia Elena Rouse
Princeton University and NBER

October, 2007

We thank the many dedicated principals, teachers and staff of the school districts that participated in this project as well as Gadi Barlevy, Thomas Cook, Jonas Fisher, Jean Grossman, Brandi Jeffs, Alan Krueger, Lisa Krueger, Sean Reardon, Jesse Rothstein, Pei Zhu, and seminar participants at Columbia University, Duke University, the Federal Reserve Bank of Chicago, McMaster University, Queens University, and the University of Notre Dame for helpful conversations and comments. Benjamin Kaplan, Katherine Meckel, Kyung-Hong Park, Ana Rocca, and Nathan Wozny provided expert research assistance. Funding for this project was generously provided by the Education Research Section at Princeton University. Any views expressed in this paper do not necessarily reflect those of the Federal Reserve Bank of Chicago or the Federal Reserve System. Any errors are ours.

Abstract

Because a significant portion of U.S. students lacks critical mathematic skills, schools across the country are investing heavily in computerized curriculums as a way to enhance education output, even though there is surprisingly little evidence that they actually improve student achievement. In this paper we present results from a randomized study in three urban school districts of a well-defined use of computers in schools: a popular instructional computer program which is designed to teach pre-algebra and algebra. We assess the impact of the program using statewide tests that cover a range of math skills and tests designed specifically to target pre-algebra and algebra skills. We find that students randomly assigned to computer-aided instruction score at least 0.17 of a standard deviation higher on a pre-algebra/algebra test than students randomly assigned to traditional instruction. We hypothesize that the effectiveness arises from increased individualized instruction as the effects appear larger for students in larger classes and those in classes in which students are frequently absent.

I. INTRODUCTION

Mathematical achievement is arguably critical both to individuals and to the future of the U.S. economy. For example, research by Grogger (1996) and Murnane, Willet, and Levy (1995) suggests that math skills may account for a large portion of wage inequality including the African-American-white wage gap. And yet, in spite of recent progress, levels of proficiency remain dramatically low (U.S. Dept. of Education, 2006 – *National Assessment of Educational Progress* (NAEP) report). Compounding the problem of poor mathematics performance is the fact that many school districts report difficulty recruiting and retaining teachers, particularly in the fields of math and science, where schools must compete with (non-education) private sector salaries (Murnane and Steele 2007). While the evidence on the importance of teacher qualifications on student achievement is mixed in many subjects, the students of more qualified math teachers appear to perform better (See, e.g., Braswell et al. 2001, Boyd et al 2007).

In response policymakers, parents, and schools are actively seeking creative and effective approaches to improving students' math skills. And, not surprisingly, many school districts are turning to advances in computer technology. By 2003 nearly all public schools had access to the internet, and the number of public school students per instructional computer with internet access had fallen from 12.1 in 1998 to 4.4.[1] Despite this trend, research on the success of computer technology in the classroom has yielded mixed evidence at best. In economics most studies have focused on the impact of subsidies for schools to invest in computer technology. For example, Angrist and Lavy (2002) show a decrease in math achievement among 8th graders after the introduction of a computer adoption program in Israeli schools. Goolsbee and Guryan (2006) study the impact of the E-rate – a program to subsidize school investment in the internet –

---

[1] Table 416 of the *Digest of Education Statistics: 2006*.

and conclude that while it has substantially increased internet investment, it has had no significant impact on student achievement thus far. In contrast, Machin, McNally, and Silva (forthcoming) find that a government program to encourage investment in information and computer technology in schools in the United Kingdom led to improved performance in English and possibly science but not in math in primary schools. While it is important to understand whether and how public subsidies are used and whether they achieve their intended goals, because the use of computers by the schools in these studies is either unknown or vaguely defined, they do not provide direct evidence on the effectiveness of computer technology as an input in the education production function.

Other literature has studied the impact of computer technology on student achievement more directly.[2] A relatively recent study of the NELS88 data showed that multimedia and calculating aids had a strong positive correlation with math achievement while it had little to no effect in any other subject (Wang, Wang, and Ye 2002). In contrast, Wenglinsky (1998) finds that, on average, computer use in math instruction is negatively related to student math achievement in the 8th grade. A potential problem with this second group is that there are few studies that use a randomized controlled study design, or employ a credible strategy for controlling for factors such as individual teacher effects and student ability, that might be

[2] Kirkpatrick and Cuban (1998) define three uses of computers in instruction: computer-assisted instruction (CAI), computer-managed instruction (CMI), and computer-enhanced instruction (CEI). CAI provides drill exercises and tutorials. CMI is more elaborate in diagnosing areas in which students need more instruction, guiding students in their own learning, and recording progress for the teacher. CEI uses the Internet or other computer programs, such as graphics or word-processing, to enhance lessons and projects directed by the teacher. The type of computerized instruction we study is best characterized as computer-aided instruction, although it also contains elements of computer-managed instruction. We use the terms computer-aided instruction and computerized instruction interchangeably.

correlated with both use of computers in the classroom and student outcomes.[3] For example, given that computer technology may be used either to help poorly performing students or to enhance the learning of high achievers, it is unclear whether selection bias would generate upward or downward biased estimates of the average impact of computers technology on student achievement in poorly designed studies.

Three notable exceptions include a randomized evaluation of computer-assisted instruction conducted in the late 1970s by the Educational Testing Service and the Los Angeles Unified School District that consisted of drill and practice sessions in mathematics, reading, and language arts (Ragosta et al., 1982); the study found educationally large effects in math and reading. More recently, using a randomized study design Banerjee et al. (2005) conclude that computer-assisted mathematics instruction boosted the math scores of fourth-grade students in Vadodara, India. In contrast, after randomly assigning students to be trained using a computer program known as *Fast ForWord*, which is designed to improve language and reading skills, Rouse and Krueger (2004) conclude that while use of the computer program may have improved some aspects of students' language skills, such gains did not appear to translate into a broader measure of language acquisition or into actual reading skills. Overall, one can conclude that this literature is also mixed, although there may be more support for the effectiveness of computer technology in the instruction of math than in reading. Notably, however, few studies offer

[3] In an oft-cited, and somewhat controversial, review of the literature, Cuban (2001) concludes, "When it comes to higher teacher and student productivity and a transformation of teaching and learning … there is little ambiguity. Both must be tagged as failures. Computers have been oversold and underused, at least for now." (p. 179). Others argue for a more nuanced view of the literature that computers can be effective in certain situations, such as when used by teachers with skill and experience in using computers themselves (see, e.g., Brooks (2000)).

evidence on why the technology may help or hinder student achievement and the most recent evidence for math may not apply to U.S. students.

In this paper we present results from a new randomized study in three urban school districts in the U.S. of a well-defined use of computers in schools: a popular instructional computer program which is designed to improve pre-algebra and algebra skills. We assess the impact of the program using both statewide tests that cover a range of math skills and tests designed specifically to target pre-algebra and algebra skills. We find that students randomly assigned to classes using the computer lab score at least 0.17 of a standard deviation higher on tests of pre-algebra and algebra achievement than students assigned to traditional classrooms. The estimated effect rises to 0.25 of a standard deviation when we estimate the effect for students who actually use the computer-aided instruction. We find some evidence for the hypothesis that the effectiveness arises from increased individualized instruction as the effects appear larger for students in larger classes and those in classes in which students have poor attendance records.

In the next section we discuss why and in which circumstances CAI may be more effective than traditional instruction. Section III presents the empirical model, research design and data. Section IV presents the results, in Section V we evaluate the cost effectiveness of CAI, and Section VI concludes.

II. WHY MIGHT CAI BE MORE EFFECTIVE THAN TRADITIONAL INSTRUCTION?

A key question is why CAI may be more effective than traditional classroom teaching, on average. Some classroom research suggests computers can offer highly individualized instruction and allow students to learn at their own pace (e.g. Lepper and Gurtner 1989, Means and Olson 1995, Sandholz et al 1997, Heath and Ravits 2001). While we do not have a direct test, we hypothesize that if CAI allows for more individualized instruction, then it may be more beneficial for struggling students who cannot keep up with the pace of the lectures in traditional classrooms or for more advanced students who could progress faster at their own pace.[4] Further, we might expect CAI to be more effective for students with poorer rates of attendance. In a traditional classroom, students missing class will miss all of the material covered in class that day. In contrast, the computer always picks up where the student left off the last time she was in class regardless of whether it was the day before or 5 days before. Similarly, in classes in which many students have poor attendance records or in larger classes, we might expect a bigger effect of CAI as teachers would struggle to find the appropriate level at which to pitch lectures. Finally, one might think that individualized instruction provided by CAI avoids some of the disruption effects of having peers with poor attendance rates or being in larger classes as modeled by Lazear (2001).

More formally we can follow Brown and Saks (1984) and think of the teacher as allocating class time to different types of instruction. In the traditional classroom, the teacher divides class time between group instruction time, T_G, and individual instruction time, T_i, such that,

$$T_G + \sum_i T_i \leq \overline{T}, \qquad (1)$$

---

[4] Other forms of self-paced instruction may offer a similar educational advantage. However, a very small, older, literature suggests that computerized self-paced instruction is more effective than other self-paced instruction. See, e.g., Enochs, Handley, and Wollenberg (1986) and Surber et al (1977) for randomized studies involving college-age students.

where $\overline{T}$ is the total class time available. Thus, total instruction time for student *i* equals

$$T_G + T_i \leq \overline{T}. \tag{2}$$

As long as other students in the class receive some individual instruction time, the total instruction time for student *i* is strictly less than the total class time available.

In the CAI classroom, the teacher also allocates class time between group and individual instruction, but computer-aided instruction effectively increases the productivity of individual instruction time. Namely, while the teacher spends time working with student *j*, student *i* can be working on the computer and receiving additional instruction. In contrast to individual instruction time, student *i* can receive an additional minute of CAI time (C_i) without reducing the total amount of instruction time available to student *j*. Total instruction time for student *i* equals

$$T_G + T_i + C_i \leq \overline{T} \tag{3}$$

Let student achievement, S_i, be a function of instruction time and individual characteristics, Z_i, so that

$$S_i = f(T_G, T_i, C_i, Z_i), \tag{4}$$

and $f_1 \geq 0$, $f_2 \geq 0$, and $f_3 \geq 0$. Since $f_3 \geq 0$, student *i*'s achievement in the CAI classroom will be greater than or equal to student *i*'s achievement in the traditional classroom for any given allocation of T_i and T_G, i.e.,

$$f(T_i, T_G, C_i, Z_i) \geq f(T_i, T_G, 0, Z_i). \tag{5}$$

Note that the relative advantage of computerized instruction will depend on the suitability of the curriculum for the students in question which will affect the magnitude of f_3. Suppose further that the teacher maximizes her utility by allocating each student the same amount of individual instruction time. For a class of N students,

$$T_i \leq \frac{\left(\overline{T} - T_G\right)}{N}. \tag{6}$$

Thus, for a given time allocation to group instruction, T_G, T_i decreases as class size increases. In the CAI class this means that $C_i \leq \frac{(N-1)}{N}\left(\overline{T} - T_G\right)$ so the potential gain in total instruction time for student i of moving from a traditional class to a CAI class is increasing in class size N.

Similarly, one might assume instead that individual instruction time (or at least some of it) is non-productive and related to the teacher needing to deal with individual student behavioral problems. Assuming that student j's disruptive behavior reduces group instruction time and/or individual instruction time but does not also disrupt student i's ability to work on the computer, the gain in total instruction time for student i of moving from a traditional class with a disruptive student to a CAI class with a disruptive student is greater than the gain from changing classroom types with a class with no disruptive students.

III. EVALUATING COMPUTER-AIDED INSTRUCTION (CAI)

A. The Empirical Model

The primary research question we examine is whether mathematics instruction is more effective when delivered via computer programs or using traditional ("chalk and talk") methods. In designing the study, we were concerned about two sources of bias that might arise using observational data in which we simply compared the outcomes of students taught using CAI to those taught using more traditional methods. The first is that principals and/or teachers may choose to put students they believed would particularly benefit from computerized instruction into the labs. This bias would overstate the effect of CAI relative to traditional instruction.

A second source of bias is that more (or less) motivated teachers may be more willing to try computerized instruction than their less (or more) motivated peers who would prefer to continue teaching using traditional methods. Thus, a key concern with the existing literature on the effectiveness of computer-aided instruction is that the students taught by teachers willing to teach using the computerized instruction would have outperformed their classmates who were taught by other teachers, regardless of whether or not the students had been in the computer lab. That is, the previous researchers may have confounded a teacher effect with the effectiveness of the computer program.

To control for both types of selection bias, we implemented a within-school random assignment design at the classroom level. We randomly assigned classrooms of students (in which the classroom is the group of students taught by a particular teacher during a particular class period in a particular school) to be taught in the computer lab or using "chalk and talk."[5]

[5] Note that randomly assigning *students* to be taught in the computer lab or not answers a slightly different question: whether being taught in the computer lab – regardless of how classes are typically formed within schools – would generate improvement relative to traditional instruction. Our approach of randomly assigning classes comes much closer to the policy question faced by school principals and superintendents, which is whether instruction for a particular class should

Because classes (with the assigned teacher) will have been randomly assigned, the observed – and unobserved – characteristics of the students and teachers assigned to the computer lab will be identical to those that were not, on average.

Our first empirical model that takes advantage of the randomization generates estimates of the "intent to treat" effect of using computerized instruction. In these models, the test scores of students in classes randomly assigned to the computer lab are compared to the test scores of students in classes randomly assigned to the control group, whether or not the students remained in their original class assignments. To estimate the intent-to-treat effect, we estimate ordinary least squares (OLS) regressions of the following model:

$$Y_{ikj} = \alpha + X_i\beta + \gamma R_{ikj} + \rho_j + \epsilon_{ikj} \tag{7}$$

where Y_{ikj} represents student *i* with teacher *k* in period *j*'s score on one of the follow-up tests, R_{ikj} indicates whether the student was assigned to a class that was randomly assigned to a computer lab, X_i represents a vector of student characteristics (including, in most specifications, the student's baseline test scores), ρ_j is the randomization pool[6], ϵ_{ikj} is a random error term, and α, β, and γ represent coefficients to be estimated. The coefficient γ represents the "intent to treat"

---

occur in the computer lab or in a traditional classroom. We also note that it would be a logistical nightmare to randomly assign students and teachers to classes at the middle or high school level irrespective of their other classroom scheduling needs. That said, the districts in which we conducted this study all use computer software to assign students to classes and they claim this assignment is basically random, as discussed in footnote 11 below.

[6] As described below, in most cases the randomization pool is the class period of the class (within a particular school).

effect and estimates the effect of assigning students to be taught using CAI on the outcome in question.

As noted, above, because we randomly select classrooms, our research strategy should generate estimates of the intent-to-treat effect that are not affected by potential self-selection of teachers into the lab. However, this is only strictly true in large samples and so one might also be concerned that – by chance – the more (or less) motivated teachers ended up in the computer lab. If more motivated teachers ended up being selected to teach in the computer lab, then OLS estimates of the effect of CAI on student outcomes will be biased upwards. One could control for this bias by comparing the achievement of students with teachers who teach both in and out of the lab. That is, one can control for a teacher fixed effect. Indeed, in their meta analysis of the research, Kulik and Kulik (1991) concluded that studies in which the same teacher taught both the computer-aided class and the comparison class, the differences in achievement were much lower than when the two types of classes had different teachers which is consistent with teacher selection bias.

At the same time, this result – that the effect of CAI is lower in the presence of teacher fixed effects – would also obtain if there are spillovers in teaching techniques such that teachers import lessons learned from the lab to their traditional classes. In this case, the spillover will attenuate the estimated impact of computerized instruction. In our study some of the participating teachers taught both in a computer lab and using traditional methods while others taught exclusively in the lab or exclusively out of the lab.[7] This variation allows us to control for

[7] An issue that can arise in studies of this kind is that the teachers and associated staff are unfamiliar with the intervention and therefore not properly trained to use it effectively. All three districts had been using this CAI program on a small scale before our study began (Districts 2 and

the quality of the teacher (by including a teacher fixed effect) and to compare results with and without the teacher fixed effects.[8]

A potential problem with the intent-to-treat estimation is that school staff may "contaminate" the experiment by assigning students from the control group (or from outside of the study) to a CAI lab class. Or, they may assign students originally in a computerized class to a traditionally-taught class. While throughout the study we emphasized the importance of maintaining the original student assignments and the principals and teachers indicated that they understood this importance, some contamination did occur. While the intent-to-treat effect represents the gains that a policymaker can realistically expect to observe with the program (since one cannot fully control whether students initially assigned to a class in the lab actually remain in that class), it does not necessarily represent the effect of the program for those who actually complete it.

Therefore, we also implement instrumental variables (IV) models in which we used whether the student was in a class randomly assigned to a computer lab as an instrumental variable for actual participation. The random assignment is correlated with actual participation in a computer lab but uncorrelated with the error term in the outcome equation (since it was

---

3 for at least one year before our study, and District 1 since 1995), and therefore some of the teachers had already been trained and were familiar with the program. Further, all CAI teachers received training and support from both the company and district support staff throughout the study.

[8] Unfortunately, if we find that the estimated impact of CAI is smaller when we control for fixed effects than when we do not, we will not be able to distinguish whether this is due to more motivated teachers having been selected to be in the lab or to the existence of spillovers from the CAI instruction to traditional instruction. Obviously, if we find that the impact is larger in the presence of teacher fixed effects, we might conclude that, at a minimum, the less motivated teachers were assigned to the lab, by chance, and that this effect was not outweighed by any potential spillovers.

determined randomly). In this case, the second-stage (outcome) equation is represented by models such as,

$$Y_{ikj} = \alpha' + X_i\beta' + \delta CAI_{ikj} + \rho'_j + \epsilon'_{ikj} \tag{8}$$

where CAI_{ikj} indicates whether the student completed at least one lesson in a computer lab, δ indicates the effect of being taught through computerized instruction on student outcomes, and the other variables and coefficients are as before. Through the use of instrumental variables one can generate a consistent estimate of the effect of computerized instruction on student outcomes.

Note that random assignment occurred at the classroom level even though we have data available for each student. Therefore, we adjust our standard errors to account for the fact that the randomization occurred at the classroom level.[9]

B. Computer-Aided Instruction

We study the effectiveness of computer-aided instruction by focusing on a group of computer programs known as *I Can Learn*© (or "Interactive Computer Aided Natural Learning") distributed by JRL Enterprises. The system is composed of both a software and hardware computer package that is designed to deliver instruction through technology on a one-on-one basis to every student; the curricula is designed to meet the National Council of Teachers of Mathematics (NCTM) standards. In addition to the interactive teaching system, the software

[9] In addition, we have estimated our models using data aggregated to the classroom level, and using classroom random effects, with similar results.

package also includes a classroom management tool for educators and the company provides on-site support for administrators and teachers.

The CAI program allows students to study math concepts while advancing at their own pace, enabling them to spend the necessary time on each subject lesson. Each lesson has five independent parts – a pretest, a review (of prerequisites needed for the lesson), the lesson, a cumulative review, and comprehensive tests. Students that do not pass the pretest or review are made to repeat the lesson until they receive a certain degree of mastery. Each student's performance is recorded in a grade book and teachers can monitor students' progress through a series of reports. The teacher's role in this environment is to provide targeted help to students when they need additional assistance. In addition, the computer program covers many administrative aspects such as lesson planning, grading and homework assignment so that teachers may spend more time on individual instruction with struggling students. Previous quasi-experimental studies of the effectiveness of this group of computer programs have yielded mixed results (see, e.g. Brooks 2000, Kerstyn 2001, Kirby 1995, and Kirby 2004).

C. The Research Design

1. *The Sites*

We conducted the study in three large urban school districts: one in the northeast, one in the midwest and one located in the south. Each of these districts had slightly different demographics but suffer similar problems in the areas of underachievement and teacher recruitment. As shown in Table 1, these districts have a high proportion of minority students who are considerably poorer than the national average District 1 has a student enrollment of nearly 68,000 students; 94 percent of whom are African American and 1% percent of whom are Hispanic. District 2 serves just over 22,000 students; 40% of whom are African American and 54% of whom are Hispanic. District 3 serves approximately 97,000 students, 59% of whom are African American and 18% of whom are Hispanic.

2. *Implementation*

To implement our randomized design, near the beginning of the academic year the participating schools provided us with their schedule of pre-algebra and algebra classes.[10] We then randomly selected the treatment classes (taught using CAI) and the control classes (taught traditionally). Officials in the schools were not informed of the outcome of our randomization until they had finished assigning students to classes to protect against students being assigned to classes on the basis of whether it would be taught using traditional methods or in the computer lab.[11] Once students were assigned to classes, we informed the schools which classes should use CAI and which should be taught using a traditional method.

We conducted the study during the 2004-2005 school year in 8 high schools and 2 middle schools in District 1; and during the 2003-2004 school year in 4 high schools in District 2 and in

[10] The schools were given the option of eliminating particular teachers and/or classes from the study before the randomization. The extent to which the schools exercised this option varied.

[11] That said, the schools claimed that the process by which they assigned students was basically random. We have assessed this claim by comparing the standard deviation of baseline test scores within the observed classes with the mean standard deviation that one would obtain if students were assigned to classes randomly (within a particular level). Consistent with the schools' claims, we found that the observed variation in baseline "ability" within classes was similar to that which would obtain if students were randomly assigned. Similarly, the spread of baseline test scores was much larger than what one would have expected if students were strictly "tracked."

3 high schools in District 3. As shown in Table 2, the demographic characteristics of students in the schools in our study in District 1 had a slightly higher percentage of African American students (97%) compared to the schools in the district; the study schools in District 2 were roughly similar to those in all schools in the district; and the schools in District 3 had a larger percentage of African American students (93%) and a smaller percentage of Hispanic students (1.2%) compared to the district average. In most cases, the students in the classes within the schools that participated in the study were representative of the students in the schools (with the exception that in District 1 the average percentage of students that were African American in the study was smaller than that in the schools (88% vs. 97%)).

As shown in Appendix Table 1, our study originally included a total of 17 schools, 147 classes, and 61 teachers. These 147 classes were grouped into 60 "randomization pools" which represented the groups of classes from which we randomly selected candidates for the treatment and control groups. These pools mostly represented a class period, although in a few cases, there were not enough classes from which to randomly pick one to go into the lab and so we combined classes from two periods.[12] Because of mobility, our analysis sample – which is limited to students with follow-up test scores using our main outcome (that on a specially designed algebra test, see below) – is comprised of 17 schools, 141 classes, 59 teachers, and 60 randomization pools.[13]

---

[12] Typically there was only one or two computer labs in each school (one school had three labs) such that there were more math classes than labs available in any one period.

[13] When we further limit the sample to students with baseline test scores on our main outcome we have 17 schools, 137 classes, 57 teachers, and 60 randomization pools, as shown in Appendix Table 1.

D. Data

1. *Academic Outcomes*

We primarily assess the impact of CAI on student achievement using test instruments. First, we sought an exam that was closely aligned with the material in the mathematics courses.[14] Thus, we contracted with the Northwest Evaluation Association (NWEA), a non-profit organization that has partnered with more that 2,300 school districts (serving more that 2 million students) to provide assessments, reports, classroom resources and professional development. NWEA designed a customized paper and pencil exam that targeted specific pre-algebra and algebra skills outlined in the district's course objectives and the CAI curriculum. (In theory, the CAI curriculum was adapted to meet each district course objectives.) NWEA created a 30-item multiple choice exam for both pre-algebra and algebra. The same exams were created for Districts 2 and 3. Slightly different exams were created for District 1 to match the district's standards. However, the exams in District 1 were designed to match the exams used in the other two districts to allow for pooled analysis.

We observe post-test scores for 1,872 students across all three districts (1,165 in District 1, 477 in District 2, and 230 in District 3). However, in some analyses we also control for the student's pre-test. Thus, in the sample that includes both pre- and post-NWEA tests we have 1,585 students (973 in District 1, 412 in District 2, and 200 in District 3). Further, we convert

[14] Note that we did not administer the Terra Nova algebra test, a common nationally-normed mathematics test, because many of the district officials were concerned it does not contain sufficient items related to pre-algebra and lower-level algebra.

the baseline and follow-up test scores to standard deviation units using the standard deviation of the baseline test score.[15]

We also assess the impact of CAI using the statewide tests administered by each state. In District 1, we only have post-treatment state test data for the students in the 8th grade; we use the district-administered Iowa Test of Basic Skills (ITBS) from the 7th grade as the pretest. At the time of our study, students in Districts 2 and 3 were tested in mathematics on state-wide tests in 4th, 8th and 10th grades. Since in these districts the students in the study were primarily in 9th grade, we use the 8th grade statewide test as the pre-test and the 10th grade test as the post-test. The mean of the (standardized) baseline statewide test in District 1 is 9.2; that in District 2 is 6.7; and that in District 3 is 16.7. Again, the test scores were standardized to have a baseline standard deviation of one within each district.[16]

---

[15] We standardize using the standard deviation of the baseline test score for all students across the three districts which is 9.20. We have also used "national" standard deviations which range from 16.7 for 8th grade students to 17.4 for grades 10 and higher. Not surprisingly, this cuts the estimated effect sizes by roughly one-half. We chose to present the effects using the standard deviation within the study for two reasons. First, we have also estimated the effects using "growth norm" gains – the effect of CAI on the expected one-year growth in test scores (this norming takes into account that initially-low scoring students typically make larger yearly gains than initially higher-scoring students). Translated, these estimates are more similar to the effect sizes using the district standard deviation than the national standard deviation, reflecting that our sample of students are by-and-large initially low achieving. As such, the study standard deviation better reflects the population in question. In addition, we only have district (or study) standard deviations for some of the outcomes such that the results are more consistently presented across outcomes when we use the district or study standard deviation. The results using both the growth-norms and national standard deviation are available on request.

[16] Before we standardize the test scores, the standard deviation of the baseline statewide test in District 1 is 23.3; that in District 2 is 31.7; and that in District 3 is 39.1. For District 1 we standardize the 8th grade follow-up test score using the standard deviation of the 8th grade test for the study 9th graders because the pre- and post tests are not the same test. The standard deviation of the 9th graders' 8th grade test is 44.7.

In addition, pre-algebra students in District 1 took mini-math exams – benchmark pre-algebra exams – throughout the semester. These tests were intended for use by the teacher and district to track students' progress. The initial benchmark test has a mean of 18.7 and a standard deviation of 5.7. We standardize the initial benchmark test to have a standard deviation of one and also standardized the 2nd and 3rd quarter benchmark tests using the initial test score standard deviation.

Because we do not have a way of standardizing the state tests across the districts, we analyze these data separately by district. The sample size of students in District 1 with both pre- and post-tests is 237; that in District 2 is 341, and that in District 3 is 199. Further, the sample size for the benchmark tests in District 1 is 230. We emphasize that while the state tests have the advantage of being high-stakes and therefore of great importance to the districts, as little as 10% of the state exams in mathematics contain test items related to pre-algebra and/or algebra. As such, they may have low power to detect effects of a pre-algebra/algebra intervention.[17]

Despite the fact that only a fraction of the state tests focuses on pre-algebra and algebra, the three test assessments are reasonably highly correlated. For example, the correlation between the baseline NWEA test and the state math tests range from 0.30 (in District 1) to 0.73 (in District 2). Further in District 1 the correlation between the baseline algebra test and the baseline benchmark test is 0.57 and that between the state math test and the baseline benchmark

[17] In one of the districts we were able to identify individual test items that were related to pre-algebra and algebra. Not surprisingly, our estimates were quite noisy given that there were very few test items on which to measure the students' performance.

pre-algebra test is 0.62. Thus, while two of our three assessments are not based on nationally normed exams, they nonetheless appear to be correlated with the high-stakes state tests.[18]

2. *Other Data*

The statistical office in each district also provided us with administrative data on students. The data included student identifiers, limited characteristics (such as the student's sex, race/ethnicity, and eligibility for a free or reduced-price lunch). In two of the three districts we also obtained data on the number of days the students attended school the previous year and the year in which we conducted the study; and we have limited information on in- and out-of-school suspensions. In addition, we gauge each student's engagement with the program and the time-on-task through tracking data that comes with the computerized program. Importantly, these data allow us to determine which students ever actually trained in the computer lab versus in a traditional classroom for the analysis estimating the effect of the treatment on the treated.

IV. Results

A. Descriptive Statistics

The first order of business is to determine if assignment to the computer lab appears random. Table 3 shows the mean of student characteristics by whether or not the student's class was assigned to the CAI lab or was assigned to receive traditional instruction. The top panel

[18] For comparison, Figlio and Rouse (2006) report that in a subset of Florida districts the correlation between student performance on a nationally-normed test (the NRT) and the FCAT curriculum-based assessments (known as the Sunshine State Standards (FCAT-SSS) examinations) is approximately 0.8.

uses the full sample of students who were randomly assigned at the beginning of the academic year. We see that the proportion of female, African American, and Hispanic students are quite similar using the full sample. Further, the baseline test scores are identical.

However, there is significant mobility among students in the districts such that we were unable to post-test all of the students. A major concern is that the attrition between the beginning and end of the study was uneven between the treatment group and the control group thereby introducing statistical bias into the analysis. We therefore compare the observable characteristics of the students in the treatment and control groups using the sample of students for whom we also have both the baseline and follow-up data on the NWEA test in the bottom panel. Again, there is no difference in the baseline pre-algebra/algebra test score, however there are small differences in the percentage of students that are African American and Hispanic that are statistically significant at the 6% level.[19] As a result, in most specifications we control for the sex, race and ethnicity of the student.

B. Overall Intent-to-Treat and Treatment-on-the-Treated Estimates

Table 4a presents the OLS estimates of the intent-to-treat effects of CAI represented by equation (1) as well as an instrumental variables (IV) estimate of the effect of treatment-on-the-treated using the NWEA test as an outcome. Column (1) presents the straightforward mean difference in the post-test between students learning algebra using CAI and those learning in a traditional classroom adjusted only for dummy variables representing the randomization pool.

[19] We note, however, that these differences in race and ethnicity arise in only one district (District 2).

The standard errors reported allow for within-classroom correlation. We estimate that, on average, students in CAI scored 0.17 of a standard deviation higher on the post-test than did those in a traditional classroom, and this difference is statistically significant at the 5% level. When we add controls for the sex and race/ethnicity of the student, in column (2), the random assignment effect does not change.

In column (3) we present the same specification as that in column (1) but restrict the sample to those students who also had a pre-test. The basic effect of CAI is slightly higher – 21% of a standard deviation – among the subset of students with baseline test scores, although the estimate is within a standard error of that in column (1).[20] Note that the coefficient estimate falls slightly when we include the baseline test score (columns (4) and (5)), although this difference is not statistically different from that in column (3). Thus, we estimate that the effect of being placed in a CAI classroom relative to a traditional classroom is an educationally and statistically significant 0.17 of a standard deviation. To interpret this effect differently, when we use the growth-normed test scores, we find that students assigned to a CAI classroom achieve 26% of a grade-level more than their peers at the end of the semester.

However, if some contamination occurred in the study, these OLS estimates will understate the potential educational gains by students who are actually taught in the lab. To the extent that students assigned to classrooms to be taught using traditional methods spent time in the lab and students assigned to the lab did not receive their algebra instruction there, the intent-to-treat estimates may be too small. Table 5 shows the number of lessons students were

[20] Further, when we regress whether the student is missing the baseline test score on a variety of student characteristics, none of the characteristics significantly differ between those with and without baseline test scores.

expected to complete given the course taken; the percentage of students completing no lessons, more than 10 lessons and more than 20 lessons in the CAI; the number of lessons the student actually completed; and the number of lessons completed as a fraction of the CAI course expectations by whether the student was assigned to the treatment group or the control group.

Note, first, that there is no difference in the number of CAI lessons that students would have been expected to complete based on the level of their math class and the school's schedule. However, there is evidence of some, although not extensive, contamination. For example, 84% of students assigned to the lab completed at least 10 lessons in the lab; 15% of those assigned to classes to be taught using traditional instruction completed at least 10 lessons in the lab as well. Similarly, while treatment students completed an average of 33 lessons using CAI, the control group students completed an average of 5.6 lessons. And, while the treatment students appear to have completed about 64% of the lessons they would have been expected to complete using CAI, the control students completed 10%.

We address this contamination by using IV to estimate equation (2), the results of which are in column (6). In this specification we identify students who were "treated" as those who completed at least one lesson in the computer lab and instrument for this indicator with the random assignment of the student's class.[21] This strategy provides a consistent estimate of the effect of "treatment-on-the-treated." We estimate that students who actually receive instruction using CAI score 0.25 of a standard deviation higher than those who received instruction in a traditional classroom, and the difference is statistically significant.

---

[21] We have used alternative definitions of students receiving treatment, such as whether the student completed at least 5 lessons in the lab and whether the student completed at least 10 lessons in the lab. The results were robust to these alternative definitions.

As noted above, although we have nearly 60 teachers who participated in the analysis, we also sought to understand whether these impacts result because we, by chance, selected more motivated teachers to teach in the lab. Thus, we exploit the fact that just over one-half of the teachers taught both in and out of the computer lab and include teacher fixed effects in the analysis. These results are presented in Table 4b which is otherwise identical in layout to Table 4a. The within-teacher coefficient estimates are uniformly greater than those without teacher fixed effects. Thus, we estimate that, controlling for (time invariant) teacher quality, the effect of being assigned to a computer lab increases student math achievement. The intent-to-treat effect is nearly 30% of a standard deviation; when we adjust for non-compliance using IV the effect of CAI increases to 40% of a standard deviation. These effects are educationally large and statistically significant and (translated) suggest that students who actually completed lessons in the lab gained roughly 50 percent of a year more than those taught in a traditional classroom.[22]

We next consider whether we detect similar effects of CAI on student math achievement using other math test instruments. Because these instruments were not standardized across the districts, we present the results separately by district. Table 6a shows the intent-to-treat effect of CAI in which we use four outcomes in District 1. The first (column (1)) is the pre-algebra and algebra test developed by NWEA that was also used as the outcome in Tables 4a and 4b; the second and third are the second and third quarter benchmark tests conducted by the district

[22] Part of the reason for the larger estimated coefficients in Table 4b derive from the fact that the intent-to-treat effect of CAI is larger when we limit the sample to the subset of teachers who taught both in- and out- of the lab (i.e., those observations from which the fixed effects analysis is identified). When we conduct the analysis on this subsample of teachers and do not include teacher fixed effects the intent-to-treat effect (similar to that in column (4) in Table 4a) is 0.27 and the IV estimate (similar to that in column (6) in Table 4a) is 0.44.

(columns (2) and (3)); and the final column (column (4)) is the statewide math test. We present the results in two panels: the top panel uses the maximum available sample for each outcome and the lower panel constrains the sample to be constant across them.

In District 1, when we allow for the maximum possible sample, the intent-to-treat effect using the NWEA pre-algebra/algebra test is approximately 0.23 of a standard deviation. We see a larger gain of 0.4 of a standard deviation using the 2nd quarter benchmark test and a gain of 0.6 of a standard deviation using the 3rd quarter benchmark test. Importantly, we also detect an effect of 0.26 on the state mathematics test. All of these gains are educationally large and statistically significant at the 5% level. Further, the coefficient estimates in the bottom panel suggest that the gains are not simply driven by changes in the sample size across the specification as they are even larger.

Analogous results for Districts 2 and 3 are presented in Table 6b (note that benchmark tests were not administered in these districts). Columns (1) and (3) show the effect of CAI using the NWEA test; those in columns (2) and (4) report the effect using the statewide test for each of the districts. In District 2 we detect an effect of 0.2 of a standard deviation using the algebra test with a p-value of 0.13; the effect is much smaller on the state test – less than 10% of a standard deviation – and not statistically different from zero. That said, these results are not unexpected given that most of the state math test is not geared towards pre-algebra and algebra. Note that the results do not appear to depend on whether or not the sample is restricted to be the same in both specifications. In contrast, we estimate a negative intent-to-treat effect of CAI on student achievement in District 3 using both the NWEA test and the state math test, although neither coefficient estimate is statistically different from zero (in fact the standard errors are much larger than the coefficient estimates).[23]

While the magnitude of the intent-to-treat effect is largest in District 1, the effect (based on the algebra test) is not statistically distinguishable from that in District 2.[24] Further, we note that the negative effect in District 3 is driven by the results from only one randomization pool. If we exclude this pool from the analysis the point estimate in column (3) of the top panel of Table 6b rises to 24 percent of a standard deviation and that in column (4) rises to 15 percent of a standard deviation. These estimates are not statistically different from those estimated in Districts 1 and 2.[25] In addition, in the districts in which CAI appears most effective, the test improves student achievement on more than simply one math test.

C. Empirical Evidence on Why is CAI More Effective

The discussion in Section II suggested that CAI may more effective for some students than others and in classes in which individualized instruction may be particularly advantageous. In the following tables, we look for patterns of impacts that are consistent with this

---

[23] We have also estimated IV models by district for all of the outcomes. In general the coefficient estimates are larger but not qualitatively different from the OLS estimates. These results are available on request.

[24] This inference is based on a combined regression in which we interact the intent-to-treat effect with dummy variables indicating the school district.

[25] The subsequent results are qualitatively similar with or without this one randomization pool in District 3. A complete set of results without the randomization pool are available on request.

interpretation.[26,27] In Table 7 we estimate whether the effect of CAI is different for pre-algebra versus algebra or students of different ability as measured by baseline (NWEA) test scores.[28] Each column of the table represents estimates of the effect of CAI for a different subset of the analysis sample. We present estimates for the three districts combined (column (1)), districts 1 and 3 combined (column (2)), and district 1, 2, and 3 separately in columns (3), (4), and (5), respectively. The top panel estimates differential effects by pre-algebra and algebra and the bottom panel estimates the CAI effect by student ability as measured by the baseline test score quartile.[29]

We have study students in algebra and pre-algebra classes in all three districts with roughly 23 percent in pre-algebra classes.[30] Pooling all three districts we estimate that the effect

---

[26] We have conducted all of the subsequent analysis using the statewide tests rather than the NWEA pre-algebra and algebra test designed for this study. The biggest problem is that the sample sizes are much smaller generating results that are quite imprecise. However, many of them are qualitatively similar to those presented in the paper. These results are available from the authors on request.

[27] We have also tested whether the effectiveness of CAI differs by sex or race/ethnicity and find no systematic differences. The results are available from the authors on request.

[28] Each column in each panel represents a separate regression.

[29] Test score quartiles for all specifications are defined within district and algebra level. All specifications additionally control for student demographic characteristics as described above and indicators for the randomization pool. The top panel also includes the baseline test score while the bottom panel includes, instead, indicators for the baseline test score quartile. We also include main effects for the level of math class in the top panel. We emphasize that these results are qualitatively similar when use growth-normed scores suggesting that they are not an artifact of the test score scaling and the possibility that students at different parts of the distribution would naturally have differential gains over the course of the year.

[30] In the analysis sample, 30 percent of District 1 students are in pre-algebra, 12 percent of District 2 students are in pre-algebra, and 9 percent of District 3 students are in pre-algebra.

of CAI for pre-algebra students is significantly larger than the effect for algebra students (the p-value of the difference between the two effects equals 0.001). Pre-algebra students in CAI score 0.48 standard deviations higher than pre-algebra students in traditional classes while algebra students in CAI score less than 1 percent of a standard deviation higher and the effect is not statistically different from zero. Note, however, that the effect of CAI for algebra students is being driven toward zero by the negative effect of CAI for algebra students in districts 2 and 3. That said, even in District 1 we find evidence that CAI has a larger effect among pre-algebra students than algebra students. In District 1 we estimate that CAI pre-algebra students score 0.44 standard deviations higher than traditionally taught pre-algebra students while CAI algebra students score only 0.13 standard deviations higher than traditionally taught algebra students. For each district the p-value for the test that the pre-algebra effect of CAI equals the algebra effect of CAI is less than 0.07.[31] Thus, this CAI treatment appears more effective for pre-algebra students than for algebra students.

In the bottom panel we allow the effect of CAI to differ by prior student math achievement.[32] A promised benefit of CAI is that the instruction is completely individualized in the sense that students can move at their own pace in covering the material. In contrast, students in a traditional classroom cover all lessons at the same pace. This could mean that CAI is

[31] Statistically, we can reject that the effectiveness of CAI for algebra students is the same in district 2 or 3 as in district 1. The effectiveness of CAI for pre-algebra students in district 2 is very similar to and not statistically different from that in district 1, and although the estimated CAI effect for pre-algebra students in district 3 is larger than in district 1, we also cannot reject that it is same as in district 1.

[32] In the bottom part of this table and in the subsequent tables we combine pre-algebra and algebra students to increase our statistical power. The results are qualitatively similar if we limit the sample to pre-algebra students. Such results are available on request.

differentially effective for students of different math ability. For example, suppose traditional classroom teachers always teach pre-algebra and algebra at the pace that is appropriate for the highest ability students in the class. In this case, we might expect to see that high ability students do equally well in CAI and traditional classrooms while those with lower math ability do better in CAI because they can take more time to cover each lesson and therefore learn the material better even if they do not cover as many lessons. Alternatively, if traditional classroom teachers always teach pre-algebra and algebra at the pace that is appropriate for the lowest ability students then high ability students may do better in CAI because they can cover more material than covered in a traditional classroom. While a possibility, when we pool either all three districts (column (1)) or Districts 1 and 2 (column (2)) we estimate that CAI is roughly equally effective for students with the lowest and highest prior math achievement students (p-value>0.60). Thus, we find no evidence that CAI is more or less effective for students with stronger or weaker backgrounds in math as measured by the baseline algebra test.

Tables 8a and 8b test for different CAI effects by attendance characteristics of individual students and for the class based on attendance data from the prior academic year. As noted earlier, we only have data on student attendance for Districts 2 and 3. While the pooled data suggest that, indeed, CAI is more effective for students with worse attendance rates we cannot reject that there are no differences at standard levels of significance. We find some statistically significant differences by attendance quartile using District 3 alone, but the pattern of results are not fully consistent with hypothesis that the individualized instruction of CAI mitigates the negative effects of poor attendance rates.

Table 8b presents estimates allowing the effect of CAI to differ with the average attendance rate of the students in the classroom.[33] For Districts 2 and 3 either pooled or individually we find a larger CAI effect for classrooms with lower average attendance rates. For students in a classroom with average attendance rates, the CAI effect is less than 6 percent of a standard deviation and not statistically different from zero. In contrast, the CAI effect for students in a classroom with attendance rates one standard deviation below the mean is 0.35 of a standard deviation (p-value equals 0.08).

Next, we examine whether CAI is more effective for larger classes. Here we measure class size based on the initial class assignment rosters used for random assignment; thus, class size is available for all three districts. The average class sizes in these districts range from 24 to 29 students. Pooling all three districts, we find that the CAI effect is larger for larger classrooms; unfortunately this marginal effect is not statistically significant at standard levels (p-value equals 0.19). However, pooling only Districts 1 and 2 we find that the CAI effect is about twice as large and statistically significant at the 10% level (the p-value is 0.067). Based on this estimate, for a classroom of 25 students the effect of CAI is 0.21 of a standard deviation (p-value < 0.001). For a class of 15 students there is no difference between CAI and traditional instruction (0.01 of a standard deviation with a p-value of 0.89). Class size effects are positive for District 1 (p-value = 0.09) and District 2 (p-value = 0.80), individually. The coefficient estimate is very small and negative with a large standard error in District 3. We cautiously

[33] For each student we calculate the average attendance rate of her classmates using attendance data for the prior year and excluding her own attendance rate from the calculation.

conclude there is some evidence CAI is more effective in larger classes, consistent with the idea that the main benefit of CAI is the individualization of the instruction.

Finally, we examine whether CAI effects are larger in classrooms with greater heterogeneity in terms of baseline math achievement. Specifically, we allow the CAI effect to depend on the baseline test score standard deviation for the class. The top panel of Table 10 presents overall results. While the estimate of the coefficient on the interaction term for District 1 is negative, those for Districts 2 and 3 individually, are positive, consistent with the idea that the benefit of CAI is through individualized instruction. However, regardless of sample, none of the coefficients on the interaction between CAI and baseline standard deviation are statistically significant.

One potential explanation for the results only being weakly supportive of the importance of individualized instruction is that heterogeneity, in-and-of itself, may not hinder effective teaching. Rather, in certain circumstances – such as in small classrooms – heterogeneity in student ability may be quite manageable in a traditional classroom. In this case, the relative advantage of CAI (and hence more individualized instruction) may only become apparent in large and heterogenous classes. To test this hypothesis, in the second panel of Table 10 we add a third level interaction – that between CAI, the baseline standard deviation in student test scores, and an indicator for whether the class is "large" (defined as more than 24 students).[34] We now find there is a large, statistically significant, relative advantage to being assigned to CAI for

[34] The results are robust to small changes in the definition of a large class. For example, the result is similar if we define large classes as those with more than 20 students (the 30th percentile based on classrooms), but they are not similar at the 60th percentile (more than 26 students). We also obtain qualitatively similar results when we define class size as a continuous variable.

large, heterogeneous classes which is consistent with the hypothesis that CAI benefits primarily accrue through increased individualization of instruction.

V. COST-BENEFIT SIMULATION

Of course, gains from computerized instruction do not come for free as the computer labs required for CAI are costly and are dedicated to CAI. In our example, a 30-seat lab costs \$100,000 with an additional \$150,000 for pre-algebra, algebra, and classroom management software and roughly \$17,000 per year for training, support, and maintenance of the lab.[35] According to the company's website a lab lasts 7-10 years so a CAI lab may cost nearly \$53,000 per year.[36]

Given that providing instruction through CAI may serve as a substitute for reduced class sizes, one way to evaluate its cost effectiveness is to compare its cost to the compensation cost of hiring additional teachers to reduce class size. Using pre-algebra/algebra test scores measured in national standard deviation units we find that a student in an average-sized class (24 pupils) using CAI in our largest district (District 1) scores 11 percent of a standard deviation higher than a student in a similarly-sized traditional classroom. Because the gains from CAI are larger for larger classes, the benefit of CAI equals zero when the average class size is reduced to 13

[35] Information on the cost of a CAI lab comes from one of the districts in our study.

[36] The company estimates the annual cost per pupil at just over \$100. However, we can only get close to this per-pupil estimate if we assume that the lab would serve 400 students per year over a 7 year period and that the district would not pay for training, support, and maintenance cost after the initial three years. We generate our own estimates because we believe this cost per pupil to be unrealistically low.

students. Thus we compare the per-pupil cost of CAI to the cost of reducing class sizes to 13 students.

Begin with an estimate of the cost of reducing class size using all of the schools in District 1 that are in our analysis sample.[37] The average class size for all District 1 classes represented in the study is 23.5. Although District 1 has eight periods per day, by contract teachers do not teach every period. The typical teacher in our sample teaches 6 periods. As a result, the District would have to hire about 24 more pre-algebra and algebra teachers to reduce the average class size to 13. Using an estimate of the starting salary for teachers in District 1, adjusted to reflect "total compensation," we estimate that the cost of class size reduction would be \$241 per pupil per year.[38] (See the Simulation Appendix and Appendix Table 3 for details.)

The key determinants of whether CAI is more cost effective than class size reduction are the average number of students per class in the lab and the number of periods in the day a lab can be used. If the district implements CAI and keeps the average class size in the lab at 23.5 students, the annual per pupil cost is about \$279. Per pupil costs of CAI are lowest when the lab can be used every period of the day and each class has 30 pupils in it. If 30 students were assigned to classes in the lab, the per pupil cost decreases to about \$218 which is slightly lower

---

[37] We only report estimates using the analysis sample in District 1 because we have a good understanding of the typical number of periods in each school; we must make more assumptions when we using our entire analysis sample. That said, the estimated annual cost per pupil of CAI would be about \$274 using the entire analysis sample and the estimated cost of reducing class size to 13 students would be about \$246.

[38] The cost of reducing class size in this simulation is much lower than the estimates of the cost of class size reduction for elementary schools as in Tennessee STAR (e.g., nearly \$5000 per pupil in Schanzenbach 2006). This is primarily because when class sizes are reduced at the elementary school level, it is for all subjects, not just algebra and pre-algebra.

than the estimated cost of class size reduction. More generally, the per pupil cost of CAI is estimated to be less than or equal to the cost of class size reduction as long as the district increases the average class size in the lab to between 27 and 30 pupils.

For individual schools in District 1 with larger average class sizes, our estimates of the cost of implementing CAI are less than our compensation cost estimates of reducing class size, even without increasing the average class size in the lab. For example, School B has an average class size of 26.8. In this case, cutting the average class size in half costs roughly $278 per pupil compared to $245 per pupil to implement CAI without changing the average class size. The benefits of CAI are the most attractive in School A where the cost of reducing class sizes is over $100 more per student than that of adopting CAI.

In general our calculations suggest that the costs of reducing pre-algebra and algebra classes to 13 students and adopting CAI are quite comparable. However, we suspect that our estimates of the cost of class size reduction are more severely underestimated compared to those for CAI. The reason is that they only reflect increased costs in terms of teacher compensation while, in fact, there would likely be additional costs such as recruiting costs and capital expenditures that have not been taken into account. As a result, CAI may be the more cost-effective way for school districts to raise mathematics achievement. Furthermore, in urban and rural districts that have difficulty hiring highly qualified mathematics teachers, CAI may be much easier to implement than a drastic reduction in class size.

VI. CONCLUSION

Our results suggest that CAI may increase student achievement in pre-algebra and algebra by at least 0.17 of a standard deviation, on average, with somewhat larger effects for students in larger classes. Put differently, students learning pre-algebra and algebra through CAI are 26% of a school year ahead of their classmates in traditional classrooms after one year. In interpreting these results, one must keep in mind that the outcomes were measured relatively soon after the intervention ended such that we do not know how long they would "last." At the same time, it is not clear to us how one might measure such longer run outcomes, particularly since mathematics is not necessarily cumulative at the secondary school level, students in the control group may go on to use CAI, and all of the students may have been involved in other enrichment programs. In addition, this represents only one use of computers for teaching pre-algebra and algebra and not all CAI hardware and software may be equally effective. That said, this study suggests that CAI has the potential to significantly enhance student mathematics achievement in middle and high school, that the gains are comparable to those achieved with drastic class size reduction, and that the costs are likely somewhat lower than the full cost of reducing the average class size for all algebra and pre-algebra classes. At the very least, our results suggest that CAI deserves additional rigorous evaluation and policy attention, particularly since it may be much easier for schools and districts to implement than large scale class size reduction.

SIMULATION APPENDIX

In this appendix we present more detailed information on the cost calculations for CAI and class size reduction using information on all algebra and pre-algebra classes for two schools in District 1. We also present the same calculations for all District 1 algebra and pre-algebra classes in the analysis sample.[39] Thus the top panel of the table presents cost estimates for implementing CAI while the bottom panel presents cost estimates for reducing class size to 13 students. The cost estimates vary because of differences across the schools in the average class size.

The first three columns are identical in each panel and represent the total number of pre-algebra and algebra classes, total number of students, and the average class size, respectively. Column (4) lists the number of periods the lab is in use (top panel) or the teacher is teaching (bottom panel). For CAI we assume that the average class size is equal to the observed average class size or a maximum of 30 students (column (5) in the top panel). For class size reduction, we assume that classes are reduced to 13 students. Column (5) in the bottom panel equals the total number of new classes required to generate an average class size of 13. Column (6) then presents the number of labs the school (district) needs to put all algebra and pre-algebra classes in CAI (top panel) or the number of additional teachers needed to reduce algebra and pre-algebra class size to 13 given the assumption that the new teachers teach for 6 of the 8 periods in the day. Finally, we assume the lab involves a fixed cost of $250,000 for hardware and software and $50,000 for 3 years of support, training, and maintenance and that the lab is good for 7 years. For

---

[39] As noted in the text, we only present results using the analysis sample in District 1 because we have specifics about the structure of the school day. To use the entire analysis sample we must make more assumptions.

the compensation cost of each teacher we use the salary of a new teacher in district 1 with zero years of experience and further assume that salary is 70 percent of the total compensation cost.

For a large school in our sample (School A), the cost of CAI is $218 per pupil compared to $329 per pupil to reduce class size to 13 students. For a smaller school in our sample (School B), the cost per pupil is roughly $245 for CAI compared to $278 for class size reduction. The final row in each panel presents cost estimates using information for all algebra and pre-algebra classes in District 1 that are represented in the analysis sample.[40] In this case, our per pupil cost of CAI is nearly $280 compared to a per pupil cost of reducing class size that is closer to $240.

When we consider the analysis sample for all three districts, we assume that teachers typically teach 6 out of a total of 8 class periods during the day in all three districts and that teacher salaries are the same as in District 1. Thus, since the average class size for all classes in the analysis sample (23.9) is quite similar to the average for District 1 classes (23.5), the estimates of the cost of CAI and the cost of class size reduction are quite similar to the estimates for District 1, $274 per pupil for CAI and $246 per pupil for class size reduction. This is likely an overestimate for CAI and an under estimate for class size reduction. For some of the schools in districts 2 and 3, it appears that teachers may actually teach fewer than 6 classes per day, and some schools may actually have more than 8 possible periods during the day. Also, teacher salaries may be somewhat higher in District 2 than in Districts 1 and 3.

---

[40] Most of schools in District 1 operate on a block schedule; however, classes could be organized either in 4 blocks for 1 semester or 8 periods over 1 year. For simplicity we assume classes are organized into 8 periods over 1 year for all schools.

References

Angrist, Joshua and Victor Lavy. "New Evidence on Classroom Computers and Pupil Learning," *The Economic Journal*, no. 112, October, 2002, pp. 735-765.

Banerjee, Abhijit, Shawn Cole, Esther Duflo, and Leigh Linden. "Remedying Education: Evidence from Two Randomized Experiments in India," *Quarterly Journal of Economics* (forthcoming).

Boyd, Donald, Daniel Goldhaber, Hamilton Lankford, and James Wyckoff. "The Effect of Certification and Preparation on Teacher Quality" in *The Future of Children*, vol. 17 no. 1 (Spring 2007): 45-68.

Braswell, James S., Anthony D. Lutkus, Wendy S. Grigg, et al. *The Nation's Report Card: Mathematics 2000*. (Washington, D.C.: National Center for Education Statistics), August 2001.

Brooks, Cormell. Evaluation of Jefferson Parish Technology Grant I CAN Learn Algebra I, submitted to Elton Lagasse, Superintendent, Jefferson Parish Public Schools. September, 2000.

Brown, Byron W. and Daniel H. Saks. "The Microeconomics of Schooling: How Does the Allocation of Time Affect Learning and What Does It Reveal about Teacher Preferences?" Unpublished manuscript, March 1984.

Cuban, Larry. *Oversold and Underused: Computers in the Classroom*. Cambridge, MA: Harvard University Press. (2001)

Enochs, J.R., H.M. Handley, and J.P. Wollenberg. "Relating Learning Style, Reading Vocabulary, Reading Comprehension, and Aptitude for Learning to Achievement in the Self-Paced and Computer-Assisted Instructional Modes." *Journal of Experimental Education*, vol. 54, no. 3 (Spring 1986): 135-139.

Figlio, David and Cecilia Elena Rouse. "Do Accountability and Voucher Threats Improve Low-performing Schools?." *Journal of Public Economics* 90, nos. 1-2 (January 2006): 239-255.

Goolsbee, Austan and Jonathan Guryan. "The Impact of Internet Subsidies in Public Schools," *The Review of Economics and Statistics,* 88 no. 2 (May 2006): 336-347.

Grogger, Jeffrey. "Does School Quality Explain the Recent Black/White Wage Trend?" *Journal of Labor Economics*, 14 (1996): 231-253.

Heath, Marilyn and Ravitz, Jason. “Teaching and Learning Computing: What Teachers Say.” Presented at ED-MEDIA 2001 World Conference on Educational Multimedia, Hypermedia and Telecommunications, 2001.

Kerstyn, Christine. “Evaluation of the *I Can Learn* Mathematics Classroom: First Year of Implementation (2000-2001 School Year)” Hillsborough County Public Schools mimeo, 2001.

Kirby, Peggy C., “I Can Learn Algebra I” Pilot Project Evaluation Report II, submitted to JRL Enterprises, December 1995.

Kirby, Peggy, C., “Comparison of I CAN Learn® and Traditionally-Taught 8th Grade Student Performance on the Georgia Criterion-Referenced Competency Test. Unpublished manuscript, November 2004.

Kirkpatrick, Heather and Larry Cuban. “Computers Make Kids Smarter – Right?” Technos 7, (2), Summer 1998, pp. 26-31.

Kulik, Chen-Lin C. and James A. Kulik. “Effectiveness of Computer Based Instruction: An Updated Analysis.” Computers in Human Behavior. vol. 7, pp. 75-94 (1991)

Lazear, Edward P. “Educational Production.” *Quarterly Journal of Economics*. 116 (3), pp. 777-803.

Lepper, Mark, R. and Jean-Luc Gutner. Children and Computers: Approacahing the Twenty-First Century. American Psychologist, V44, n2, Feb 1989, p170-78.

Machin, Stephen, Sandra McNally, and Olmo Silva. “New Technology in Schools: Is There a Payoff?” *Economic Journal* (forthcoming).

Means, Barbara. and Olson, Kerry. “Technology’s Role in Education Reform.” Menlo Park, CA: SRI International. (1995)

Murnane, Richard J. and Jennifer L. Steele. “What is the problem? The Challenge of Providing Effective Teachers for All Children,” *The Future of Children*, vol. 17, no. 1 (Spring 2007), forthcoming.

Murnane, Richard J., John B. Willet, and Frank Levy. “The Growing Importance of Cognitive Skills in Wage Determination.” *Review of Economics and Statistics* 77 (1995), pp. 251-266.

Ragosta, M. et al. “Computer-Assisted Instruction and Compensatory Education: The ETS/LAUSD Study Final Report, Project Report 19.” Princeton, NJ: Educational Testing Service, 1982.

Rouse, Cecilia Elena, Alan B. Krueger, with Lisa Markman. "Putting Computerized Instruction to the Test: A Randomized Evaluation of a 'Scientifically-based' Reading Program." *Economics of Education Review* 23, no. 4 (August 2004): 323-338.

Schanzenbach, Diane Whitmore. What Have Researchers Learned from Project STAR? Harris School Working Paper Series 06.06. (August 2006)

Surber, Colleen F. and others. "Self-Pacing Versus Pacing Requirements: Criterion Measures, Student Evaluations, and Retention. Paper presented at the Annual Meeting of the American Psychological Association (Washington, D.C., September 3-7, 1976).

Snyder, Thomas D., Sally A. Dillow, and Charlene M. Hoffman. *Digest of Education Statistics: 2006*. (Washington, DC: National Center for Education Statistics, 2007).

U.S. Department of Education. "The Nation's Report Card. Mathematics 2005" Washington, D.C. 2006.

Wang, Xiaoping, Tingyu Wang, and Renmin Ye. "Usage of Instructional Materials in High Schools: analyses of NELS Data." Presented at Annual Meeting of American Educational Research Association. 2002.

Wenglinsky, Harold. "Does it Compute? The Relationship Between Educational Technology and Student Achievement in Mathematics." Princeton, NJ: Policy Information Center, Research Division, Educational Testing Service. (ERIC Document Reproduction Service No. ED425191) 1998.

Table 1: Districts in Study Compared to National Average

	United States 100 Largest Districts	3 Districts Combined	District 1	District 2	District 3
Average # of students in a district (all grades)	112,807	~63,000	~68,000	~22,000	~97,000
% Female	48.8	49.4	49.7	48.8	49.3
% African American	28.1	69.5	93.6	40.3	59.4
% Hispanic	34.1	16.2	1.1	54.3	18.0
% Native American	0.6	0.5	0.1	0.1	0.9
% Asian	7.1	3.1	1.9	0.8	4.4

Source: Authors' calculations based on the National Center for Education Statistics Common Core of Data, 2003-2004 school year, 100 largest districts by total enrollment. Percentages are based only on schools reporting. (Data on sex are missing for Knox County, Memphis City, Nashville-Davidson County, Philadelphia City, Portland, and Shelby County School Districts. Data on race and ethnicity are missing for Memphis City, Nashville-Davidson County, and Shelby County School Districts.) Demographic characteristics for the 3 districts combined are enrollment-weighted averages of the individual district means.

Table 2: Schools and Students in Study Compared to the Overall District Averages

	District 1			District 2			District 3		
	Relevant Schools	Schools in Study	Students in Study	Relevant Schools	Schools in Study	Students in Study	Relevant Schools	Schools in Study	Students in Study
number of students	29,603	8,148	973	5,270	4,476	412	27,572	3,540	200
students per school	604	815	97	659	1119	103	484	1180	67
% grade 8	19.3	16.8	40.4	2.3	0.0	0.0	1.4	0.0	3.5
% grade 9	18.0	18.3	47.2	38.0	40.0	52.7	35.6	40.0	91.5
% grade 10	15.1	17.8	9.9	22.0	23.2	31.8	23.3	25.1	3.0
% female	50.5	49.0	52.0	48.4	48.2	46.7	49.9	47.6	47.7
% African American	94.2	97.2	87.8	43.6	42.0	47.1	61.1	92.5	94.5
% Hispanic	1.0	0.8	0.8	50.1	51.2	44.7	15.2	1.2	0.5
% white	2.6	0.4	0.1	5.5	5.9	6.6	18.3	4.0	1.5
% Native American	<0.1	<0.1	0.0	0.2	0.1	0.2	1.1	0.4	0.0
% Asian	2.2	1.6	1.8	0.7	0.8	0.5	4.5	1.9	3.0
% missing demographic data			9.6			0.2			0.5

Source: Authors' calculations based on the National Center for Education Statistics. Common Core of Data, 2003-2004 school year. There are 49 "relevant" schools in District 1, 8 in District 2, and 57 in District 3. Relevant schools in District 1 are defined as schools in the CCD with a level of middle school, high school, or other; relevant schools in District 2 and District 3 have a level of high school or other. We drop middle schools in District 1 for which the highest grade offered is less than grade 8. There are 10 schools in the study in District 1, 4 schools in District 2, and 3 schools in District 3. Characteristics on the students in the study come from data made available to the authors by the school districts.

Table 3: Randomization of Treatment and Control Using Full and Analysis Samples

	Random Assignment		
	Traditional Instruction	Computer-Assisted Instruction	p-value of difference
	Full Sample		
Baseline algebra test score	24.7	24.7	0.494
Percent female	47.2	47.1	0.637
Percent African American	80.0	83.2	0.561
Percent Hispanic	15.9	13.5	0.195
Class size	25.8	25.7	0.860
Number of Observations	1133	1145	
	Analysis Sample		
Baseline algebra test score	24.7	24.8	0.304
Percent female	51.1	48.9	0.148
Percent African American	81.9	84.0	0.060
Percent Hispanic	13.8	12.1	0.061
Class size	25.8	26.2	0.549
Number of Observations	785	800	

Notes: All test scores are scaled scores converted to standard deviation units. The test for a difference in mean characteristic by random assignment is based on a regression of the characteristic on an indicator for random assignment and randomization pool fixed effects allowing for correlation in standard errors at the classroom level. We report the p-value for the t-test that the coefficient on the random assignment indicator equals zero.

Table 4a: Ordinary Least Squares and Instrumental Variable Estimates of the Effect of Computer-Assisted Instruction (CAI) on Algebra Achievement (without Teacher Fixed Effects)

	OLS					IV
	(1)	(2)	(3)	(4)	(5)	(6)
CAI	0.173 (0.076)	0.172 (0.074)	0.212 (0.077)	0.172 (0.060)	0.173 (0.059)	0.249 (0.086)
Baseline algebra test score				0.500 (0.035)	0.493 (0.034)	0.491 (0.034)
Female		0.081 (0.044)			0.095 (0.041)	0.087 (0.041)
African American		-0.671 (0.180)			-0.506 (0.137)	-0.498 (0.138)
Hispanic		-0.540 (0.211)			-0.390 (0.159)	-0.370 (0.159)
Observations	1872	1872	1585	1585	1585	1585

Notes: Each column represents a separate regression. Test scores are scaled scores converted to standard deviation units. Each regression also controls for the randomization pool as well as an indicator equal to one if sex is missing and an indicator equal to 1 if race/ethnicity is missing for those regressions that include demographic information. For the IV estimates of the effect of treatment on the treated we define treatment as completing at least one lesson in computerized algebra instruction. We report standard errors that allow for correlation within classroom in parentheses.

Table 4b: Ordinary Least Squares and Instrumental Variable Estimates of the Effect of Computer-Assisted Instruction (CAI) on Algebra Achievement (with Teacher Fixed Effects)

	OLS					IV
	(1)	(2)	(3)	(4)	(5)	(6)
CAI	0.373 (0.071)	0.367 (0.067)	0.423 (0.074)	0.284 (0.053)	0.283 (0.053)	0.417 (0.080)
Baseline algebra test score				0.483 (0.035)	0.477 (0.034)	0.468 (0.034)
Female		0.108 (0.041)			0.125 (0.041)	0.115 (0.041)
African American		-0.619 (0.155)			-0.449 (0.129)	-0.433 (0.131)
Hispanic		-0.498 (0.185)			-0.351 (0.154)	-0.315 (0.152)
Observations	1872	1872	1585	1585	1585	1585

Notes: Each column represents a separate regression. Test scores are scaled scores converted to standard deviation units. Each regression also controls for the randomization pool, an indicator equal to one if sex is missing, and an indicator equal to 1 if race/ethnicity is missing for those regressions that include demographic information, and teacher fixed effects. For the IV estimates of the effect of treatment on the treated we define treatment as completing at least one lesson in computerized algebra instruction. We report standard errors that allow for correlation within classroom in the parentheses. The p-values of the F-tests on the statistical significance of the teacher effects equal zero for all specifications.

Table 5: Amount of Time in the Computer Lab
by the Random Assignment of the Student's Class

	Random Assignment	
	Traditional Instruction	CAI
Number of lessons students are expected to complete based on the course level	52.7 (14.4)	55.3 (15.3)
Percent of students completing no lessons in CAI	80.1 (39.9)	9.1 (28.8)
Percent of students completing more than 10 lessons in CAI	14.8 (35.5)	83.8 (36.9)
Percent of students completing more than 20 lessons in CAI	10.3 (30.4)	70.3 (45.7)
Number of lessons completed in CAI	5.6 (15.2)	33.0 (23.9)
Number of CAI lessons completed as a percent of course expectations	10.0 (27.8)	64.5 (50.4)
Number of observations	785	800

Notes: District 1 has 62 school days in the study while classes in districts 2 and 3 generally have 180 days in the study. One exception is that a few classes in district 3 meet only one-half of the schools days.

Table 6a: Ordinary Least Squares Estimates of the Effect of Computer-Assisted Instruction (CAI) on Algebra and Mathematics Achievement in District 1 Using Different Tests

	Algebra Scale Score	2nd Qtr Benchmark Algebra Test	3rd Qtr Benchmark Algebra Test	State Mathematics Test
	(1)	(2)	(3)	(4)
	Maximum available sample			
CAI	0.226	0.381	0.604	0.260
	(0.071)	(0.127)	(0.286)	(0.119)
Observations	973	230	239	454
	Constraining sample students to be the same across specification			
CAI	0.374	0.462	0.946	0.381
	(0.168)	(0.173)	(0.482)	(0.139)
Observations	185	185	185	185

Notes: Standard errors that allow for correlation within classroom are in parentheses. The dependent variable in the first column is the normalized scale score for the algebra test; that in the second column is the 2nd quarter district-wide 8th-grade math test score; that in the third column is the 3rd quarter district-wide 8th-grade math test score; and that in the fourth column is the state mathematics test. All test scores are scale scores converted to standard deviation units. Each regression also includes controls for baseline test scores, the randomization pool, demographic characteristics, and an indicator equal to one if sex is missing, and an indicator equal to 1 if race/ethnicity is missing. The algebra and state mathematics tests were administered in the spring. The baseline algebra tests were given in the beginning of the academic year. The baseline benchmark algebra test was given in the 1st quarter of the academic year. The baseline state test was given in the spring of the preceding academic year.

Table 6b: Ordinary Least Squares Estimates of the Effect of Computer-Assisted Instruction (CAI) on Algebra and Mathematics Achievement in Districts 2 and 3 Using Different Tests

	District 2		District 3	
	Algebra Scale Score	State Mathematics Test	Algebra Scale Score	State Mathematics Test
	(1)	(2)	(3)	(4)
	Maximum sample available			
CAI	0.200	0.089	-0.124	-0.062
	(0.130)	(0.094)	(0.122)	(0.118)
Observations	412	341	200	199
	Constraining sample students to be the same across specification within school district			
CAI	0.400	0.082	0.031	-0.202
	(0.171)	(0.112)	(0.182)	(0.109)
Observations	229	229	107	107

Notes: Standard errors that allow for correlation within classroom are in parentheses. The dependent variable in the first and third columns is the normalized scale score for the algebra test; those in the second and fourth column results are the respective state mathematics test. All test scores are scale scores converted to standard deviation units. Each regression also includes controls for baseline test scores, the randomization pool, demographic characteristics, and an indicator equal to one if sex is missing, and an indicator equal to 1 if race/ethnicity is missing. The algebra tests were administered in the spring. The baseline algebra tests were given in the beginning of the fall. For district 2 the state mathematics test was administered in the spring of the students' 10th grade year. For district 3 the state mathematics test was administered in the fall of the students 10th grade year. For both districts, the baseline state tests were given in the fall of the students' 8th grade year.

Table 7: Differential Intent to Treat Effects of the Computerized Instruction on Pre-Algebra and Algebra Achievement by Class Type and Baseline Test Score Quartile

	All 3 Districts	Districts 1 and 2	District 1	District 2	District 3
	(1)	(2)	(3)	(4)	(5)
CAI effect for Algebra	0.005	0.069	0.130	-0.307	-0.230
	(0.059)	(0.065)	(0.066)	(0.218)	(0.100)
CAI effect for pre-Algebra	0.481	0.453	0.442	0.513	1.360
	(0.119)	(0.120)	(0.155)	(0.187)	(0.690)
CAI effect for bottom baseline test	0.216	0.288	0.280	0.136	-0.199
score quartile	(0.091)	(0.095)	(0.095)	(0.235)	(0.287)
CAI effect for 2nd baseline test	0.242	0.273	0.343	0.150	-0.090
score quartile	(0.100)	(0.104)	(0.115)	(0.212)	(0.282)
CAI effect for 3rd baseline test	0.171	0.199	0.090	0.522	0.004
score quartile	(0.105)	(0.117)	(0.125)	(0.260)	(0.161)
CAI effect for top quartile	0.155	0.245	0.218	0.358	-0.436
	(0.106)	(0.112)	(0.124)	(0.237)	(0.259)
Number of observations	1585	1385	973	412	200

Notes: Each column of each panel represents a separate regression. All test scores are scale scores converted to standard deviation units. Regressions in the top panel also includes baseline test scores. Each regression also controls for the randomization pool, demographic characteristics, an indicator equal to one if sex is missing, and an indicator equal to 1 if race/ethnicity is missing. Baseline test score quartiles are defined within district and class type (algebra or pre-algebra).We report standard errors that allow for correlation within classroom in the parentheses.

Table 8a: Differential Intent to Treat Effects of the Computerized Instruction on Pre-Algebra and Algebra Achievement by Individual Attendance Rates

	Districts 2 and 3	District 2	District 3
	(1)	(2)	(3)
CAI effect for bottom baseline attendance quartile	0.439 (0.287)	0.112 (0.395)	0.797 (0.353)
CAI effect for 2nd baseline attendance quartile	-0.221 (0.208)	-0.136 (0.328)	-0.578 (0.267)
CAI effect for 3rd baseline attendance quartile	-0.051 (0.175)	-0.053 (0.256)	-0.068 (0.293)
CAI effect for top baseline attendance quartile	-0.020 (0.197)	-0.119 (0.313)	0.146 (0.255)
Number of observations	372	221	151

Notes: Each column and panel represents a separate regression. Test scores are scaled scores converted to standard deviation units. Each regression also controls for the randomization pool, the baseline test scores, demographic characteristics, an indicator equal to one if sex is missing, and an indicator equal to 1 if race/ethnicity is missing. We report standard errors that allow for correlation within classroom in the parentheses. Each student's attendance rate is calculated as the percent of enrolled days that the student is in attendance. Attendance quartiles are calculated within district.

Table 8b: Differential Intent to Treat Effects of the Computerized Instruction on Pre-Algebra and Algebra Achievement by Class Characteristic: Attendance

	District 2 and District 3	District 2	District 3
CAI	2.131 (1.017)	2.261 (1.220)	2.808 (1.970)
CAI × Average class attendance	-0.025 (0.012)	-0.025 (0.014)	-0.034 (0.022)
Mean (std. deviation) of class attendance rate	83.287 (11.803)	82.513 (13.605)	84.695 (7.307)
Number of observations	564	364	200

Notes: See notes for table 9a. Average class attendance is based on individual student attendance data for the year preceding the year of the experiment.

Table 9: Differential Intent to Treat Effects of the Computerized Instruction on Pre-Algebra and Algebra Achievement by Class Size

	All 3 Districts	Districts 1 and 2	District 1	District 2	District 3
	(1)	(2)	(3)	(4)	(5)
CAI	-0.097	-0.281	-0.266	-0.035	-0.033
	(0.215)	(0.254)	(0.250)	(0.920)	(0.694)
CAI × Class size	0.010	0.020	0.019	0.011	-0.004
	(0.008)	(0.011)	(0.011)	(0.042)	(0.022)
Mean class size	26.005	25.623	26.420	23.740	28.650
(standard deviation)	(6.623)	(6.122)	(6.330)	(5.135)	(8.976)
Number of observations	1585	1385	973	412	200

Notes: Each column represents a separate regression. Test scores are scaled scores converted to standard deviation units. Each regression also controls for the randomization pool, the baseline test scores, demographic characteristics, an indicator equal to one if sex is missing, and an indicator equal to 1 if race/ethnicity is missing. We report standard errors that allow for correlation within classroom in the parentheses.

Table 10: Differential Intent to Treat Effects of the Computerized Instruction on Pre-Algebra and Algebra Achievement by Class Baseline Test Score Standard Deviation

	All 3 Districts	Districts 1 and 2	District 1	District 2	District 3
	(1)	(2)	(3)	(4)	(5)
CAI × baseline standard deviation for the class	0.110 (0.391)	-0.064 (0.387)	-0.118 (0.443)	0.600 (0.921)	0.583 (0.934)
	(6)	(7)	(8)	(9)	(10)
CAI × baseline standard deviation for the class	-1.100 (0.560	-0.620 (0.529)	-0.559 (0.594)	-0.514 (1.352)	-3.804 (0.369)
CAI × class baseline standard deviation × I(large class)	1.512 (0.892)	0.688 (0.870)	0.485 (0.907)	9.257 (2.085)	4.136 (0.711)
Mean class baseline standard deviation (standard deviation)	0.781 (0.160)	0.782 (0.154)	0.774 (0.157)	0.802 (0.147)	0.773 (0.193)
Number of observations	1585	1385	973	412	200

Notes: See notes for table 10. The coefficients in top and bottom panels are from different specifications. The median class size in the overall sample is 24 students. A large class is defined as having more than 24 students. A small class is defined as having 24 or fewer students.

Appendix Table 1: Numbers of Schools Classes, Teachers, and Randomization Pools

	Combined	District 1	District 2	District 3
	Full Sample			
Number of schools	17	10	4	3
Number of randomization pools	60	31	19	10
Number of classes	151	81	46	24
Number of teachers	61	39	15	7
Number of students	3541	1870	1062	609
	Analysis Sample			
Number of schools	17	10	4	3
Number of randomization pools	60	31	19	10
Number of classes	141	74	44	23
Number of teachers	57	36	14	7
Number of students	1585	973	412	200

Appendix Table 2a: Randomization of Treatment and Control (Using Full Sample)

	Random Assignment		
	Traditional Instruction	Computerized Instruction	p-value of difference
District #1			
Baseline algebra test score	24.6	24.7	0.285
Baseline state test score	9.2	9.2	0.990
Baseline district test score	3.0	3.7	0.107
Female	51.5	47.8	0.128
African American	98.0	97.8	0.260
Hispanic	0.6	0.8	0.821
Class size	25.3	25.5	0.949
District #2			
Baseline algebra test score	24.6	24.7	0.823
Baseline state test score	6.6	6.7	0.558
Female	43.9	44.8	0.561
African American	51.3	44.8	0.566
Hispanic	42.6	48.1	0.204
Class size	24.1	24.6	0.369
District #3			
Baseline algebra test score	25.0	24.9	0.904
Baseline state test score	16.7	16.7	0.992
Female	43.2	48.2	0.482
African American	92.7	95.6	0.126
Hispanic	0.7	0.8	0.792

Class size	30.4	28.0	0.547

Notes: All test scores are scaled scores converted to standard deviation units. The test for a difference in mean characteristic by random assignment is based on a regression of the characteristic on an indicator for random assignment and randomization pool fixed effects allowing for correlation in standard errors at the classroom level. We report the p-value for the t-test that the coefficient on the random assignment indicator equals zero. For district #1: baseline algebra test scores are available for 700 treatment students and 624 controls; baseline state test scores are available for 474 treatment students and 387 controls; baseline district test scores are available for 110 treatment students and 147 controls; and demographic data are available for 831 treatment students and 689 controls. For district #2: baseline algebra test scores are available for 280 treatment students and 351 controls; baseline state test scores are available for 243 treatment students and 348 controls; and demographic data are available for 397 treatment students and 556 controls. For district #3: baseline algebra test scores are available for 165 treatment students and 158 controls; baseline state test scores are available for 151 treatment students and 172 controls; and demographic data are available for 249 treatment students and 287 controls.

Appendix Table 2b:
Assessing Random Assignment with the Analysis Sample

	Random Assignment		
	Traditional Instruction	Computerized Instruction	p-value of difference
District #1			
Baseline algebra test score	24.7	24.7	0.487
Baseline state test score	9.3	9.5	0.854
Baseline district test score	3.2	3.7	0.093
Female	53.6	50.6	0.092
African American	96.9	97.2	0.239
Hispanic	0.7	1.1	0.977
Class size	26.0	26.8	0.481
District #2			
Baseline algebra test score	24.7	25.0	0.274
Baseline state test score	6.7	6.9	0.200
Female	48.0	45.1	0.634
African American	49.3	44.5	0.061
Hispanic	43.7	46.2	0.054
Class size	23.5	24.0	0.353
District #3			
Baseline algebra test score	25.1	25.0	0.320
Baseline state test score	16.9	16.8	0.437
Female	48.0	47.5	0.808
African American	94.0	94.9	0.290
Hispanic	0.0	1.0	0.161
Class size	30.2	27.1	0.462

Appendix Table 3:
Cost Comparisons

The cost of CAI

School	Number of Classes	Total number of Students	Class size	Periods	CAI class size	CAI labs needed	Annual cost per lab	Cost per student
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
School A	22	730	33.2	8	30.0	3.0	$52,381	$218
School B	12	321	26.8	8	26.8	1.5	$52,381	$245
District 1 analysis sample	74	1736	23.5	8	23.5	9.3	$52,381	$279

The cost of reducing class size to 13 students

School	Number of Classes	Total number of Students	Class size	Periods	New total math classes	New teachers required	Salary + benefits per teacher	Cost per student
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
School A	22	730	33.2	6	56.2	5.7	$42,143	$329
School B	12	321	26.8	6	24.7	2.1	$42,143	$278
District 1 analysis sample	74	1736	23.5	6	133.5	9.9	$42,143	$241

Notes: The information on number of classes and number of students for schools A and B apply to all algebra and pre-algebra classes in the school while the information on the number of classes and students for the analysis samples only applies to classes that are represented in our analysis sample. The number of CAI labs needed equals the total number of students divided by the number of students each lab serves each day. We assume that the computer lab can be used for the number of periods specified in column (4) of the top panel and that each CAI class is equal to average class size with a maximum of 30 students (column 5). We assume the cost of the lab equals $250,000 in fixed costs plus $50,000 every 3 years for training, support, and maintenance and that the lab will be good for 7 years. New total math classes in column (5) of the bottom panel equals the number of math classes needed for an average class size of 13 students. Assuming each teacher teaches the number of periods in column (4), column (6) represents the number of new teachers needed to reduce class size to 13 students. Salary is based on the salary schedule for teachers in district 1 with no experience. We assume that salary equals 70 percent of total compensation costs.

Working Paper Series

A series of research studies on regional economic issues relating to the Seventh Federal Reserve District, and on financial and economic topics.

Working Paper Series *(continued)*

Working Paper Series *(continued)*

Firm-Specific Capital, Nominal Rigidities and the Business Cycle *David Altig, Lawrence J. Christiano, Martin Eichenbaum and Jesper Linde*	**WP-05-01**
Do Returns to Schooling Differ by Race and Ethnicity? *Lisa Barrow and Cecilia Elena Rouse*	**WP-05-02**
Derivatives and Systemic Risk: Netting, Collateral, and Closeout *Robert R. Bliss and George G. Kaufman*	**WP-05-03**
Risk Overhang and Loan Portfolio Decisions *Robert DeYoung, Anne Gron and Andrew Winton*	**WP-05-04**
Characterizations in a random record model with a non-identically distributed initial record *Gadi Barlevy and H. N. Nagaraja*	**WP-05-05**
Price discovery in a market under stress: the U.S. Treasury market in fall 1998 *Craig H. Furfine and Eli M. Remolona*	**WP-05-06**
Politics and Efficiency of Separating Capital and Ordinary Government Budgets *Marco Bassetto with Thomas J. Sargent*	**WP-05-07**
Rigid Prices: Evidence from U.S. Scanner Data *Jeffrey R. Campbell and Benjamin Eden*	**WP-05-08**
Entrepreneurship, Frictions, and Wealth *Marco Cagetti and Mariacristina De Nardi*	**WP-05-09**
Wealth inequality: data and models *Marco Cagetti and Mariacristina De Nardi*	**WP-05-10**
What Determines Bilateral Trade Flows? *Marianne Baxter and Michael A. Kouparitsas*	**WP-05-11**
Intergenerational Economic Mobility in the U.S., 1940 to 2000 *Daniel Aaronson and Bhashkar Mazumder*	**WP-05-12**
Differential Mortality, Uncertain Medical Expenses, and the Saving of Elderly Singles *Mariacristina De Nardi, Eric French, and John Bailey Jones*	**WP-05-13**
Fixed Term Employment Contracts in an Equilibrium Search Model *Fernando Alvarez and Marcelo Veracierto*	**WP-05-14**
Causality, Causality, Causality: The View of Education Inputs and Outputs from Economics *Lisa Barrow and Cecilia Elena Rouse*	**WP-05-15**

Working Paper Series *(continued)*

Competition in Large Markets — **WP-05-16**
Jeffrey R. Campbell

Why Do Firms Go Public? Evidence from the Banking Industry — **WP-05-17**
Richard J. Rosen, Scott B. Smart and Chad J. Zutter

Clustering of Auto Supplier Plants in the U.S.: GMM Spatial Logit for Large Samples — **WP-05-18**
Thomas Klier and Daniel P. McMillen

Why are Immigrants' Incarceration Rates So Low?
Evidence on Selective Immigration, Deterrence, and Deportation — **WP-05-19**
Kristin F. Butcher and Anne Morrison Piehl

Constructing the Chicago Fed Income Based Economic Index – Consumer Price Index:
Inflation Experiences by Demographic Group: 1983-2005 — **WP-05-20**
Leslie McGranahan and Anna Paulson

Universal Access, Cost Recovery, and Payment Services — **WP-05-21**
Sujit Chakravorti, Jeffery W. Gunther, and Robert R. Moore

Supplier Switching and Outsourcing — **WP-05-22**
Yukako Ono and Victor Stango

Do Enclaves Matter in Immigrants' Self-Employment Decision? — **WP-05-23**
Maude Toussaint-Comeau

The Changing Pattern of Wage Growth for Low Skilled Workers — **WP-05-24**
Eric French, Bhashkar Mazumder and Christopher Taber

U.S. Corporate and Bank Insolvency Regimes: An Economic Comparison and Evaluation — **WP-06-01**
Robert R. Bliss and George G. Kaufman

Redistribution, Taxes, and the Median Voter — **WP-06-02**
Marco Bassetto and Jess Benhabib

Identification of Search Models with Initial Condition Problems — **WP-06-03**
Gadi Barlevy and H. N. Nagaraja

Tax Riots — **WP-06-04**
Marco Bassetto and Christopher Phelan

The Tradeoff between Mortgage Prepayments and Tax-Deferred Retirement Savings — **WP-06-05**
Gene Amromin, Jennifer Huang,and Clemens Sialm

Why are safeguards needed in a trade agreement? — **WP-06-06**
Meredith A. Crowley

Working Paper Series *(continued)*

Taxation, Entrepreneurship, and Wealth **WP-06-07**
Marco Cagetti and Mariacristina De Nardi

A New Social Compact: How University Engagement Can Fuel Innovation **WP-06-08**
Laura Melle, Larry Isaak, and Richard Mattoon

Mergers and Risk **WP-06-09**
Craig H. Furfine and Richard J. Rosen

Two Flaws in Business Cycle Accounting **WP-06-10**
Lawrence J. Christiano and Joshua M. Davis

Do Consumers Choose the Right Credit Contracts? **WP-06-11**
Sumit Agarwal, Souphala Chomsisengphet, Chunlin Liu, and Nicholas S. Souleles

Chronicles of a Deflation Unforetold **WP-06-12**
François R. Velde

Female Offenders Use of Social Welfare Programs Before and After Jail and Prison:
Does Prison Cause Welfare Dependency? **WP-06-13**
Kristin F. Butcher and Robert J. LaLonde

Eat or Be Eaten: A Theory of Mergers and Firm Size **WP-06-14**
Gary Gorton, Matthias Kahl, and Richard Rosen

Do Bonds Span Volatility Risk in the U.S. Treasury Market?
A Specification Test for Affine Term Structure Models **WP-06-15**
Torben G. Andersen and Luca Benzoni

Transforming Payment Choices by Doubling Fees on the Illinois Tollway **WP-06-16**
Gene Amromin, Carrie Jankowski, and Richard D. Porter

How Did the 2003 Dividend Tax Cut Affect Stock Prices? **WP-06-17**
Gene Amromin, Paul Harrison, and Steven Sharpe

Will Writing and Bequest Motives: Early 20th Century Irish Evidence **WP-06-18**
Leslie McGranahan

How Professional Forecasters View Shocks to GDP **WP-06-19**
Spencer D. Krane

Evolving Agglomeration in the U.S. auto supplier industry **WP-06-20**
Thomas Klier and Daniel P. McMillen

Mortality, Mass-Layoffs, and Career Outcomes: An Analysis using Administrative Data **WP-06-21**
Daniel Sullivan and Till von Wachter

Working Paper Series *(continued)*

Working Paper Series *(continued)*

Conflict of Interest and Certification in the U.S. IPO Market — **WP-07-09**
Luca Benzoni and Carola Schenone

The Reaction of Consumer Spending and Debt to Tax Rebates –
Evidence from Consumer Credit Data — **WP-07-10**
Sumit Agarwal, Chunlin Liu, and Nicholas S. Souleles

Portfolio Choice over the Life-Cycle when the Stock and Labor Markets are Cointegrated — **WP-07-11**
Luca Benzoni, Pierre Collin-Dufresne, and Robert S. Goldstein

Nonparametric Analysis of Intergenerational Income Mobility — **WP-07-12**
with Application to the United States
Debopam Bhattacharya and Bhashkar Mazumder

How the Credit Channel Works: Differentiating the Bank Lending Channel — **WP-07-13**
and the Balance Sheet Channel
Lamont K. Black and Richard J. Rosen

Labor Market Transitions and Self-Employment — **WP-07-14**
Ellen R. Rissman

First-Time Home Buyers and Residential Investment Volatility — **WP-07-15**
Jonas D.M. Fisher and Martin Gervais

Establishments Dynamics and Matching Frictions in Classical Competitive Equilibrium — **WP-07-16**
Marcelo Veracierto

Technology's Edge: The Educational Benefits of Computer-Aided Instruction — **WP-07-17**
Lisa Barrow, Lisa Markman, and Cecilia Elena Rouse