NCEE 2009-4041 U.S. DEPARTMENT OF EDUCATION

# Effectiveness of Reading and Mathematics Software Products

## Findings From Two Student Cohorts

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

# Effectiveness of Reading and Mathematics Software Products

## Findings From Two Student Cohorts

**February 2009**

**Larissa Campuzano**
**Mark Dynarski**
**Roberto Agodini**
**Kristina Rall**
Mathematica Policy Research, Inc.

**Audrey Pendleton**
*Project Officer*
Institute of Education Sciences

NCEE 2009-4041
U.S. DEPARTMENT OF EDUCATION

NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

**U.S. Department of Education**
Arne Duncan
*Secretary*

**Institute of Education Sciences**
Sue Betka
*Acting Director*

**National Center for Education Evaluation and Regional Assistance**
Phoebe Cottingham
*Commissioner*

**February 2009**

The report was prepared for the Institute of Education Sciences under Contract No. ED-01CO0039/0007. The project officer is Audrey Pendleton in the National Center for Education Evaluation and Regional Assistance.

IES evaluation reports present objective information on the conditions of implementation and impacts of the programs being evaluated. IES evaluation reports do not include conclusions or recommendations or views with regard to actions policymakers or practitioners should take in light of the findings in the reports.

**To order copies of this report,**

- Write to ED Pubs, Education Publications Center, U.S. Department of Education, P.O. Box 1398, Jessup, MD 20794-1398.
- Call in your request toll free to 1-877-4ED-Pubs. If 877 service is not yet available in your area, call 800-872-5327 (800-USA-LEARN). Those who use a telecommunications device for the deaf (TDD) or a teletypewriter (TTY) should call 800-437-0833.
- Fax your request to 301-470-1244.
- Order online at www.edpubs.org.

This report also is available on the IES website at http://ies.ed.gov/ncee.

Upon request, this report is available in alternate formats such as Braille, large print, audiotape, or computer diskette. For more information, please contact the Department's Alternate Format Center at 202-260-9895 or 202-205-8113.

# Acknowledgments

The study team wants to thank the many districts, schools, teachers, students, and publishing companies who took part in this effort. Their cooperation, support, and willingness to be part of a random-assignment experiment are much appreciated.

Staff from SRI International had a significant role in shaping the study and in designing and collecting the teacher training data described in this report. We thank Barbara Means and Robert Murphy for leading that effort. The team also wants to thank the study's Technical Working Group for its suggestions and support. The members of the Technical Working Group are Jere Brophy, Thomas Cook, Douglas Clements, John Cooney, Michael Kamil, Chris Lonigan, Gary Phillips, and Steven Ross.

Led by Nancy Carey, a large team of survey researchers, data collectors, and programmers worked to design, gather, and process the data. We thank them for their efforts. We also thank our word-processing and editing staff for their expert support.

# Disclosure of Potential Conflicts of Interest

The research team for this evaluation consists of a prime contractor, Mathematica Policy Research, and a major subcontractor, SRI International. Mathematica and its key staff have no financial interests that could be affected by findings from the study. None of the Technical Working Group members have financial interests that could be affected by findings from the study.

# Contents

**Chapter** **Page**

# Tables

**Table** **Page**

# Figures

# Executive Summary

# Effectiveness of Reading and Math Software Products: Findings from Two Student Cohorts

In the No Child Left Behind Act, Congress called for the U.S. Department of Education (ED) to conduct a rigorous study of the conditions and practices under which educational technology is effective in increasing student academic achievement. A 2003 design effort by ED working with educational technology and research experts recommended focusing the study on software products used to support reading and math instruction. The study team set up a competitive process and worked with ED to select reading products to be studied in the first and fourth grades, pre-algebra products in the sixth grade, and algebra I products in high school (or possibly in middle school). The team implemented an experimental design in which teachers in the same school were randomly assigned to use or not to use a software product, and the team collected test scores and other data to assess effectiveness of the software products.

A report was released in April 2007 presenting study findings for the 2004-2005 school year (Dynarski et al. 2007). The findings indicated that, after one school year, differences in student test scores were not statistically significant between classrooms that were randomly assigned to use products and those that were randomly assigned not to use products. School and teacher characteristics generally were not related to whether products were effective.

The study also collected test scores and other data in the 2005-2006 school year, in which teachers who continued with the study had a new cohort of students and a year of experience using software products. Data from the second cohort enable the study to address the question of whether software products are more effective in raising test scores after teachers have a year of experience using them.

The first-year report presented average effects of four groups of products on student test scores, which supported assessing whether products were effective in general. School districts and educators purchase individual products, however, and knowing whether individual products are effective is important for making decisions supported by evidence. This report presents findings on the effects of 10 products on student test scores.

### Study Design

The second year of the study was shaped by the structure of its first year. For the first year, the study team identified 16 products for the study, as noted above, and recruited 33 school districts to implement the products. In turn, districts identified a total of 132 schools to implement the selected products, and the study randomly assigned 428 volunteering teachers in the schools to either use or not use the products in their classrooms. Students were allocated to classrooms by their schools in whatever manner schools conventionally used. Students were tested in these classrooms in both the fall and spring of the 2004-2005 school year (a total of 9,458 students). The study also observed classrooms at three points during the school year, and supplemented the test scores and observational data with data about students from school records, a questionnaire completed by teachers in the study, and school data from the *Common Core of Data* maintained by the National Center for Education Statistics (NCES).

Collecting a second year of student data, while staying within resource constraints, required modifying the study in five ways compared to the first year. Products that had been implemented in only a few schools were dropped, classrooms were not observed in the second year, one treatment classroom and one control classroom were randomly sampled within schools that had more than one, some districts provided their test scores rather than having the study team test students, and some items were collected from school records. These changes in the data collection strategy reduced the amount of data collected in the 2005-2006 school year, and precluded the study from exploring the same range of questions it explored in the first year. The second year of the study included 10 products, 23 districts, 77 schools, 176 teachers, and 3,280 students.

The second-year study also should be understood as two different but related sub-studies. One objective of the second-year study is to assess whether the experience of a second year of use of software products increased the effects products had on student test scores. Another objective is to report on the effectiveness of individual software products in raising student test scores. Addressing the first objective requires restricting the sample to teachers who participated in both the first and second years of the study. Addressing the second objective requires data from teachers who participated in either the first or second year. Because the samples of teachers and students differ between the two substudies, estimates of sample characteristics and product effects also differ.

### Collecting Achievement and Product Usage Data

The study's analyses rely on data from student test scores. Scores came from two sources. The data collection strategy was to collect district scores to the extent they were available and were consistent with the study's analytic approach, and for the study to administer its own tests if districts could not provide a fall or spring score (the study used the previous spring scores in place of fall scores if districts could provide them). In first, fourth, and sixth grades, if districts did not administer a standardized test with national norms in a grade level, the study administered a student test in the fall and spring of the 2005-2006 school year. It used the Stanford Achievement Test (version 9) reading battery for first graders, the Stanford Achievement Test (SAT-10) reading battery for fourth graders, and the Stanford Achievement Test (SAT-10) math battery for sixth graders. The study used the Educational Testing Services' (ETS) End-of-Course Algebra Assessment (1997) for algebra I (which is not administered by districts in the study).

For district tests, in first grade one district provided scores on the Iowa Tests of Basic Skills for fall scores, and another district provided scores on the Stanford Achievement Test for spring scores. For fourth grade, one district provided scores on the Iowa Tests of Basic Skills as fall scores. For sixth grade, one district provided fall scores on the Iowa Tests of Basic Skills and another provided fall and spring scores on the New Mexico Standards Based Assessment. For algebra I, one district provided fall scores on the Iowa Tests of Basic Skills. With the exception of scores on the ETS algebra test, scores were converted to normal curve equivalent (NCE) units to standardize the measures across tests and cohorts. Algebra I scores for the ETS test are reported as percent correct.

Data from product records provided information about usage of the products. Eight of the 10 products included in the study used databases to track the time when each student was logged on. The usage measure reported in the study is actual student logged-on time for a school year, as reported by the product database. Usage by more than one student at a time, such as in a group activity, is counted only for the logged-on student. Time spent doing activities that are related to product use but occur when students are not logged on, such as reading materials related to a computer lesson, is not counted as usage.

**Software Products**

The products included in the second year are a subset of the products used in the first year. Some products that had been studied in the first year had been implemented in too few schools for individual effects to be reported on them in the second year. For two products that were just below the threshold needed for reasonable sample sizes, the study added districts and schools in the second year.

The second-year study included four reading software products for first grade, Destination Reading (Riverdeep 2008), the Waterford Early Reading Program (Pearson School 2008), Headsprout (Headsprout 2008), and Plato Focus (Plato Learning Corporation 2008). Three of the four products provided supplemental instruction and Plato Focus was used as the core reading curriculum. The second-year study also included two reading products for fourth grade, LeapTrack (LeapFrog Schoolhouse 2008) and Academy of Reading (Autoskill International 2008). These products supplemented the core reading curriculum with tutorials, practice, and assessment geared to specific reading skills.

For math, the second-year study included two math products for sixth grade, Larson Pre-Algebra (Houghton-Mifflin 2008) and Achieve Now (Plato Learning 2008). The products supplemented the core math curriculum with provided tutorial and practice opportunities and assessed student skills. The study included two algebra I products: Cognitive Tutor Algebra I (Carnegie Learning 2008) and Larson Algebra I (Houghton-Mifflin 2008). The Larson product supplemented algebra I instruction and the Cognitive Tutor product was the core algebra I curriculum. Students at a variety of high school grade levels can take algebra I, and many middle schools also teach algebra I. In the study, 9 percent of students were in eighth grade, 87 percent were in ninth grade, and 4 percent were in grades 10, 11, or 12.

Standardized Tests the Study Used to Measure Achievement Outcomes

First grade reading test: The version-9 reading battery of the Stanford Achievement Test (Pearson 1996a).

Fourth grade reading test: The version-10 reading battery of the Stanford Achievement Test (Pearson 2003b).

Sixth grade math test: The version-10 math battery of the Stanford Achievement Test (Pearson 2003c).

Algebra test: The Educational Testing Service (ETS) End-of-Course Algebra Assessment (Educational Testing Service 1997).

The reading and math products supplemented the core curriculum or, as was the case for Cognitive Tutor, were the core curriculum. Products generally were for whole classes and were not implemented only to remediate skills for students who were lagging.

### Findings from First Year of Study

The implementation analysis for the first-year study focused on how products were used in classrooms, their extent of usage, issues that resulted from their use, and how their use affected classroom activities. The analysis found that nearly all teachers received training on using products and believed the training prepared them adequately to use them. Technical difficulties using products mostly were minor. They included issues with students logging in, computers locking up, or hardware problems such as headphones not working. Most of the technical difficulties were easily corrected or worked around. When asked whether they would use the products again, 88 to 92 percent of teachers indicated that they would (the percentage depended on the grade level).

Comparing student test scores for treatment teachers using study products and control teachers not using study products is the study's measure of product effectiveness. Effects on test scores were estimated using a statistical model that accounts for correlations of students within classrooms and classrooms within schools. Below we summarize the key first-year findings.

### First-Year Effects of First Grade Technology Products

- **Effects on Test Scores Were Not Statistically Different from Zero.** Overall reading scores for students in treatment and control classrooms were 50.2 and 49.5, respectively (in normal curve equivalent units).[1] The difference was not statistically different from zero.

- **Most School and Classroom Characteristics Were Uncorrelated with Effects.** Classroom characteristics (teaching experience, teacher gender, teacher education level, whether there were problems getting access to the product, whether teachers had adequate time to prepare to use the product, whether the product was used in the classroom, and whether the teacher participated in technology professional development in the past year) were not correlated with product effects for the overall SAT-9 score. School characteristics (percentage of students eligible for free lunch, whether the school is in an urban area, percentage of students that were African American, percentage that were Hispanic) also were not correlated with product effects on the overall SAT-9 score. The one exception was the student-teacher ratio. Time of study product usage did not have a statistically significant correlation with effects for the overall score or subtest scores.

### First-Year Effects of Fourth Grade Technology Products

- **Differences in Test Scores Were Not Statistically Different from Zero.** Overall reading scores for students in treatment and control classrooms were 42.1 and 41.7, respectively (in normal curve equivalent units). The difference was not statistically different from zero.

- **Some Classroom and School Characteristics Were Correlated with Product Effects.** For the overall score, a statistically significant correlation was found between product effects and product usage. For the word study skills score, statistically significant correlations were found between product effects and teaching experience, whether the product was used in the classroom, whether teachers had participated in technology professional development, and the percentage of students that were African American.

### First-Year Effects of Sixth Grade Technology Products

- **Effects on Test Scores Were Not Statistically Different from Zero.** Overall math scores for students in treatment and control classrooms were 52.2 and 50.8, respectively (in normal curve equivalent units). The difference was not statistically different from zero.

- **School and Classroom Characteristics Were Not Related to Product Effects.** Time of product use and other school and classroom characteristics were uncorrelated with product effects.

[1] A normal curve equivalent (NCE) score converts the scaled test score into the range 1 to 99, with 50 being the average for the nationally normed sample. Unlike percentiles, NCE scores can be averaged, which makes them more appropriate for statistical analyses and estimation of product effects.

### First-Year Effects of Algebra I Technology Products

- **Effects on Test Scores Were Not Statistically Different from Zero.** Overall math scores for students in treatment and control classrooms were 37.3 percent correct and 38.1 percent correct, respectively. The difference was not statistically different from zero.

- **Classroom and School Characteristics Were Uncorrelated with Product Effects.** The algebra I study included fewer schools, which limited the ability to estimate moderator effects. None of the classroom and school characteristics included in the model was statistically significant.

### Does Experience Increase Product Effects?

The first hypothesis addressed in the second year of the study is whether product effects on student test scores are larger in the second year than the first, after teachers have had one year to use products in their classrooms. To test the hypothesis, the study created a merged data file that was restricted to 115 teachers who continued with the study for a second year (27 percent of the number that participated in the first year). Teachers who moved to other schools or grade levels, or left teaching, did not continue with the study. The merged file included 5,345 students combined across the first year and the second year for the 115 teachers.

The study estimated statistical models in which student test scores were related to treatment status (whether the teacher was assigned to use a product). To test the effect of experience, the models estimated product effects on student test scores in each of the two years, and then tested statistically to determine if the two differed by more than what would be expected due to sampling variance. The models also included student fall test scores, age, and gender; and teacher experience and education level. Effects of individual products are not reported.

Figure 1 shows experience effects, which are the difference between the second-year effect of products on test scores and the first-year effect, for the reading products used in first and fourth grades. Figure 2 shows the experience effects for the math products used in sixth grade and algebra I. These figures show product effects in each of the two years, and the arrow between the product effects represents the experience effect (the difference between second-year and first-year effects).

Evidence is mixed for the hypothesis that an additional year of experience using the software products improves product effects on test scores. In first grade, the measured product effect in the second year is not statistically significantly different from the product effect in the first year. Similarly, in fourth grade, the measured product effect in the second year is not statistically significantly larger than the effect in the first year. In sixth grade, the product effect in the second year is more negative than in the first year (the effect is negative in both years) and the difference between the two negative effects is statistically significant. In algebra I, the product effect in the second year is larger than in the first year and the difference is statistically significant.

**Figure 1. Reading Product Effects Differences in the First and Second Years**

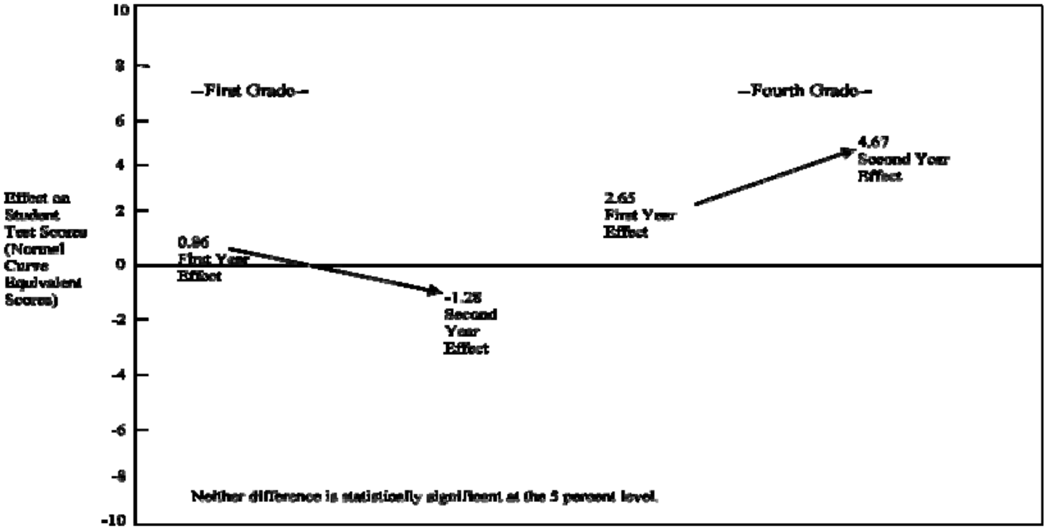
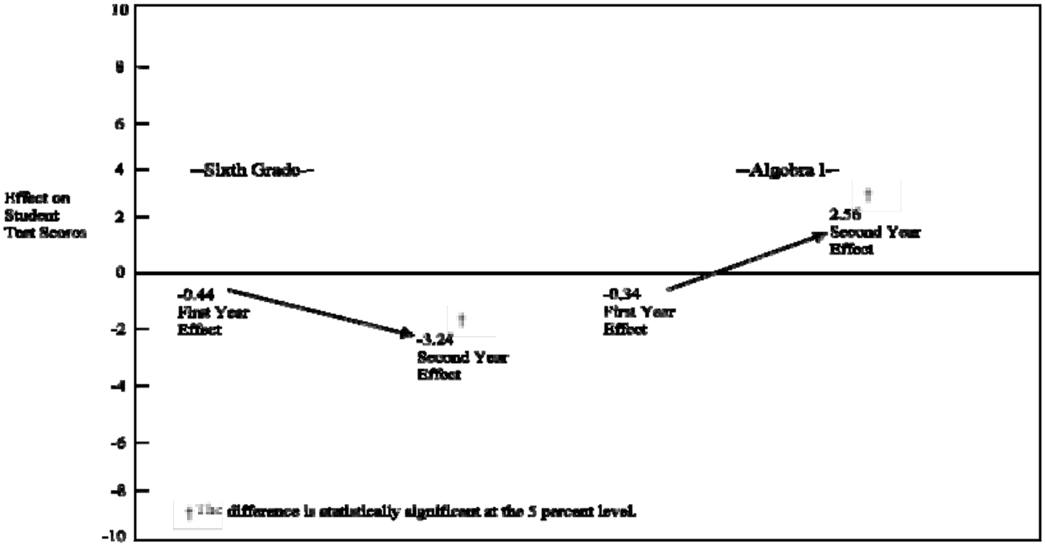


**Figure 2. Math Product Effects Differences in the First and Second Years**



The study investigated the relationship between product usage and product effects in the two years. Usage data were gathered from product records and are accurate to the extent that student logged-in time represents product usage. (If students used other materials related to the product while not being logged on, the additional time is not reflected in the usage data.) Average first grade student usage went from 2,556 minutes in the first year to 1,182 minutes in the second year. Average fourth grade student usage went from 720 minutes in the first year to 936 minutes in the second year. Average sixth grade student usage went from 852 minutes in the first year to 678 minutes in the second year. Average algebra I student usage went from 1,308 minutes in the first year to 1,452 minutes in the second year. All differences between years were statistically significant. The relationship between changes in effects between the two years and changes in usage was not statistically significant.

Because the study did not observe classrooms or interview teachers in the second year, it has no information about how teachers may have modified their use of products from one year to the next beyond examining usage times that are captured by the products being studied. For the same reason, the study has no information about whether control group teachers modified their use of other software products in their classrooms.

### Effects of Individual Products

Another objective of the study's second year is to report effects of software products separately. As done in the analysis of experience effects, the study used statistical models to estimate product effects on student test scores that accounted for student fall test scores, age, and gender, and teacher experience and education. Data for all students, teachers, and schools who participated in the study either in the first or second year were used in the analysis. Models were estimated separately for each of the 10 products.

Figure 3 presents the results for six reading products, with the product effect displayed in the middle of its 95 percent confidence interval. The product effect in Figure 3 is the estimated difference in student test scores between classrooms using products and classrooms not using products in the two years of the study. For example, the effect shown for Destination Reading means that an average first grade student in a classroom that used Destination Reading is estimated to have a spring test score that is higher by 1.91 NCE units than if the student were in a classroom not using that product. This effect is equivalent to moving an average student from the 50th percentile on the test score to the 54th percentile. A positive and statistically significant effect was found for one of the six reading products (Leap Track, fourth grade). The remaining five product effects were not statistically significant. Of these, four were positive and one was negative.

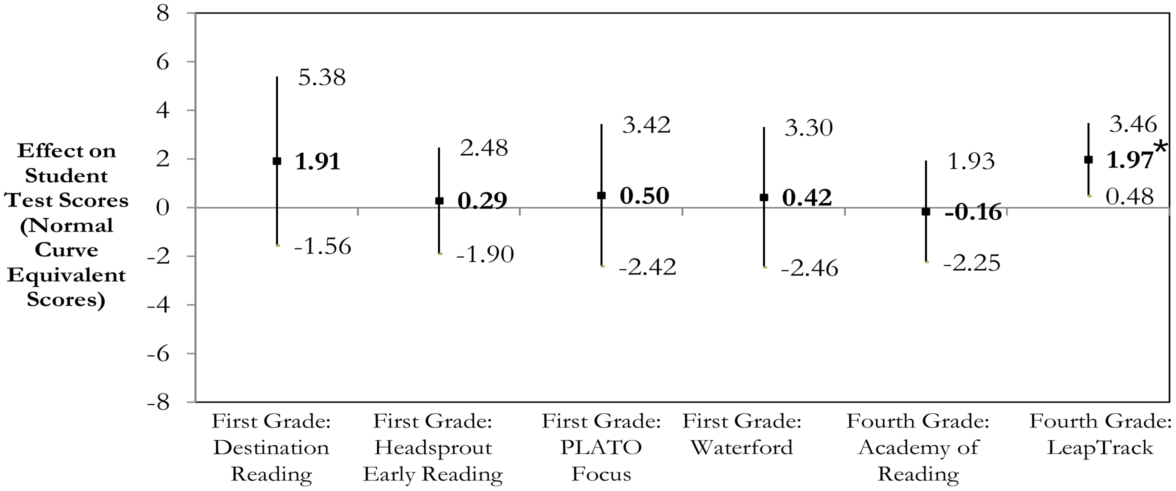
Figure 4 presents analogous results for the four math products. None of the effects is statistically significant. Three of the estimated effects were negative and one was positive.

Presenting product effects on test scores in this way does not mean that the study results indicate that products with larger estimated effects are more desirable than products with smaller estimated effects. Characteristics of districts and schools that volunteered to implement the products differ, and these differences may relate to product effects in important ways. The findings do not adjust for differences in schools and districts that go beyond measured characteristics but may be related to outcomes.

### Summary

The second year of the study examined whether an additional year of teaching experience using the software products increased the estimated effects of software products on student test scores. The evidence for this hypothesis is mixed. For reading, there were no statistically significant differences between the effects that products had on standardized student test scores in the first year and the second year. For sixth grade math, product effects on student test scores were statistically significantly lower (more negative) in the second year than in the first year, and for algebra I, effects on student test scores were statistically significantly higher in the second year than in the first year.

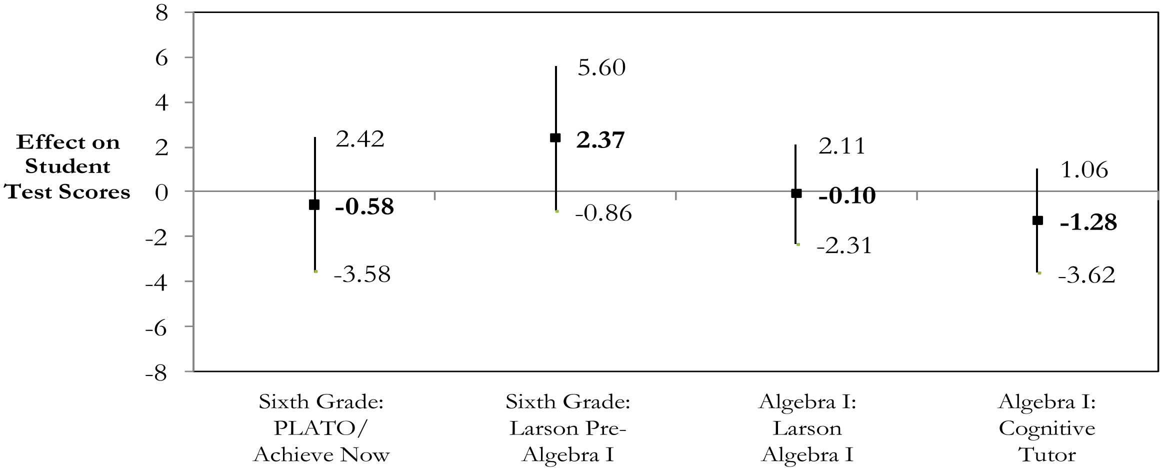
Figure 3. Effects of Reading Software Products



Note: Vertical lines represent 95 percent confidence intervals.

*The effect is statistically significant at the 5 percent level.

Figure 4. Effects of Math Software Products



Note: Vertical lines represent 95 percent confidence intervals.

Test scores are normal curve equivalent units for sixth grade and percent correct units for algebra.

The study also tested whether using any of the 10 software products increased student test scores. One product had a positive and statistically significant effect. Nine did not have statistically significant effects on test scores. Five of the insignificant effects were negative and four were positive.

The study's findings should be interpreted in the context of its design and objectives. It examined a range of reading and math software products in a range of diverse school districts and schools. But it did not study many forms of educational technology and it did not include many types of software products. How much information the findings provide about the effectiveness of products that are not in the study is an open question. Products in the study also were implemented in a specific set of districts and schools, and other districts and schools may have different experiences with the products. The findings should be viewed as one element within a larger set of research studies that have explored the effectiveness of software products.

# Chapter I

# Introduction

In the No Child Left Behind Act, Congress called for the U.S. Department of Education (ED) to conduct a rigorous study of the conditions and practices under which educational technology is effective in increasing student academic achievement. A 2003 design effort by ED working with educational technology and research experts recommended focusing the study on software products used to support reading and math instruction. The study team set up a competitive process and worked with ED to select reading products to be studied in the first and fourth grades, pre-algebra products in the sixth grade, and algebra I products in high school and possibly in middle school. The study team implemented the products in a range of school districts and schools and collected data at the beginning and end of the 2004-2005 school year.

Over the past two decades, numerous studies comparing computer-based and conventional instruction in reading and mathematics have been conducted. Both qualitative research syntheses (Schacter 2001; Sivin-Kachala 1998) and formal meta-analyses of these studies (Blok et al. 2002; Kulik and Kulik 1991; Kulik 1994; Kulik 2003; Murphy et al. 2001; Pearson et al. 2005; Waxman et al. 2003) found that computer-assisted instruction in reading and mathematics generally had a positive effect. Kulik's 1994 meta-analysis, for example, found a positive effect on test scores (the effect size—the effect as a proportion of the standard deviation of test scores—was 0.30).

Murphy et al. (2001) examined a wide range of research studies from the published literature and from software vendors. Of the 195 experimental or quasi-experimental studies conducted between 1993 and 2000 that met the criteria for inclusion, 31 studies met minimum methodological requirements for inclusion in the synthesis. For these studies, researchers estimated an average effect size of .35 for reading and .45 for mathematics.

Despite the fairly sizable number of studies and generally positive findings, researchers have noted weaknesses or design flaws in many of the studies (Murphy et al. 2001; Pearson et al. 2005). Of the technology studies reviewed by Waxman et al. (2003), for example, half had sample sizes of fewer than 50 students. Many studies had no control groups or equivalent comparison groups, leading to questionable validity for claims of effects. Studies with stronger research designs showed smaller effects (Pearson et al. 2005).

Against this backdrop, the national study was designed to be both rigorous and large, enabling it to make causal statements about the effects of technology with a reasonable degree of statistical precision. The study's first report was released in April 2007 (Dynarski et al. 2007) and indicated that, after one school year, differences in student test scores were not statistically significant between classrooms that were randomly assigned to use products and those that were randomly assigned not to use products. School and teacher characteristics generally were not related to whether products were effective.

The study also collected data in the subsequent 2005-2006 school year, in which teachers who continued with the study had a new cohort of students and a year of experience using the products. Data from this second cohort enable the study to address the question of whether software products are more effective in raising test scores after teachers have a year of experience using them.

The first-year report estimated average effects of groups of products and did not report effects for individual products. School districts and educators purchase individual products, however, and knowing whether individual products are effective is an important ingredient for making decisions supported by evidence. This report presents findings on the effects of 10 products on which data were collected in the first and second years.

The rest of this chapter provides an overview of the study's first-year design and how the design changed for its second year, and notes key aspects of data collection.

### First-Year Study Design

The second year of the study is based on the first year and the main features of the first year study are summarized here. The study was based on voluntary participation of educational software developers, districts and schools, and teachers. The study team worked to ensure that teachers received appropriate training on using products and that schools' technology infrastructures were adequate, and product vendors were responsible for providing the training and technical assistance and for working with schools and teachers on how to use their products.

***Software Products***. Before products could be selected, decisions were made about the study's focus. A design team (ED staff, researchers from MPR, and national experts in evaluation design and educational technology) recommended that the study:

- Focus on software products that support reading and math instruction in low-income schools serving kindergarten to 12th grade

- Use an experimental design to ensure that measured achievement gains could be attributed to the technology products

- Analyze standardized test scores as measures of student academic achievement

A report provides discussions and rationales for the design team's recommendations (Agodini et al. 2003).

The team used a public process to select products for the study. It invited publishers and product developers to submit proposals to include their products in the evaluation. A total of 160 submissions were received in response to a public invitation made in September 2003. Proposals were rated using a two-step process. First, the study team rated the submissions on evidence of effectiveness (based on previous research conducted by the companies or by other parties), whether products could operate on a scale that was suitable for a national study, and whether companies had the capacity to provide training to schools and teachers on the use of their products. Second, a list of candidate products was reviewed by external panels of experts, one for reading and one for math, to arrive at a short list of candidate products. ED selected 16 products for the study from that short list and announced the choices in January 2004. ED also identified four grade levels for the study, deciding to study reading products in first and fourth grades, math products in sixth grade, and algebra I, which typically is taught in ninth grade.

***Recruiting Districts and Schools***. After products were selected, the study team began recruiting school districts to participate. It focused on school districts that had low student achievement and large proportions of students in poverty, but these were general guidelines rather than strict eligibility criteria. The study sought districts and schools that did not already use products like those in the study so that there would be a contrast between the use of software products in treatment and control classrooms. Generally, schools were identified by senior district staff based on broad considerations including whether schools had adequate technology infrastructure and whether schools were participating in other initiatives.

***Participants***. By September 2004, 33 districts and 132 schools had been recruited to participate in the first year of the study. Five districts implemented products in two or more grade levels, and one district implemented products in all four grade levels, resulting in 45 combinations of districts and product implementations. Table I.1 shows the sample sizes by subject and grade level.

Consistent with the recruitment focus on low-income districts and schools, participating sites had a higher percentage of students eligible for free or reduced-price lunch than the average district and school. Free and reduced-price lunch rates were 44 percent for districts using reading products and 57 percent for those using math products, compared to 36 percent nationwide (Table I.2). Study districts also were more likely to be in urban areas (38 percent of districts in the study compared to about 9 percent of districts nationwide) and were larger than the average district along several measures (for example, districts using reading products had an average of about 79 schools and those using math products about 126 schools, compared to about 6 schools in the average district).

**Table I.1. Sample Sizes, First-Year Study**

| Subject and Grade Level | Number of Districts | Number of Schools | Number of Teachers[a] | Number of Students[b] |
|---|---|---|---|---|
| Reading (Grade 1) | 14 | 46 | 169 | 2,619 |
| Reading (Grade 4) | 11 | 43 | 118 | 2,265 |
| Math (Grade 6) | 10 | 28 | 81 | 3,136 |
| Math (Algebra I) | 10 | 23 | 71 | 1,404 |
| Total | 45 | 140 | 439 | 9,424 |
| Unduplicated Total[c] | 33 | 132 | 439 | n.a. |

[a] The number of teachers includes the treatment and control teachers.

[b] The number represents students in the analysis sample who were tested in both fall 2004 and spring 2005. The total number of students who were tested at either point in time is larger because some students tested in the fall moved out of their school district by the time of the spring test and some students tested in the spring had moved into study classrooms after the fall test. The total number of students tested was 10,659 in the fall and 9,792 in the spring.

[c] Because nine districts and eight schools are piloting more than one product for the study, the unduplicated total gives the number of unique districts and schools in the study.

n.a. = not applicable.

Similarly, the particular schools recruited for the study had a higher percentage of students eligible for free or reduced-price lunch and were more likely to be in urban areas (Table I.3). Schools participating in the first and fourth grade studies were elementary schools. Schools participating in the sixth grade study mostly were middle schools (almost 90 percent). Three other schools in the sixth grade study were elementary schools that included sixth grade. In all three schools, sixth grade was the highest grade in the school. Schools participating in the algebra I study were mostly high schools (77 percent) with some middle schools (23 percent) (these percentages are not shown in Table I.3).

***Implementing the Classroom-Level Experimental Design.*** Teachers in schools that volunteered to participate in the study were assigned to either use one of the study's products (the "treatment" group) or not use one of the study's products (the "control" group).[2] Teachers in the treatment group were to implement the designated product as part of their reading or math instruction. Teachers in the control group were to teach reading or math as they would have normally, possibly using software products already available to them, though as the first-year report showed, use of other products in control classrooms was lower than use of the products in treatment classrooms. Because the only difference on average between the treatment and control groups is whether teachers were assigned to use one of the study's products, student test-score differences could be attributed to the technology product provided to treatment teachers, after allowing for sampling variability.

[2] Teachers who volunteered for the study were asked to sign a consent form indicating they understood that they would be part of a research study and would implement their school's product if assigned to the treatment group.

**Table I.2. Characteristics of Districts in the First Year of the Study**

| Characteristics[a] | Average U.S. District | Districts Using Reading Products | Districts Using Math Products |
|---|---|---|---|
| Number of Title I schools[b] | 3.3 | 34.8 | 78.5 |
| District location (percentage) | | | |
| Urban | 8.7 | 38.1 | 37.5 |
| Urban fringe | 24.9 | 52.4 | 43.8 |
| Town | 14.7 | 4.8 | 6.3 |
| Rural area | 51.7 | 4.8 | 12.5 |
| Number of schools per district | 5.9 | 78.6 | 126.4 |
| Number of full-time teachers per district | 170 | 3,642 | 5,828 |
| Number of students per district | 2,988 | 61,660 | 103,426 |
| Percentage of students eligible for free or reduced-price lunch[c] | 36.1 | 44.4 | 56.6 |
| **Number of Districts** | **15,417** | **21** | **16** |

Source: Study tabulations by MPR from the 2001–2002 *Common Core of Data.*

Note: Four districts used both reading and math products.

[a] Data include districts with one or more regular schools (excluding schools focused primarily on special education, vocational/technical or career education, or alternative programs).

[b] Data missing for 6 percent of study districts and 9 percent of districts nationwide.

[c] Data missing for 6 percent of study districts and 10 percent of districts nationwide.

***Data Collection***. The first-year analysis relied on student test scores, data from classroom observations, and teacher questionnaires and interviews. The study team relied on the SAT-10 test for three of the four grade levels:

- First grade reading test: The reading battery of the Stanford Achievement Test (version 9) and the Test of Word Reading Efficiency (TOWRE), a short and reliable one-on-one test of reading ability, for first graders to augment measures of reading skills provided by the SAT-9 (Torgesen et al. 1999)

- Fourth grade reading test: The reading battery of the Stanford Achievement Test (version 10)

- Sixth grade math test: The math battery of the Stanford Achievement Test (version 10)

- Algebra I test: The Educational Testing Service (ETS) End-of-Course Algebra Assessment

**Table I.3. Characteristics of Schools in the First Year of the Study.**

| Characteristics[a] | Average for U.S. | Schools in First Grade Study | Schools in Fourth Grade Study | Schools in Sixth Grade Study | Schools in Algebra I Study |
|---|---|---|---|---|---|
| School location (percentage) | | | | | |
| Urban | 24 | 45 | 52 | 36 | 55 |
| Urban fringe | 32 | 45 | 48 | 43 | 45 |
| Town | 12 | 0 | 0 | 4 | 0 |
| Rural area | 32 | 10 | 0 | 18 | 0 |
| Number of students per teacher | 16 | 16 | 16 | 15 | 15 |
| Number of students per school | 543 | 626 | 572 | 1,073 | 1,352 |
| Percentage of schools receiving Title I funding | 59 | 76 | 88 | 64 | 23 |
| Percentage of students eligible for free or reduced-price lunch | 42 | 49 | 64 | 71 | 54 |
| Student race/ethnicity (percentage) | | | | | |
| White | 64 | 44 | 17 | 21 | 29 |
| Black | 15 | 31 | 57 | 33 | 45 |
| Hispanic | 15 | 22 | 23 | 42 | 19 |
| Asian | 3 | 2 | 3 | 3 | 7 |
| Native American | 3 | <1 | <1 | <1 | <1 |
| **Number of Schools[b]** | **88,542** | **46** | **43** | **28** | **23** |

Source: Study tabulations by MPR from the 2003–2004 *Common Core of Data* (CCD).

[a]Data include regular schools only (excluding schools focused primarily on special education, vocational/technical or career education, or alternative programs).

[b]CCD data are missing for 10 study schools.

Tests were administered in fall 2004 and spring 2005, so gains in achievement made by treatment and control classrooms could be compared.[3]

The study team conducted classroom observations, which were used to assess product implementation. Each classroom was visited three times during the school year, and observers collected data using a protocol that was designed to gather similar information in

[3] To create a baseline measure of algebra skills, the study worked with ETS to separate its algebra assessment, which essentially is a final exam, into two components that had equal levels of difficulty.

both treatment and control classrooms and across the different grade levels and subject areas in the study.

Teacher interviews complemented the observations and provided an opportunity to collect information about implementation issues teachers may have experienced. Background information about teachers was also gathered from a questionnaire that teachers completed in November and December 2004.

### Second-Year Study Design

The second year of the study relied heavily on the context established in the first year of the study highlighted above, but it was modified to stay within resource constraints.

1. ***Fewer products and school districts.*** In the first year, some products were implemented in only a handful of schools, or only had a few schools that would agree to participate in the second year of the study. To achieve a reasonable degree of statistical power for the reporting of individual product effects, the study did not include those products in the second year.

2. ***Less data.*** The study did not conduct classroom observations or teacher interviews, used one reading test rather than two for first graders (dropping the Test of Word Reading Efficiency), did not collect extensive information about students from school records, used district standardized-test scores to the extent possible, and sampled teachers within schools that had more than one treatment or control teacher, reducing the sample to one treatment and one control teacher per school.

***Software Products.*** The second-year study included 10 of the 16 products included in the first-year study. As part of the second-year study, effects are presented separately for each software product. Reporting at the product level means that sample sizes are smaller and statistical power to detect effects is thereby lower than when effects are reported as a group as in the first-year report. Four products that had been implemented in only a few schools in the first year of the study, and for which the expectation was that it was unlikely the study could add schools to reach the target level of statistical power (detecting an effect size of 0.35), had to be dropped from the second-year study. One other product could not be included in the second-year study because a district decided not to participate in the second year and the remaining number of schools was too few to include it. Finally, one developer decided not to participate in the second-year study for one grade level.

The second-year study included four reading software products for first grade, Destination Reading (Riverdeep 2008), the Waterford Early Reading Program (published by Pearson Education 2008), Headsprout (Headsprout 2008), and Plato Focus (Plato 2008). Three of the four products provided supplemental instruction and Plato Focus was used as the core reading curriculum. The second-year study also included two reading products for fourth grade, LeapTrack (LeapFrog Schoolhouse 2008) and Academy of Reading (Autoskill International 2008). These products supplemented the core reading curriculum with tutorials, practice, and assessment geared to specific reading skills.

For math, the second-year study included two math products for sixth grade, Larson Pre-Algebra (Houghton-Mifflin 2008) and Achieve Now (Plato 2008). The products supplemented the core math curriculum with provided tutorial and practice opportunities and assessed student skills. The study included two products for algebra I, Cognitive Tutor Algebra I (Carnegie Learning 2008) and Larson Algebra I (Houghton-Mifflin 2008). The Larson product supplemented algebra I instruction and the Cognitive Tutor product was the core algebra I curriculum. Students at a variety of high school grade levels can take algebra I, and many middle schools also teach algebra I. In the study, 9 percent of algebra students were in eighth grade, 87 percent were in ninth grade, and 4 percent were in grades 10, 11, or 12. For the remainder of the report, algebra I will be referred to as "algebra."

The reading and math products supplemented the core curriculum or, as was the case for Cognitive Tutor, were the core curriculum. Products generally were not implemented only to remediate skills for those students who were lagging.

***Participating in the Second-Year Study***. As in the first year, participation in the second year of the study was voluntary. As noted above, some districts were not included because they had implemented products that were not used in enough schools to attain a reasonable degree of statistical power to report the effects of the individual product.[4] The study team also added two districts and added schools in one district for products that agreed to remain in the study but were at the margin of adequate statistical power.

The second-year study included 23 districts and 77 schools that participated in the study during the second year (2005-2006 school year). Two districts and 18 schools were new to the study and the rest had participated in the previous year of the study. Table I.4 shows the sample sizes by product and grade level for the second year of the study. More information on teacher sample sizes and year of participation appears in Appendix A, Table A.1.

Consistent with the first-year study, Tables I.5 and I.6 show that districts and schools participating in the second year had a higher percentage of students eligible for free or reduced-price lunch than the average district and school. Free and reduced-price lunch rates were 39 percent for districts using reading products and 59.7 percent for those using math products, compared to 36.1 percent nationwide. As in the first year, districts in the second-year study were more likely to be in urban areas and were larger than the average district along several measures (for example, districts using reading products had an average of about 51 schools and those using math products about 166 schools, compared to about 6 schools in the average district). Similarly, the particular schools that participated in the second-year study had a higher percentage of students eligible for free or reduced-price lunch and were more likely to be in urban areas (Table I.6).

[4] The first-year study had a target effect size of 0.25. Because individual product effects are based on smaller sample sizes, the study used a target effect size of 0.35 for individual products and calculated the number of additional schools that would be needed to reach that target effect size, which translated to about 24 teachers in the treatment and control groups. If the number of additional teachers needed was deemed too large to be feasible, the product was not included in the second year. If the number was feasible to reach by adding more schools within the study's time frame, the product was included in the second year conditional on the success of recruiting efforts to identify more schools. The study was able to recruit additional schools for two of three products.

**Table I.4. Sample Sizes, Second-Year Study**

| | Number of Districts | Number of Schools | Number of Teachers[a] | Number of Students[b] |
|---|---|---|---|---|
| First Grade | | | | |
| Destination Reading | 2 | 9 | 25 | 453 |
| Headsprout | 3 | 7 | 18 | 268 |
| Plato Focus | 3 | 8 | 18 | 319 |
| Waterford Early Reading Program | 3 | 9 | 20 | 331 |
| Total | 11 | 33 | 81 | 1,371 |
| Fourth Grade | | | | |
| Academy of Reading | 2 | 7 | 14 | 282 |
| LeapTrack | 2 | 4 | 8 | 181 |
| Total | 4 | 11 | 22 | 463 |
| Sixth Grade | | | | |
| Achieve Now | 3 | 8 | 20 | 313 |
| Larson Pre-Algebra | 3 | 8 | 18 | 386 |
| Total | 6 | 16 | 38 | 699 |
| Algebra I | | | | |
| Cognitive Tutor | 4 | 9 | 18 | 276 |
| Larson Algebra I | 3 | 8 | 17 | 471 |
| Total | 7 | 17 | 35 | 747 |
| **Total** | **28** | **77** | **176** | **3,280** |
| **Unduplicated Total** [c] | **23** | **77** | n.a. | n.a. |

[a] The number of teachers includes the treatment and control teachers.

[b] The number represents students in the analysis sample who had fall 2004 and spring 2005 test scores or for which fall score was imputed. The total number of students who were tested or for whom the district provided test scores at either point in time is larger because some students tested in the fall moved out of their school district by the time of the spring test and some students tested in the spring had moved into study classrooms after the fall test. Table A.3 in the Appendix presents the number of students that were eligible for participating in the second-year study.

[c] Because three districts are using more than one product for the study, the unduplicated total gives the number of unique districts in the study.

n.a. = not applicable.

**Table I.5. Characteristics of Districts in the Second Year of the Study**

| Characteristics[a] | Average U.S. District | Districts Using Reading Products | Districts Using Math Products |
|---|---|---|---|
| Number of Title I schools[b] | 3.4 | 19.9 | 117.1 |
| District location (percentage) | | | |
| Urban | 8.7 | 26.7 | 23.1 |
| Urban fringe | 25.0 | 60.0 | 69.2 |
| Town | 14.7 | 6.7 | 7.7 |
| Rural area | 51.6 | 6.6 | 0.0 |
| Number of schools per district | 6.1 | 51.1 | 165.6 |
| Number of full-time teachers per district | 180 | 2,337 | 9,121 |
| Number of students per district | 3,159 | 36,820 | 134,017 |
| Percentage of students eligible for free or reduced-price lunch[c] | 36.1 | 39.0 | 59.7 |
| **Number of Districts** | **15,450** | **15** | **13** |

Source: Study tabulations by MPR from the 2001–2002 *Common Core of Data.*

Note: Four districts used both reading and math products.

[a] Data include districts with one or more regular schools (excluding schools focused primarily on special education, vocational/technical or career education, or alternative programs).

[b] Data missing for 6 percent of study districts and 9 percent of districts nationwide.

[c] Data missing for 6 percent of study districts and 10 percent of districts nationwide.

**Table I.6. Characteristics of Schools in the Second Year of the Study**

| Characteristics[a] | Average U.S. School | Schools in First Grade Study | Schools in Fourth Grade Study | Schools in Sixth Grade Study | Schools in Algebra I Study |
|---|---|---|---|---|---|
| School location (percentage) | | | | | |
| Urban | 25 | 61 | 27 | 25 | 53 |
| Urban fringe | 33 | 33 | 55 | 62 | 47 |
| Town | 10 | 0 | 0 | 0 | 0 |
| Rural area | 32 | 6 | 18 | 13 | 0 |
| Students per teacher | 15 | 16 | 12 | 14 | 10 |
| Number of students per school | 535 | 536 | 582 | 1,158 | 1,325 |
| Percentage of schools receiving Title I funding | 60 | 76 | 82 | 63 | 24 |
| Percentage of students eligible for free or reduced-price lunch | 44 | 55 | 64 | 75 | 53 |
| Student race/ethnicity (percentage) | | | | | |
| White | 62 | 49 | 41 | 9 | 28 |
| Black | 16 | 24 | 28 | 35 | 53 |
| Hispanic | 16 | 23 | 26 | 52 | 15 |
| Asian | 4 | 3 | 5 | 4 | 4 |
| Native American | 2 | 1 | <1 | <1 | <1 |
| **Number of Schools[b]** | **90,020** | **33** | **11** | **16** | **17** |

Source: Study tabulations by MPR from the 2003–2004 *Common Core of Data* (CCD).

[a] Data include regular schools only (excluding schools focused primarily on special education, vocational/technical or career education, or alternative programs).

[b] CCD data are missing for 10 study schools.

***Creating Treatment and Control Groups of Teachers.*** For the second year of the study, teachers, schools, and districts volunteered to remain in the study for a second year. Teachers who had been randomly assigned to the treatment group in the first year retained their status. Teachers assigned to the control group in the first year also retained their status, unless there were no treatment teachers remaining in the study in their school, in which case they were randomly assigned to treatment or control groups. As in the first year, random assignment was done within schools. Teachers who were new to the study were randomly assigned to treatment or control status with the same probability of assignment used in the first year.

As in the first year, teachers in the treatment group were to implement the designated product as part of their reading or math instruction. Teachers in the control group were to teach reading or math as they would have normally, possibly using technology products already available to them, though as the first-year report showed, use of other products in control classrooms was lower than use of the products in treatment classrooms. For example, the first report indicated that first grade treatment teachers used reading software products for 52 hours during the year and first grade control teachers reported using reading software products for 10 hours ($p < .01$). Differences in product use were similar in other grade levels (Dynarski et al. 2007).

As mentioned above, the construction of the second-year sample was designed to accomplish two objectives: to examine whether a year of teacher experience using products increased effectiveness, and to examine product effects separately. Adding schools and teachers to the sample improved the study's ability to detect individual product effects, but the teachers who were added were not part of the study of experience because they were first-time users of the products, similar to teachers in the study's first year. As a result, teacher and student sample sizes differ for the two parts of the analysis. Appendix A discusses in more detail the second-year sample and the subsample that participated in the two years of the study, which was used to study teacher experience effects. Appendix B discusses the sample used to study individual product effects, which included both the first- and second-year study samples in order to increase the statistical power of the study.

***Data Collection.*** Data collection procedures differ from those used in the first year. For teachers, classroom observations and teacher interviews were not collected in the second year. Background information about teachers new to the study was gathered from a questionnaire that teachers completed in November and December 2005. For schools, *Common Core* data were matched to schools that had entered in the second year of the study. For most districts, the second-year analysis relied on student test scores from tests administered by the study. Some districts already administered a nationally normed test as part of their regular district testing program; in these cases, the study used those scores. The study team administered the following student tests:

- First grade reading tests: The reading battery of the Stanford Achievement Test (version 9).[5]
- Fourth grade reading tests: The reading battery of the Stanford Achievement Test (version 10).[6]
- Sixth grade math test: The math battery of the Stanford Achievement Test (version 10).[7]
- Algebra I test: The Educational Testing Service (ETS) End-of-Course Algebra Assessment.[8]

[5] For fall of first grade, the test is known as the Stanford Early School Achievement Test (SESAT). See Pearson Education (1996a,).

[6] See Pearson Education (2003a).

[7] See Pearson Education (2003b).

To reduce costs, tests were administered to a subsample of classrooms. In each school, one treatment classroom and one control classroom were randomly sampled if more than one teacher was in either group. Appendix A provides details about the sample of teachers and students that were tested and their response rates (see Tables A.1 and A.4).

Test scores collected from districts that already administered a nationally normed test varied as follows:

- First grade using reading products. One district provided fall test scores on the reading section of the Iowa Tests of Basic Skills (ITBS).[9] One district provided spring test scores on the reading section of the Stanford Achievement Test (version 10).
- Fourth grade using reading products. Two districts provided fall test scores, one on the reading section of the Iowa Tests of Basic Skills and one on the California Achievement Test, Sixth Edition.[10]
- Sixth grade using math products. One district provided fall test scores from the math section of the ITBS. One district provided fall and spring test scores on the math section of the New Mexico Standards Based Assessment.[11]
- Algebra I students. One district provided fall test scores on the math section of the Iowa Tests of Basic Skills (ITBS).

The study used district test scores only if the tests were commercially available tests with national norms. District scores were collected for 484 students (the study tested 1,760 students). Table I.7 summarizes characteristics of the tests used in the second year. The study converted test scores in first, fourth, and sixth grades to normal curve equivalent (NCE) units to standardize the measures across tests and cohorts.[12] Algebra I scores for the ETS test are reported as percent correct. As noted before, the use of various tests in place of the study's tests required the estimation models to incorporate indicator variables as to which test students had taken. Using district tests in this way is unlikely to affect the findings unless the tests have differing levels of sensitivity to the effects of software compared to the SAT-10.

[8] See Educational Testing Service (1997).

[9] See Riverside (2001).

[10] See CTB/McGraw-Hill (2001a).

[11] See Pearson Education (2006).

[12] A normal curve equivalent score converts the scaled test score into the range 1 to 99, with 50 being the average for the nationally normed sample. Unlike percentiles, NCE scores can be averaged, which makes them more appropriate for statistical analyses and estimation of product effects.

The last source of data comes from product records which were used to provide information about product use. Eight of the 10 products included in the study used databases to track the time when each student was logged on. The usage measure reported in the study is actual student logged-on time for a school year, as reported by the product database. Usage by more than one student at a time, such as in a group activity, is counted only for the logged-on student. Time spent doing activities that are related to product use but occur when students are not logged on, such as reading materials related to a computer lesson, are not counted as usage.[13]

[13] The first-year study analyzed usage from product records and also usage as reported by teachers in interviews. However, teachers were not interviewed in the second year.

**Table I.7. Features of Tests Used in this Study**

| Test | Grade and Subject | General Information | Norm Sample | Reliability and Validity |
|---|---|---|---|---|
| Stanford Early School Achievement Test (SESAT 2, Form S) | First Grade-Reading (fall: 10 districts) | Test used to measure what students learn in their first years of schooling. | The norm sample included more than 700 subjects at each level, but demographic information is not reported. | The only validity measures available are between the SESAT and the Otis-Lennon School Ability Test (OLSAT). Within the third edition, the correlation was .81, with an N of 5,967, with the reading composite correlates .62 and .61 with OLSAT verbal and non-verbal components. Internal consistency reliability coefficient for the Level 2 exam is .96. |
| Stanford Achievement Test, Ninth Edition (SAT-9) | First Grade-Reading (spring: 10 districts) | Commercially available, used by a large number of states and school districts. | National norms, based on samples of 250,000 students in spring 1995 and 200,000 in fall 1995.<br>The average student in the norm sample has a normal curve equivalent score of 50, and the standard deviation of normal curve equivalent score is 21.06. | Internal consistency (KR-20) coefficients ranged from the .80s to the .90s for full multiple-choice battery test and subtests. Evidence of content, criterion-related, and construct validity. |
| Iowa Tests of Basic Skills (ITBS) | First Grade-Reading (fall: 1 district)<br>Fourth Grade-Reading (fall: 1 district)<br>Sixth Grade-Math (fall: 1 district)<br>Algebra I (fall: 1 district) | Group-administered, norm-referenced battery of achievement tests for students in kindergarten through eighth grade. The tests are ordered by levels ranging from Level 5 to Level 14. The levels correspond to target ages and grade levels based on academic achievement. | The national standardization was based on the 2000 spring and fall administrations of the tests. The total unweighted sample was approx. 170,000 students for the spring 2000 sample administration and approximately 76,000 for the fall 2000 sample. | Internal consistency (KR-20) and equivalent forms reliabilities are in the expected range with most reliability coefficients ranging from the middle .80s to low .90s.<br>ITBS uses other measures of validity as well as summaries of conventional item analyses and reliability coefficients. |
| Stanford Achievement Test, Tenth Edition (SAT-10) | Fourth Grade-Reading (fall: 2 districts; spring: 4 districts)<br>Sixth Grade-Math (fall and spring: 5 districts) | Commercially available, used by a large number of states and school districts. | National norms, based on samples of 250,000 students in spring 2002 and samples of 100,000 in fall 2003.<br>The average student in the norm sample has a normal curve equivalent score of 50, and the standard deviation of normal curve equivalent score is 21.06. | Internal consistency (KR-20) coefficients are .80 and .90 for full multiple-choice battery test and subtests. Evidence of content, criterion-related, and construct validity. |

Table I.7 *(continued)*

| Test | Grade and Subject | General Information | Norm Sample | Reliability and Validity |
|---|---|---|---|---|
| California Achievement Test, Sixth Edition (CAT/6) | Fourth Grade-Reading (fall: 1 district) | The California Achievement Test (CAT), also called TerraNova, is a norm- and criterion-referenced achievement test for students in kindergarten through 12th grade. | Norm data were collected in the 1999-2000 academic year. The norming sample consisted of 280,000 students from 429 schools in the fall, 202 schools in the winter, and 689 schools in the spring. Individual schools were the sampling unit. | All internal consistency estimates for subareas are in the mid- to low .90s. |
| New Mexico Standards Based Assessment (NMSBA) | Sixth Grade-Math (fall and spring: 1 district) | Criterion-referenced test. | The sample consisted of all New Mexico (NM) students in grades 3-9 and 11 who were administered the NMSBA—virtually the entire population of students in grades 3-9 and 11 in NM, which is the target population for this assessment. | Inter-rater reliability for items on the English version of the sixth grade math exam was 84.81% with a standard deviation of 7.52. Items were from CTB items pools, and were aligned to the New Mexico K-12 content standards, benchmarks, and performance standards. |
| Educational Testing Service End-of-Course Algebra Test (ETS) | Algebra 1 (fall: 5 districts; spring: 6 districts) | Full form is commercially available. Test is based on algebra I standards of the National Council of Teachers of Mathematics. For the study, ETS separated the test items into two balanced halves with equal levels of difficulty, such that one could be administered in the fall and the other in the spring. | Not nationally normed. | In 2003, information from 20,506 test takers indicated a mean score of 23.3 questions correct out of 50, with reliability of .87 and a standard error measurement of 3.1. The two halves of the test used in the study had similar reliability characteristics. |

# Chapter II

# Effects in the First and Second Years of the Study

The question in this part of the study is whether student scores on standardized reading or math tests are related to teachers' experience using the software products for a second year. To address this question, the study restricted the sample of teachers to those who had continued to teach in their school and grade level for another year. For each teacher, we included students who were in the teacher's classroom either in the first year or in the second year of the study.

The analysis does not answer the question of how the results from the study's first year differ with a year of experience. It only can answer the question of how the results from the study's first year differ with a year of experience *for teachers who participated in the second year.* The reduction in the number of products from the first year to the second year represents a key difference in the two years of the study that affects how results are interpreted.

The study uses an experimental design in the sense that teachers were assigned randomly at the beginning of the first year to use or not use a product. However, if teacher decisions to exit grade levels, their school, or teaching *per se,* are related to the use or non-use of products in the classrooms, the integrity of the experimental design is reduced. Tables presented below (Tables II.1, II.5, II.9, and II.13) indicate that teachers in the treatment and control groups who continued in the second year generally had similar characteristics, but whether unobserved characteristics are similar cannot be known.

### Approach for Estimating Experience Effects

The study calculated the "effect" of teachers' experience using the software products as the difference between the second-year product effect on test scores and the first-year product effect on test scores. A "product effect" in this context is the difference in spring student test scores between treatment and control classrooms caused by the assignment of treatment classrooms to use a software product.

Product effects were estimated using multilevel models with two levels. The outcome at the first level is the spring student test score, which is modeled as a function of a student's fall test score, age, and gender, and a student random effect. The outcome at the second level is the classroom-average test score, which is modeled as a function of teacher years of experience and education level. The second level also includes a variable indicating the teacher's school and a random effect for each teacher. The models use multiple imputation methods to impute missing data, with all models estimated on five data sets and the estimates averaged. Variances are adjusted for the multiple imputation. The HLM 6.02 package was used for estimation.

The estimated models pool products together to estimate effects (a variable is entered that indicates whether teachers are assigned to use a product, regardless of which product). This approach puts greater weight on products that represent a larger proportion of the sample.

Within the statistical model, the experience effect is estimated using a "treatment by year" interaction variable that allows the first-year product effect to be shifted in the second year. This approach allows conventional tests of significance to be used to determine if the second-year product effect differs significantly from the first year product effect, by testing whether the estimated coefficient of the "treatment by year" interaction variable is statistically different from zero. Conceptually, estimating experience effects using this approach is equivalent to subtracting the difference in adjusted treatment and control group average spring test scores in the second year of the study (the second-year product effect) from the difference in adjusted treatment and control group average spring test scores in the second year of the study (the first-year product effect).

Appendix C presents details on the models.

## A. Findings for First Grade Reading Products

This section presents findings for first-grade reading software products. The analysis sample included 11 districts, 22 schools, and 43 teachers who participated in both the first and second years of the study, and 1,411 students (716 from year 1 and 695 from year 2). Teachers used one of four reading products: Destination Reading (Riverdeep 2008), Waterford Early Reading (Pearson Education 2008), Headsprout (Headsprout 2008), and Plato Focus (Plato Learning Corporation 2008). Participating schools had 48 percent of students receiving free or reduced-price lunch, 20 percent Hispanic students, 25 percent black students, and an average student/teacher ratio of 16. Forty-five percent of schools were in urban areas.

### Characteristics of Teachers and Students in Treatment and Control Classrooms

Teachers were randomly assigned to treatment or control groups in the first year. Since the sample used to study the effect of teachers' experience using products on student test scores included only teachers who participated in the study for both years of the study, attrition and mobility created the possibility of differences between teachers in the treatment and control groups, but *p*-values of tests of equivalence indicated no statistically significant differences (see Table II.1). The study did not randomly assign students to teachers. Rather, schools used their conventional approaches to allocate students to teachers. Table II.1 shows that students with the treatment and control teachers were similar on fall test scores, age, and gender. None of the differences is statistically significant. Student and teacher characteristics are entered into models to adjust for remaining differences and to increase the statistical power of the estimation.

### Teacher Training and Support During the First Year

Product vendors trained teachers who would be implementing the products on how to use them. Training generally took place in the host districts (and sometimes the host schools) during summer or early fall of 2004. Training topics included classroom management, curriculum, and standards alignment, and generally teachers had opportunities to practice using the products. Nearly all teachers (94 percent) attended the initial training, according to attendance logs. On average, vendors provided 8 hours of training, varying from 2 to 18 hours across products. Vendors also provided support during the school year. Modes for ongoing support included e-mail or telephone help desks (69 percent of teachers reported receiving this kind of help), product representatives visiting teachers (55 percent), and additional training at schools (39 percent). The study team also worked with districts to identify hardware and software needs including computers, headphones, memory, and operating system upgrades, and the study purchased the upgrades as needed. Common upgrades included desktop and laptop computers, servers, memory, and headphones. The study did not provide software or hardware support for control group teachers.

**Table II.1 Characteristics of Teachers and Students in Treatment and Control Classrooms, First Grade**

| | Year 1 | | | Year 2 | | | Years 1 and 2 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Treatment Classrooms | Control Classrooms | *p*-value of the Difference | Treatment Classrooms | Control Classrooms | *p*-value of the Difference | Treatment Classrooms | Control Classrooms | *p*-value of the Difference |
| **Teacher Characteristics** | | | | | | | | | |
| Teaching experience (years) | | | | | | | 13.17 | 11.47 | 0.58 |
| | | | | | | | (10.28) | (9.00) | |
| Has a master's degree (percentage) | | | | | | | 29.17 | 47.37 | 0.23 |
| | | | | | | | (46.43) | (51.30) | |
| Female (percentage) | | | | | | | 100.00 | 100.00 | . |
| | | | | | | | (0.00) | (0.00) | |
| **Teacher Sample Size** | | | | | | | **24** | **19** | |
| **Student Characteristics** | | | | | | | | | |
| Female (percentage) | 52.55 | 51.34 | 0.75 | 45.90 | 47.87 | 0.61 | 49.31 | 49.61 | 0.91 |
| | (49.76) | (49.93) | | (49.90) | (50.04) | | (49.91) | (49.97) | |
| Age (years) | 6.61 | 6.62 | 0.71 | 6.64 | 6.63 | 0.76 | 6.62 | 6.62 | 0.92 |
| | (0.38) | (0.38) | | (0.38) | (0.39) | | (0.38) | (0.39) | |
| Unadjusted score on fall reading test (NCE) | 49.45 | 49.72 | 0.93 | 50.28 | 51.27 | 0.65 | 49.86 | 50.50 | 0.75 |
| | (20.52) | (20.15) | | (21.29) | (19.27) | | (20.89) | (19.71) | |
| Unadjusted score on spring reading test (NCE) | 49.91 | 48.54 | 0.84 | 51.51 | 53.43 | 0.36 | 50.69 | 50.98 | 0.75 |
| | (20.01) | (20.26) | | (17.10) | (15.77) | | (18.66) | (18.30) | |
| **Student Sample Size** | **411** | **305** | | **390** | **305** | | **801** | **610** | |

Note: Standard deviations in parentheses.

Tests of treatment and control differences were conducted using a two-level hierarchical model with classroom treatment status as a fixed effect and teacher and student random effects. The *p*-value of the difference shown in the table is the *p*-value of the estimated treatment coefficient.

## Usage of First Grade Reading Products Was Lower in the Second Year

Three of the four products used in first grade included databases that tracked the time when the students were logged on. The variable "product usage in minutes" is the number of minutes that students were logged on to the product. The variable "product usage in weeks" is the number of weeks in which students used the product for at least some time.

Table II.2 presents means and standard deviations of these two variables for the subsample of students for which these data were available.[14] The average student used products for 2,556 minutes in the first year of the study and 1,180 minutes in the second year of the study (the difference in usage is statistically significant, $p < 0.01$). Using a 40-week school year as a basis, minutes of usage averaged 30 minutes per week in the second year. Products were not used in all weeks, however. Product records indicate that the average student used a product 29.6 weeks in the first year and 24.8 weeks in the second year. Using this number of actual weeks of usage as a basis, average minutes of usage was 48 minutes a week during the second year.

**Table II.2. Student Product Usage in the First and Second Year, First Grade**

| | Year 1 | Year 2 | Difference |
|---|---|---|---|
| Average minutes of student product usage | 2,556<br>(1,738) | 1,180<br>(1,213) | -1,376 |
| Average number of weeks in which students used products | 29.6<br>(7.8) | 24.8<br>(9.8) | -4.8 |
| Student sample sizes | 347 | 197 | |

Note: Standard deviations in parentheses. Source: Product records.

### Effects of First Grade Reading Products in the Second Year Are Not Statistically Different from Effects in the First Year

Table II.3 presents the product effects for the two study years and the difference of effects between the two years, which is interpreted as the effect of teachers' experience using products on test scores. The effects in the table are estimates from the two-level models (Appendix C presents estimates from the full model).

Table II.3 shows that, in the first year, the effect of software products on student reading test scores was 0.86 NCE units, corresponding to the average student going from the 50th percentile to the 52nd percentile. In the second year, the effect of software products on reading test scores was –1.28 NCE units, corresponding to a student going from the 50th percentile to the 48th percentile. Neither effect is statistically significant.

The main question in this part of the study was whether a second year of teaching experience using software products increased reading test scores. An increase in product effects from a year of experience would yield a positive difference between the second-year effect and the first-year effect. Table II.3 shows that the effect of an additional year of

[14] Data are available for 84 percent of treatment students in the first year and 51 percent of treatment students in the second year. At the teacher level, these correspond to 88 percent of treatment teachers in the first year and 58 percent of treatment teachers in the second year.

teaching experience using the product was –2.14 NCE units. This difference is not statistically different from zero ($p > 0.05$).[15]

**Table II.3. Product Effects on Spring Reading Test Scores, First Grade**

| | Treatment Group | Control Group | Product Effect | Effect Size | *p*-value |
|---|---|---|---|---|---|
| | | | | | |
| Average score on spring reading test, first year (NCE) | 49.40 | 48.54 | 0.86 | 0.04 | >0.50 |
| Average score on spring reading test, second year (NCE) | 52.15 | 53.43 | -1.28 | -0.06 | >0.50 |
| Difference between the first and second year | | | -2.14 | | 0.08 |
| Student sample sizes (first year plus second year) | 801 | 610 | | | |

Note: Details of the estimation model are presented in Appendix C and Table C.1. Variables in the model include student age, gender, whether or not the student was in the second year, fall scores, and an interaction if the student took a district test; and teachers' experience, whether teachers had a master's degree, whether teachers were assigned to the treatment group, and a variable indicating each school. In addition, the model includes student and teacher random effects.

The average score reported in the table for the treatment group is the unadjusted average score for the control group plus the product effect estimated from the model. It differs from the unadjusted average score for the treatment group.

Effect sizes are calculated by dividing the score difference (product effect) shown in the table by 21.06, which, by design, is the standard deviation of a national norm sample of NCE scores.

The study also examined whether products reduced the proportion of students who scored below the 33rd percentile on the reading test (based on the test's national norms).[16] The statistical model used for the analysis was a two-level generalized linear model for binary outcomes, where the indicator variable of whether the student's spring score was below the 33rd percentile was used as the outcome. A similar indicator variable for the student's fall score was used as a covariate along with the same set of student and teacher characteristics used in the above model.

[15] As noted in the methods discussion at the beginning of the chapter, the statistical test of the equality of effects in the two years is equivalent to the statistical test that the coefficient of the "year by treatment" interaction variable (the amount by which the effect in the second year differs from the effect in the first year) is statistically significant. Appendix C discusses the test.

[16] The percentiles referred to here are based on national norms and thus, by definition, 33 percent of students nationally fall below the 33rd percentile. In any individual study sample, more or less than 33 percent of the students may fall below the national 33rd percentile, reflecting the degree to which the study sample is generally lower- or higher-scoring than the national norming sample.

Table II.4 shows percentages of treatment and control group students scoring below the 33rd percentile on the spring reading test in the two years. Scores of both treatment and control students were higher in the second year than in the first year (fewer students in the second year scored below the 33rd percentile on the reading test). However, the difference between the percentage of treatment students scoring below the 33rd percentile and the percentage of control students scoring below the 33rd percentile was statistically significantly larger in the second year than in the first (12.5 percentage points versus 1.5 percentage points).

**Table II.4. Effect on Percentage of Students in Lowest Third of Reading Test Score, First Grade**

| | Treatment Percentage | Control Percentage | Product Effect | Effect Size | *p*-value |
|---|---|---|---|---|---|
| Percentage of students below 33rd percentile of spring reading test in first year | 37.6 | 36.1 | 1.5 | 0.04 | >0.50 |
| Percentage of students below 33rd percentile of spring reading test in second year | 33.8 | 21.3 | 12.5 | 0.39 | 0.08 |
| Difference between first and second year | | | 11.0 | | 0.03 |

Note: Other variables in the model include student age, gender, indicator for the second year, the pretest indicator that a student was in the lowest third of the standardized test, and an interaction if the student took a district test; teacher's experience, whether he or she had a master's degree, an interaction of the second year and treatment, and a variable indicating each school. The model includes student and teacher random effects.

The treatment percentage reported in the table is the unadjusted control percentage plus the product effect. It differs from the unadjusted treatment percentage. The effect size is calculated using the Cox Index (the log odds ratio divided by 1.65).

Additional analyses investigated the relationship between products effects on student test scores and usage. However, interactions between usage and product effects were not statistically significant in either year.[17]

## B. Findings for Fourth Grade Reading Products

This section presents findings for fourth grade reading software products. The analysis sample included 3 districts, 7 schools, 13 teachers who participated in both the first and second years of the study and 604 students (317 from year 1 and 287 from year 2). Teachers used two reading products: Academy of Reading (published by Autoskill), and LeapTrack (published by LeapFrog Schoolhouse). Participating schools had 60 percent of students

[17] To assess this relationship, we first created variables at the teacher level for their students' average usage in the two study years. We then used the model to estimate product effects described in Appendix C and included usage variables as covariates in the second-level equations. The two estimated coefficients for usage were statistically insignificant (*p*-values are 0.76 for the coefficient on first-year usage and 0.41 for second-year usage).

receiving free or reduced-price lunch, 20 percent Hispanic students, 28 percent black students, and a student/teacher ratio of 18. Forty-three percent of schools were located in urban areas.

### Characteristics of Teachers and Students in Treatment and Control Classrooms

Teachers were randomly assigned to treatment or control groups in the first year. Since the sample used to study the effect of teachers' experience using software products on student test scores included only teachers who participated in the study for both years of the study, attrition and mobility created the possibility of differences between teachers in the treatment and control groups, but tests of equivalence indicated no statistically significant differences (see Table II.5). The study did not randomly assign students to teachers. Schools used conventional approaches they normally used to allocate incoming students to teachers. Table II.5 shows that students with the treatment and control teachers were similar on fall test scores, age, and gender. None of the differences is statistically significant. Student and teacher characteristics are entered into models to adjust for remaining differences and increase the statistical power of the analysis.

### Teacher Training and Support During the First Year

Product vendors trained teachers to use products during summer and early fall of 2004. Trainings typically were in the host district and sometimes in the host schools. The initial training averaged 7 hours, varying from 2 hours to 17 hours depending on the product. Topics included classroom management and alignment with standards and with the local curriculum; the trainings also gave teachers the opportunity to practice using the products. Nearly all teachers (94 percent) attended the initial training, according to attendance logs. In addition, teachers received other forms of support after initial training. Product representatives visited teachers (84 percent of teachers reported being visited by a representative), teachers received support through e-mail or telephone help desks (41 percent of teachers), and additional training was provided at schools (59 percent of teachers).The study team also worked with districts to identify hardware and software needs including computers, headphones, memory, and operating system upgrades.

### Usage of Fourth Grade Reading Products Was Higher in the Second Year

The two products used in fourth grade included databases that tracked the time when the students were logged on. Table II.6 presents means and standard deviations of the two variables for the subsample of students for which these data were available.[18] The average student used products for 721 minutes in the first year of the study and 933 minutes in the second year (the difference in usage is statistically significant, $p < 0.01$). The average student also used products in 13.2 weeks of the first year and 15.5 weeks of the second year. Combined, minutes of usage in the second year averaged 60 minutes a week when products were being used.

[18] Data are available for all treatment students in the first year and 82 percent of treatment students in the second year. At the teacher level, these correspond to all treatment teachers in the first year and 86 percent of treatment teachers in the second.

Table II.5 Characteristics of Teachers and Students in Treatment and Control Classrooms, Fourth Grade

| | First Year | | | Second Year | | | First and Second Years | | |
|---|---|---|---|---|---|---|---|---|---|
| | Treatment Classrooms | Control Classrooms | *p*-value of the Difference | Treatment Classrooms | Control Classrooms | *p*-value of the Difference | Treatment Classrooms | Control Classrooms | *p*-value of the Difference |
| **Teacher Characteristics** | | | | | | | | | |
| Teaching experience (years) | | | | | | | 15.71 | 12.20 | 0.50 |
| | | | | | | | (7.65) | (9.01) | |
| Has a master's degree (percentage) | | | | | | | 28.57 | 50.00 | 0.44 |
| | | | | | | | (48.80) | (54.77) | |
| Female (percentage) | | | | | | | 71.43 | 83.33 | 0.62 |
| | | | | | | | (48.80) | (40.82) | |
| **Teacher Sample Size** | | | | | | | **7** | **6** | |
| **Student Characteristics** | | | | | | | | | |
| Female (percentage) | 48.24 | 50.00 | 0.76 | 48.41 | 47.69 | 0.90 | 48.32 | 48.94 | 0.88 |
| | (49.98) | (50.17) | | (50.13) | (50.14) | | (49.97) | (50.08) | |
| Age (years) | 9.62 | 9.55 | 0.46 | 9.71 | 9.65 | 0.56 | 9.66 | 9.60 | 0.33 |
| | (0.45) | (0.46) | | (0.48) | (0.43) | | (0.47) | (0.45) | |
| Unadjusted score on fall reading test (NCE) | 49.04 | 52.04 | 0.43 | 48.28 | 51.09 | 0.53 | 48.67 | 51.60 | 0.47 |
| | (19.55) | (17.47) | | (19.99) | (17.34) | | (19.74) | (17.39) | |
| Unadjusted score on spring reading test (NCE) | 52.39 | 52.90 | 0.78 | 52.88 | 51.99 | 0.95 | 52.63 | 52.48 | 0.93 |
| | (21.24) | (19.81) | | (21.36) | (20.68) | | (21.27) | (20.18) | |
| **Student Sample Size** | **165** | **152** | | **157** | **130** | | **322** | **282** | |

Note: Standard deviations are in parentheses. Tests of treatment and control differences were conducted using a two-level hierarchical model with classroom treatment status as a fixed effect and teacher and student random effects. The *p*-value of the difference shown in the table is the *p*-value of the estimated treatment coefficient.

Table II.6. Student Product Usage in the First and Second Year, Fourth Grade

| | Year 1 | Year 2 | Difference |
|---|---|---|---|
| Average minutes of student product usage | 721<br>(320) | 933<br>(376) | 212 |
| Average number of weeks in which students used products | 13.2<br>(4.7) | 15.5<br>(3.8) | 2.3 |
| Student sample sizes | 165 | 128 | |

Note: Standard deviations in parentheses. Source: Product records.

## Effects of Fourth Grade Reading Products in the Second Year Are Not Statistically Different from the First Year

Table II.7 presents the product effect for the two study years and the difference of effects between the two years, which is interpreted as the effect of teachers' experience using products on test scores. The effects in the table are estimates from the two-level models (Appendix C presents estimates from the full model).

Table II.7 shows that, in the first year, the effect of software products on student reading test scores was 2.65 NCE units, corresponding to the average student going from the 50th percentile to the 55th percentile. The effect is statistically insignificant. In the second year, the effect of software products on reading test scores was 4.67 NCE units, corresponding to a student going from the 50th percentile to the 58th percentile. This effect is statistically significant.

**Table II.7. Product Effects on Reading Test Scores, Fourth Grade**

| | Treatment Group | Control Group | Product Effect | Effect Size | *p*-value |
|---|---|---|---|---|---|
| Average score on spring reading test, first year (NCE) | 55.55 | 52.90 | 2.65 | 0.13 | 0.18 |
| Average score on spring reading test, second year (NCE) | 56.66 | 51.99 | 4.67 | 0.22 | 0.01 |
| Difference in effects between first and second year | | | 2.02 | | 0.29 |
| Student sample sizes (first year plus second year) | 322 | 282 | | | |

Note: Details of the estimation model are presented in Appendix C and Table C.1. Variables in the model include student age, gender, whether or not the student was in the second year, fall scores, and an interaction if the student took a district test; and teachers' experience, whether teachers had a master's degree, whether teachers were assigned to the treatment group, and a variable indicating each school. In addition, the model includes student and teacher random effects.

The average score reported in the table for the treatment group is the unadjusted average score for the control group plus the product effect estimated from the model. It differs from the unadjusted average score for the treatment group.

Effect sizes are calculated by dividing the score difference (product effect) shown in the table by 21.06, which, by design, is the standard deviation of a national norm sample of NCE scores.

The main question in this part of the study was whether a second year of teaching experience using software products increased reading test scores. An increase in product effects from a year of experience would yield a positive difference between the second-year effect and the first-year effect. Table II.7 shows that the effect of an additional year of teaching experience using the product was 2.02 NCE units. This difference is not statistically different from zero ($p > 0.05$).[19] Table II.8 shows that, although fourth grade students who used software products were less likely to score below the 33rd percentile in the spring test than control students in both years of the study, these effects were not statistically significant at the 5 percent level. Furthermore, product effects on low scorers were not statistically significantly different between the two years of the study.

Additional analyses investigated the relationship between products effects on student test scores and usage. No statistically significant interactions were estimated between usage and product effects in either year.[20]

**Table II.8. Effect on Percentage of Students in Lowest Third of Reading Test Score, Fourth Grade**

| | Treatment Percentage | Control Percentage | Product Effect | Effect Size | *p*-value |
|---|---|---|---|---|---|
| Percentage of students below 33rd percentile of spring reading test in first year | 20.9 | 28.9 | -8.0 | -0.26 | 0.10 |
| Percentage of students below 33rd percentile of spring reading test in second year | 18.2 | 33.1 | -14.9 | -0.48 | 0.12 |
| Difference between first and second year | | | -6.9 | | 0.48 |

Note: Other variables in the model include student age, gender, indicator for the second year, the pretest indicator that a student was in the lowest third, and an interaction if the student took a district test; teacher's experience, whether he or she had a master's degree, an interaction of the second year and treatment, and a variable for each school. In addition, the model includes student and teacher random effects.

The treatment percentage reported in the table is the unadjusted control percentage plus the product effect. It differs from the unadjusted treatment percentage. The effect size is calculated using the Cox Index (the log odds ratio divided by 1.65).

[19] As noted in the methods discussion at the beginning of the chapter, the statistical test of the equality of effects in the two years is equivalent to the statistical test that the coefficient of the "year by treatment" interaction variable (the amount by which the effect in the second year differs from the effect in the first year) is statistically significant. Appendix C discusses the test.

[20] To assess this relationship, we first created variables at the teacher level for their students' average usage in the two study years. We then used the model to estimate product effects described in Appendix C and included usage variables as covariates in the second-level equations. The two estimated coefficients for usage were statistically insignificant (*p*-values are 0.09 for the coefficient on first-year usage and 0.46 for second-year usage).

## C. Findings for Sixth Grade Math Products

This section presents findings for sixth grade math software products. The analysis sample included 6 districts, 16 schools, 35 teachers who participated in both years of the study, and 2,279 students (1,620 from year 1 and 659 from year 2).[21] These teachers used one of two math products: Achieve Now (published by Plato), and Larson Pre-Algebra (published by Houghton-Mifflin). The 16 participating schools had 66 percent of students receiving free or reduced-price lunch, 42 percent Hispanic students, 30 percent black students, and an average student/teacher ratio of 17. Twenty-five percent of schools were in urban areas.

### Characteristics of Teachers and Students in Treatment and Control Classrooms

Teachers were randomly assigned to treatment or control groups in the first year. Since the sample used to study the effect of teachers' experience using software products on student test scores included only teachers who participated in the study for both years of the study, attrition and mobility created the possibility of differences between teachers in the treatment and control groups, but tests of equivalence indicated no statistically significant differences (see Table II.9). The study did not randomly assign students to teachers. Schools used whatever conventional approaches they normally used to allocate incoming students to teachers. Table II.9 shows that students with the treatment and control teachers were similar on fall test scores, age, and gender. None of the differences is statistically significant. Student and teacher characteristics are entered into models to adjust for remaining differences and increase the statistical power of the analysis.

### Teacher Training and Support During the First Year

Vendor training sessions generally took place in host districts and sometimes host schools during summer or early fall of 2004. The initial training averaged 6 hours and varied by product between 4 and 8 hours. Training topics included classroom management and alignment with standards and the local curriculum, and training sessions included opportunities to practice with the technology. Nearly all teachers (98 percent) attended the initial training, according to attendance logs. Vendors delivered ongoing support in several modes. Product representatives visited teachers (66 percent of teachers reported receiving this kind of help); vendors also provided support through e-mail or telephone help desks (40 percent) and additional training at schools (30 percent). The study team also worked with districts to identify hardware and software needs including computers, headphones, memory, and operating system upgrades; it also purchased a set of mobile laptop carts for one district in which access to school computer labs was too constrained to support adequate student use.

[21] The large sample size differences in the first and second years reflect the change in the sampling procedure in the second year. Many sixth grade teachers teach multiple sections of math. In the first year, all sections were included in the sample; in the second year, the study sampled one section.

**Table II.9 Characteristics of Teachers and Students in Treatment and Control Classrooms, Sixth Grade**

| | First Year | | | Second Year | | | First and Second Year | | |
|---|---|---|---|---|---|---|---|---|---|
| | Treatment Classrooms | Control Classrooms | *p*-value of the Difference | Treatment Classrooms | Control Classrooms | *p*-value of the Difference | Treatment Classrooms | Control Classrooms | *p*-value of the Difference |
| **Teacher Characteristics** | | | | | | | | | |
| Teaching experience (years) | | | | | | | 10.45<br>(9.69) | 15.73<br>(11.75) | 0.17 |
| Has a master's degree (percentage) | | | | | | | 16.67<br>(38.35) | 35.29<br>(49.26) | 0.22 |
| Female (percentage) | | | | | | | 61.11<br>(50.16) | 82.35<br>(39.30) | 0.17 |
| **Teacher Sample Size** | | | | | | | **18** | **17** | |
| **Student Characteristics** | | | | | | | | | |
| Female (percentage) | 50.94<br>(49.56) | 51.95<br>(48.98) | 0.68 | 53.47<br>(50.02) | 48.53<br>(50.06) | 0.27 | 51.65<br>(49.68) | 50.94<br>(49.30) | 0.74 |
| Age (years) | 11.60<br>(0.47) | 11.65<br>(0.47) | 0.14 | 11.62<br>(0.46) | 11.63<br>(0.45) | 0.74 | 11.61<br>(0.46) | 11.65<br>(0.47) | 0.18 |
| Unadjusted average score on fall math test (NCE) | 49.48<br>(20.97) | 50.11<br>(21.38) | 0.84 | 48.89<br>(18.34) | 51.07<br>(19.90) | 0.83 | 49.32<br>(20.25) | 50.40<br>(20.95) | 0.94 |
| Unadjusted average score on spring math test (NCE) | 52.84<br>(20.44) | 51.88<br>(19.98) | 0.82 | 48.58<br>(18.87) | 53.28<br>(21.49) | 0.45 | 51.63<br>(20.09) | 52.29<br>(20.43) | 0.93 |
| **Student Sample Size** | **887** | **733** | | **352** | **307** | | **1,239** | **1,040** | |

Note: Standard deviations in parentheses.

## Usage of Sixth Grade Math Products

Only one of the two products used in sixth grade included databases that tracked the time when the students were logged on. Table II.10 presents means and standard deviations of usage in minutes and number of weeks of usage.[22] Student usage averaged 851 minutes in the first year of the study and 679 minutes in the second year, about 80 percent of the first-year average (the difference in usage is statistically significant, $p < 0.01$). Furthermore, using a 40-week school year as a basis, minutes of usage averaged 17 minutes per week in the second year. The actual number of weeks of usage provided by product records (weeks in which at least one student recorded usage) was 13.1 weeks. Using the number of actual weeks of usage as a basis, average minutes of usage during the second year of the study was 52 minutes a week.

[22] Data are available for 70 percent of treatment students in the first year and 34 percent of treatment students in the second year. At the teacher level, these correspond to 56 percent of treatment teachers in the first year and 33 percent of treatment teachers in the second year.

**Table II.10. Average Student Product Usage in the First and Second Year, Sixth Grade**

| | Year 1 | Year 2 | Difference |
|---|---|---|---|
| Average minutes of student product usage | 851<br>(532) | 679<br>(661) | -172 |
| Average number of weeks in which students used products | 19.1<br>(8.84) | 13.1<br>(8.51) | -6 |
| Student sample sizes | 624 | 121 | |

Note: Standard deviations in parentheses. Source: Product records.

## Sixth Grade Product Effects Were Smaller in the Second Year

The same two-level model is used for sixth grade as for first and fourth grades. Table II.11 presents the treatment effects estimated with the two-level model for each year of the study, along with their difference, which we interpret as the effect of teachers' experience using software products on student test scores.

Table II.11 presents the product effect for the two study years and the difference of effects between the two years, which is interpreted as the effect of teachers' experience using products on test scores. The effects in the table are estimates from the two-level models (Appendix C presents estimates from the full model).

**Table II.11. Product Effects on Spring Math Test Scores, Sixth Grade**

| | Treatment Group | Control Group | Product Effect | Effect Size | *p*-value |
|---|---|---|---|---|---|
| Average score on spring math test, first year (NCE) | 51.44 | 51.88 | -0.44 | -0.02 | >0.50 |
| Average score on spring math test, second year (NCE) | 50.04 | 53.28 | -3.24 | -0.15 | 0.11 |
| Difference between first and second year | | | -2.80 | | 0.02 |
| Student sample size (first year plus second year) | 1,239 | 1,040 | | | |

Note: Details of the estimation model are presented in Appendix C and Table C.1. Variables in the model include student age, gender, whether or not the student was in the second year, fall scores, and an interaction if the student took a district test; and teachers' experience, whether teachers had a master's degree, whether teachers were assigned to the treatment group, and a variable for each school. In addition, the model includes student and teacher random effects.

The average score reported in the table for the treatment group is the unadjusted average score for the control group plus the product effect estimated from the model. It differs from the unadjusted average score for the treatment group.

Effect sizes are calculated by dividing the score difference (product effect) shown in the table by 21.06, which, by design, is the standard deviation of a national norm sample of NCE scores.

Table II.11 shows that, in the first year, the effect of software products on student math test scores was -0.44 NCE units, corresponding to the average student going from the 50th percentile to the 49th percentile. The effect is statistically insignificant. In the second year, the effect of software products on math test scores was -3.24 NCE units, corresponding to a student going from the 50th percentile to the 44th percentile. This effect is statistically insignificant.

The main question in this part of the study was whether a second year of teaching experience using software products increased math test scores. An increase in product effects from a year of experience would yield a positive difference between the second-year effect and the first-year effect. Table II.11 shows that the effect of an additional year of teaching experience using the product was -2.80 NCE units. This difference is statistically different from zero ($p$ = 0.02).[23] Using the information noted above, the statistically significant experience effect is the 5-percentile-point difference (scores declining to the 49th percentile in the first year and to the 44th percentile in the second year).

Table II.12 shows that, although sixth grade students who used software products were more likely to score below the 33rd percentile in the spring test than control students in both years of the study, only the first-year effect is statistically significant at the 5 percent level. Furthermore, product effects were not statistically significantly different between the two years of the study.

**Table II.12. Effect on Percentage of Students in Lowest Third of Math Test Score, Sixth Grade**

| | Treatment Group | Control Group | Product Effect | Effect Size | *p*-value |
|---|---|---|---|---|---|
| Percentage of students below 33rd percentile of spring math test in first year | 32.9 | 31.8 | 1.1 | 0.03 | >0.50 |
| Percentage of students below 33rd percentile of spring math test in second year | 38.2 | 30.0 | 8.2 | 0.22 | >0.50 |
| Difference between second and first year | | | 7.1 | | 0.28 |

Note: Other variables in the model include student age, gender, indicator for second year, the pretest experience, whether the teacher had a master's degree, and an interaction of second year and student. In addition, the model includes student and teacher random effects.

The treatment percentage reported in the table is the control percentage plus the product effect. It differs from the unadjusted treatment percentage. The effect size is calculated using the Cox Index (the log odds ratio divided by 1.65).

[23] As noted in the methods discussion at the beginning of the chapter, the statistical test of the equality of effects in the two years is equivalent to the statistical test that the coefficient of the "year by treatment" interaction variable (the amount by which the effect in the second year differs from the effect in the first year) is statistically significant. Appendix C discusses the test.

Additional analyses investigated the relationship between product effects on student test scores and usage. However, interactions between usage and product effects were not statistically significant in either year.[24]

## D. Findings for Algebra I Products

This section presents findings for algebra I software products. The analysis sample included 6 districts, 13 schools, 24 teachers who participated in both the first and second years of the study, and 1,051 students (517 from year 1 and 534 from year 2). Teachers used one of two algebra I products: Larson Algebra I (published by Houghton-Mifflin), and Cognitive Tutor (published by Carnegie Learning). Participating schools had 50 percent of students receiving free or reduced-price lunch, 13 percent Hispanic students, 41 percent black students, and a student/teacher ratio of 17. Sixty-two percent of schools were located in urban areas.

### Characteristics of Teachers and Students in Treatment and Control Classrooms

Teachers were randomly assigned to treatment or control groups in the first year. Since the sample used to study the effect of teachers' experience using software products on student test scores included only teachers who participated in the study for both years, attrition and mobility created the possibility of differences between teachers in the treatment and control groups, but tests of equivalence show no statistically significant differences (see Table II.13). The study did not randomly assign students to teachers. Schools used whatever conventional approaches they normally used to allocate incoming students to teachers. Table II.13 shows that students with the treatment and control teachers were similar on fall test scores, age, and gender. None of the differences is statistically significant. Student and teacher characteristics are entered into models to adjust for remaining differences and increase the statistical power of the analysis.

[24] To assess this relationship, we first created two variables at the teacher level that summarized their students' average usage in each year of the study. We then used the model to estimate product effects described in Appendix C and included the usage variables as covariates in the second-level equations. The two estimated coefficients for usage were statistically insignificant (*p*-values are 0.29 for the coefficient on first-year usage and 0.73 for second-year usage).

**Table II.13 Characteristics of Teachers and Students in Treatment and Control Classrooms, Algebra I**

| | First Year | | | Second Year | | | First and Second Year | | |
|---|---|---|---|---|---|---|---|---|---|
| | Treatment Classrooms | Control Classrooms | *p*-value of the Difference | Treatment Classrooms | Control Classrooms | *p*-value of the Difference | Treatment Classrooms | Control Classrooms | *p*-value of the Difference |
| **Teacher Characteristics** | | | | | | | | | |
| Teaching experience (years) | | | | | | | 16.33<br>(9.90) | 12.48<br>(10.54) | 0.39 |
| Has a master's degree (percentage) | | | | | | | 33.33<br>(49.24) | 66.67<br>(49.24) | 0.12 |
| Female (percentage) | | | | | | | 41.67<br>(51.49) | 58.33<br>(51.49) | 0.42 |
| **Teacher Sample Size** | | | | | | | **12** | **12** | |
| **Student Characteristics** | | | | | | | | | |
| Female (percentage) | 49.26<br>(50.09) | 48.18<br>(50.07) | 0.93 | 50.54<br>(50.09) | 50.19<br>(50.10) | 0.94 | 49.91<br>(50.05) | 49.21<br>(50.04) | 0.92 |
| Age (years) | 14.84<br>(0.90) | 14.83<br>(0.72) | 0.77 | 15.27<br>(0.56) | 15.37<br>(0.79) | 0.60 | 15.06<br>(0.78) | 15.10<br>(0.80) | 0.62 |
| Unadjusted average score on fall ETS algebra test (percent correct) | 31.40<br>(11.24) | 35.27<br>(10.75) | 0.12 | 34.70<br>(11.29) | 34.73<br>(10.43) | 0.86 | 33.07<br>(11.38) | 35.00<br>(10.58) | 0.40 |
| Unadjusted average score on spring ETS algebra test (percent correct) | 33.17<br>(12.47) | 37.82<br>(12.97) | 0.27 | 38.33<br>(15.31) | 37.07<br>(13.71) | 0.75 | 35.78<br>(14.21) | 37.44<br>(13.34) | 0.44 |
| **Student Sample Size** | **270** | **247** | | **277** | **257** | | **547** | **504** | |

Note: Standard deviations are in parentheses.

Tests of treatment and control differences were conducted using a two-level hierarchical model with classroom treatment status as a fixed effect and teacher and student random effects. The *p*-value of the difference shown in the table is the *p*-value of the estimated treatment coefficient.

## Teacher Training and Support During the First Year

Treatment teachers were trained in their host districts or schools during summer or early fall of 2004. The initial training provided by the three algebra product vendors averaged 12 hours, varying between 4 and 23 hours depending on the product. Topics included classroom management and alignment with standards and the local curriculum. Teachers were also able to practice using the product. Nearly all teachers (97 percent) attended the initial training, according to attendance logs. Ongoing support was provided by vendors in various modes. Some had company representatives visit teachers (28 percent of teachers reported receiving this kind of support), supported e-mail or telephone help desks (36 percent of teachers said they were aware of or used this kind of support), and provided additional training at schools (which 14 percent of teachers reported receiving). The study team also purchased some hardware and software upgrades in host schools, but to a lesser extent than in other grade levels.

## Usage of Algebra I Products

The two algebra I products tracked minutes students were logged on. Table II.14 presents means and standard deviations of the average number of student minutes of usage and the average number of weeks during which students used products, for students for whom these data were available.[25] The average student used products 1,309 minutes in the first year and 1,450 minutes in the second year (the difference in usage is statistically significant, $p < 0.01$). Products were used in 14.4 weeks in the first year and 16.8 weeks in the second year. Average minutes of usage in the second year was 86 minutes a week when products were used.

**Table II.14. Student Product Usage in the First and Second Year, Algebra**

| | Year 1 | Year 2 | Difference |
|---|---|---|---|
| Average minutes of student product usage | 1,309<br>(1,274) | 1,450<br>(1,271) | 141 |
| Average number of weeks in which students used products | 14.4<br>(11.2) | 16.8<br>(11.4) | 2.4 |
| Student sample sizes | 254 | 161 | |

Note: Standard deviations are in parentheses. Source: Product records.

## Effects of Algebra I Products Were Larger in the Second Year

Table II.15 shows that, in the first year, the effect of software products on student algebra I test scores was -0.34 in percent correct units, corresponding to the average student going from the 50th percentile to the 49th percentile. The effect is statistically insignificant. In the second year, the effect of software products on algebra I test scores was 2.56 percent correct units, corresponding to a student going from the 50th percentile to the 56th percentile. This effect is statistically significant.

The main question in this part of the study was whether a second year of teaching experience using software products increased algebra I test scores. An increase in product effects from a year of experience would yield a positive difference between the second-year effect and the first-year effect. Table II.15 shows that the effect of an additional year of teaching experience using the product was 2.90 percent correct units. This difference is

[25] Data are available for almost all treatment students in the first year and 58 percent of treatment students in the second year. At the teacher level, these correspond to all treatment teachers in the first year and 83 percent of treatment teachers in the second year.

statistically different from zero ($p$ = 0.05).[26] The experience effect indicates that in the first year students in the treatment and control classrooms scored nearly the same on the spring test (37.48 percent compared to 37.82 percent), whereas in the second year students in treatment classrooms scored higher than students in control classrooms (39.63 percent compared to 37.02 percent).

**Table II.15. Product Effects on Spring ETS Algebra Test Scores**

| | Treatment Group | Control Group | Product Effect | Effect Size | *p*-value |
|---|---|---|---|---|---|
| Average spring ETS exam score, first year (percent correct) | 37.48 | 37.82 | -0.34 | -0.02 | >0.50 |
| Average spring ETS exam score, second year (percent correct) | 39.63 | 37.07 | 2.56 | 0.15 | 0.03 |
| Difference between first and second year | | | 2.90 | | 0.05 |
| Student sample size (first year plus second year) | 547 | 504 | | | |

Note: Details of the estimation model are presented in Appendix C and Table C.1. Variables in the model include at level 1: student age, gender, whether the student was in the second year, pretest scores, and an interaction if the student took a district test; at level 2: teachers' experience, whether he or she had a master's degree, whether he or she was a treatment teacher, and a variable for each school; the model also includes student and teacher random effects.

The treatment average score reported in the table is the unadjusted control average score plus the treatment effect estimated from the model. It differs from the unadjusted treatment score.

Effect sizes are calculated by dividing the score difference (product effect) shown in the table by 17, the standard deviation of scores reported by ETS.

Additional analyses investigated the relationship between product effects on student test scores and usage. However, interactions between usage and product effects were not statistically significant in either year.[27]

[26] As noted in the methods discussion at the beginning of this chapter, the statistical test of the equality of effects in the two years is equivalent to the statistical test that the coefficient of the "year by treatment" interaction variable (the amount by which the effect in the second year differs from the effect in the first year) is statistically significant. Appendix C discusses the test.

[27] To assess this relationship, we first created two variables at the teacher level that summarized their students' average usage in each year of the study. We then used the model to estimate product effects described in Appendix C and included the usage variables as covariates in the second-level equations. The two estimated coefficients for usage were statistically insignificant ($p$-values are 0.31 for the coefficient on first-year usage and 0.97 for second-year usage).

### Summary

For the 10 products that were studied in the second year, the analysis estimated whether product effects on test scores differed in the two years, to test the hypothesis that a year of experience using products influenced effectiveness. The findings are mixed. In the first and fourth grade levels, experience effects are statistically insignificant. For the sixth grade math and algebra levels, the estimated effects on math test scores were statistically significantly larger in the second year than in the first year. For sixth grade math, product effects were negative. For algebra, product effects were positive.

The findings should be interpreted in the context of the study's limitations. The second year of the study had six fewer products than the first year. For those products studied in the second year, the analysis includes only teachers who remained in the second year of the study and were sampled for collecting student test scores. To the extent that whether teachers leave their grade level or school may be related to their assignment to use or not use products, unobservable factors may affect the findings. Also, the study did not observe classrooms and interview teachers to ask directly about how they may have changed their use of products in the second year based on their experience using them in the first year.

# Chapter III

# Effectiveness of 10 Educational Software Products

This part of the study analyzes the effect on test scores of the 10 software products that were investigated in the second year of the study. To increase statistical power, data from both study years are used to estimate effects of the 10 products.

The study essentially is a set of studies of individual products that share the same experimental design and data collection structure. Each of the products was implemented in a set of volunteering districts, schools, and classrooms. For each product, fall and spring test scores were collected and combined with other student, teacher, and school data to estimate effects using nested models.

Notwithstanding the consistency of the study approach, comparing effects of different products should be done with caution. Each product was implemented in a different set of districts and schools and the differences may affect the findings. More research would be needed to determine whether a product that is ineffective in the set of schools and districts that implemented it in the study might be effective if it were implemented in other settings.

The chapter follows a template in reporting study findings for each product. It describes the product's main features based on information from developers and reviews of materials, provides basic cost information based on information from developers, describes characteristics of the student sample, and presents estimates of product effects on test scores. The chapter does not discuss the implementation experience and classroom observation findings for individual products from the first year. As noted in the first chapter, classroom observations were not conducted in the second year. The chapter also reports overall and second-year findings for each product as a way to separate the effects that experience may have had for the products. First-year effects are not broken out separately because, in the first year, the study operated under the guideline that individual product effects would not be reported.

### Approach for Estimating Effects of Individual Products

A "product effect" in this context is the difference in spring student test scores between treatment and control classrooms caused by the assignment of treatment classrooms to use a software product.

The study calculated the effect of each product by combining the samples of schools, teachers, and students who participated in the first year of the study, the second year, or both.

Product effects were estimated using a two-level model. The outcome at the first level is the spring student test score, which is modeled as a function of a student's fall test score, age and gender, and a student random effect. The outcome at the second level is the classroom-average test score, which is modeled as a function of teacher years of experience and education level. The second level also includes an indicator variable for the teacher's school and a teacher random effect.

The models use multiple imputation methods to impute missing subtest scores, student age, and student gender. The specific approach was the Markov Chain Monte Carlo (MCMC) method in SAS 9. The imputation was done five times, separately for students in treatment and control classrooms. Tests based on first-year data indicated that the MCMC method had a high degree of predictive power for imputing subtest scores (when subtest scores for random samples of students were set to "missing" and the MCMC method was used to impute them, correlations between the actual and imputed values ranged from 90 percent to 95 percent).

The HLM 6.02 package was used for estimation of the multilevel models. The estimation and other statistical procedures used the five imputed data sets and the HLM application produced variances of the estimates that incorporated the added variance from the imputation. Appendix C presents details on the models.

## A. First Grade Reading: Destination Reading

*Destination Reading,* published by Riverdeep, is a supplemental reading program that seeks to improve phonics, decoding, reading comprehension, and other reading skills. The product studied here is course 1, which covers material for students in kindergarten and first grade. Teachers introduce concepts to the entire class and students then work individually with the software, which provides assessments informing teachers about student progress. Teachers can change the sequence of activities for the entire class or for individual students. The product can be used in a computer lab or in the classroom, and is recommended to be used for 20 minutes at least two times a week. Teacher training takes two to three days on site, and districts can purchase additional training and coaching. The vendor provides ongoing support during the school year by phone and e-mail. The study estimated the annualized cost per student to be $78. Of that amount, 68 percent is the license fee and the remaining 32 percent is for teacher training and support, technical support, and printed materials and supplies.[28]

[28] Cost data for all products apply to the 2004-2005 school year and were provided by developers.

More information about the product, its technical requirements, and contact information can be found at http://www.riverdeep.net.

### Study Design and Context

Across the two years of the study, the study was implemented in 12 schools in two urban districts. The two districts averaged 97 schools and 61,143 students. Thirty-five first grade teachers volunteered to participate in the first or second year. Twenty-one teachers were assigned randomly to use the *Destination Reading* product and 14 were assigned not to use the product. Each school had at least one treatment teacher and one control teacher.

For the two years of the study, fall and spring reading test scores are available for 742 students who participated in the first or second year of the study. The average student in the study had reading skills at the 43rd percentile (on the fall test). Thirty-five percent of students were reading in the lowest third (their reading scores placed them below the 33rd percentile). The average age of the students was 6.7 years, and 48 percent were female. Teachers had, on average, 16 years of teaching experience, and 43 percent had a master's degree.

The product tracks the time students were logged on. During the two years of the study, the average student was logged on 615 minutes a year (s.d. 395 minutes), and used a product during 25 weeks.

In the second year, the study was implemented in nine schools in the same two districts. Fifteen teachers were assigned randomly to use the *Destination Reading* product and 10 were assigned not to use the product. Teachers had on average 13.3 years of teaching experience and 35 percent had a master's degree. A fall and spring SAT-10 reading test was administered to 453 students. The average student in the second year had reading skills at the 45th percentile in fall 2005, and 44 percent of students were reading below the 33rd percentile. The average age of the students was 6.7 years and 48 percent were female. In the second year, the average student was logged on 693 minutes and used a product during 27 weeks.

### Findings

The estimated treatment effect (in normal curve equivalent units) is 1.91 (*p*-value = 0.27). The estimated treatment effect is not statistically significant at the 0.05 level of significance. Using only second-year data, the estimated treatment effect (in normal curve equivalent units) is 2.19 (*p*-value = 0.31).

## B. First Grade Reading: Headsprout Early Reading

*Headsprout Early Reading,* published by Headsprout, is a supplemental reading program to improve skills in areas including phonemic awareness, phonics, fluency, vocabulary, and comprehension. The product consists of 80 episodes. The first 40 episodes focus on decoding, segmenting, and blending. The next 40 episodes focus on vocabulary, reading fluency, and comprehension. Students work through the program at their own pace. The product generates assessments informing teachers about students' usage and their progress through the sequence of episodes. The product can be used in a school computer lab or in the classroom, and recommended usage is 30 minutes a day at least three times a week. Teachers receive one day of initial training, which may be completed via phone, web, or in person. Ongoing support during the school year is available by phone and e-mail. The study estimates the annualized cost per student to be $146. Of that amount, 85 percent is the license fee and the remaining 15 percent is for teacher training and support, technical support, and printed materials and supplies.

More information about the product, its technical requirements, and contact information can be found at http://www.headsprout.com.

**Study Design and Context**

Across the two years of the study, the study was implemented in 3 districts and 12 schools. One district was in an urban area, one was in the urban fringe, and one was in a rural area. The average district had 59 schools and 47,723 students. Sixty-three first grade teachers volunteered to participate in the first or second year. Thirty-two teachers were assigned randomly to use the product and 31 were assigned not to use the product, with at least a pair of treatment and control teachers in each school.

For the two years of the study, fall and spring reading test scores were obtained for 1,079 students who participated in the first or second year of the study. The scores indicated that the average student in the study had reading skills at the 65th percentile on the fall test. Twenty-three percent of students were reading below the 33rd percentile. The average age of the students was 6.7 years and 48 percent were female. Teachers had on average 11 years of teaching experience and 58 percent had a master's degree.

Headsprout uses a database that tracks time students are logged on. During the two years of the study, the average student was logged on to the product 857 minutes a year (s.d. 326 minutes), and used the product during 26 weeks a year.

In the second year, the study was implemented in 3 districts and 7 schools. The districts were located in an urban area, an urban fringe, and a rural area correspondingly. Eighteen first grade teachers volunteered to participate in the second year. Nine teachers were assigned randomly to use the product and nine were assigned not to use the product, with at least a pair of treatment and control teachers in each school. Teachers had on average 13 years of teaching experience and 71 percent had a master's degree. A fall and spring SAT-10 reading test was administered to 268 students. The average student in the second-year sample had reading skills at the 65th percentile in fall 2005, and 28 percent of students were reading below the 33rd percentile. The average age of the students was 6.7 years and 48 percent were female. The average student was logged on for 772 minutes in the second year and used the product during 22 weeks.

### Findings

The estimated treatment effect (in normal curve equivalent units) is 0.29 (*p*-value = 0.79). The estimated treatment effect is not statistically significant at the 0.05 level of significance. Using only second year data, the estimated treatment effect (in normal curve equivalent units) is –4.13 (*p*-value = 0.06).

## C. First Grade Reading: PLATO Focus

*PLATO Focus,* published by PLATO Learning Corporation, is a complete reading curriculum to develop skills in phonemic awareness, phonics, fluency, vocabulary, and reading comprehension. Students spend 30 to 45 minutes on activities led by the instructor, 15 to 30 minutes on associated computer-based activities, and 30 to 45 minutes on related print-based activities. The teacher can choose the order and difficulty level for the computer-based activities. The product can be used in the classroom or in a computer lab where a reading specialist trained in PLATO is monitoring the students. The product generates progress reports for each student. Teachers receive three to six days of training, at least one day of training during the school year, and at least one in-class consultation. Ongoing support during the school year is available by phone and through a website. The study estimates the annualized cost per student to be $351. Of that amount, 27 percent is the license fee and the remaining 73 percent is for teacher training and support, technical support, and printed materials and supplies.

More information about the product, its technical requirements, and contact information can be found at http://www.plato.com/Products/PLATO-Focus-Reading-and-Language-Program.aspx.

### Study Design and Context

Across the two years of the study, the study was implemented in eight schools in three districts. Two districts were located in an urban fringe area and one was in an urban area. The average district had 13 schools and 6,966 students.

Twenty-nine first grade teachers volunteered to participate in the first or second year. Fifteen teachers were assigned randomly to use the product and 14 were assigned not to use the product, with at least a pair of treatment and control teachers in each school.

Across the two years of the study, fall and spring reading test scores were obtained for 618 students. The scores indicated that the average student in the study had reading skills at the 40th percentile on the fall test. Fifty percent were reading below the 33rd percentile. The average age of the students was 6.6 years and 52 percent were female. Teachers had on average 17 years of teaching experience and 55 percent had a master's degree. The software did not provide data on usage by student.

In the second year, the study was implemented in eight schools in three districts. Two of these districts were located in an urban fringe area and one in an urban area. The average district had 13 schools and 6,966 students. Eighteen first grade teachers volunteered to participate in the second year. Nine teachers were assigned randomly to use the product and nine were assigned not to use the product, with at least a pair of treatment and control teachers in each school. Teachers had on average 20 years of teaching experience and 65 percent had a master's degree. A fall and spring SAT-10 reading test was administered to 319 students. The average student scored at the 42nd percentile in fall 2005 and 53 percent were reading below the 33th percentile. The average age of the students was 6.7 years and 50 percent were female.

### Findings

The estimated treatment effect (in normal curve equivalent units) is 0.50 (*p*-value = 0.72). The estimated treatment effect is not statistically significant at the 0.05 level of significance. Using only second year data, the estimated treatment effect (in normal curve equivalent units) is –.10 (*p*-value = 0.95).

## D. First Grade Reading: Waterford Early Reading

*Waterford Early Reading Program,* published by Pearson Digital Learning, is a supplemental reading program with three levels. Level 1, typically used in kindergarten, covers topics including letter recognition, phonemic awareness, and print concepts. Level 2, typically used in first grade, covers topics including letter sounds, word recognition, and comprehension. Level 3, typically used in second grade, covers topics including spelling and encoding, fluency, and the writing process. A course on phonological awareness can be added to the first two levels of instruction. The product generates reports to inform teachers about student progress. All students work at their own pace through sequenced activities. Student books are sent home weekly with directions for parents. The program can be used in the classroom or in a computer lab. Recommended usage is from 17 to 30 minutes, depending on the level, at least three times a week. Teacher training takes about two days on site. Ongoing support during the school year is available by phone, e-mail, and through a website. The study estimates annualized cost per student to be $223. Of that amount, 54 percent of the cost is the license fee and the remaining 46 percent is for teacher training and support, technical support, and printed materials and supplies.

More information about the product, its technical requirements, and contact information can be found at http://www.pearsondigital.com/waterford/.

### Study Design and Context

Across the two years of the study, the study was implemented in 13 schools in 3 districts. Two districts were in urban fringe areas and one district was in an urban area. The average district had 68 schools and 49,450 students. Forty-six first grade teachers volunteered to participate in the first or second year. Twenty-eight teachers were assigned randomly to use the product and 18 were assigned not to use the product, with at least a pair of treatment and control teachers in each school.

Across the two years of the study, fall and spring reading test scores were obtained for 1,155 students. The scores indicated that the average student in the study had reading skills at the 52nd percentile on the fall test, and 31 percent were reading below the 33rd percentile. The average age of the students was 6.6 years and 48 percent were female. Teachers had 11 years of teaching experience on average and 35 percent had a master's degree.

The product includes a database that tracks time students were logged on. During the two years of the study, the average student was logged on for 3,643 minutes a year (s.d. 1,029 minutes), and usage occurred during 34 weeks.

In the second year, the study was implemented in nine schools in three districts. Two districts were located in urban fringe areas and one district was in an urban area. The average district had 68 schools and 49,450 students. Twenty first grade teachers volunteered to participate in the second year. Eleven teachers were assigned randomly to use the product and nine were assigned not to use the product, with at least a pair of treatment and control teachers in each school. Teachers had on average nine years of teaching experience and 26 percent had a master's degree. A fall and spring SAT-10 reading test was administered to 331 students. The average student in the study had reading skills at the 54th percentile in fall 2005, and 34 percent were reading below the 33th percentile. The average age of the students was 6.6 years and 46 percent were female. The average student was logged on for 2,794 minutes in the second year and used the product during 35 weeks.

### Findings

The estimated treatment effect (in normal curve equivalent units) is 0.42 (*p*-value = 0.77). The estimated treatment effect is not statistically significant at the 0.05 level of significance. Using only second-year data, the estimated treatment effect (in normal curve equivalent units) is -1.76 (*p*-value = 0.41).

## E. Fourth Grade Reading: Academy of Reading

*Academy of Reading*, published by Autoskill, Inc., is a set of exercises to improve phonemic awareness and sound-symbol association, phonics and decoding skills, fluency and comprehension, and reading proficiency. Students work through exercises at their own pace. The product provides assessments for teachers about student usage and progress through the exercises. The program is designed to be used in computer labs and recommended usage is 25 minutes a day for three or more days a week. Teachers receive one day of initial training on how to use the product and one day of training four to six weeks later as a follow-up. The publisher also provides ongoing support by phone, e-mail, and webinars. The study estimates the annualized cost per student to be $217. Of that amount, 51 percent is the license fee and the remaining 49 percent is for teacher training and support, technical support, and printed materials and supplies.

More information about the product, its technical requirements, and contact information can be found at http://www.autoskill.com/products/reading/index.php.

## Study Design and Context

Across the two years of the study, the study was implemented in 15 schools in 4 districts. Two districts were in the fringe of urban areas and two in urban areas. The average district had 95 schools and 64,342 students. Forty-one fourth grade teachers volunteered to participate in the first or second year. Twenty-two teachers were assigned randomly to use the product and 19 were assigned not to use the product, with at least a pair of treatment and control teachers in each school. Teachers averaged nine years of experience and 32 percent had master's degrees.

Across the two years of the study, fall and spring reading test scores were obtained for 899 students. The scores indicated that the average student had reading skills at the 34th percentile on the fall test, and 53 percent of students scored below the 33rd percentile. The average age of the students was 9.7 years and 50 percent were female.

The product used a database that tracked time students were logged on. During the two years of the study, the average student was logged on 624 minutes a year (s.d. 384 minutes) and usage occurred during 13 weeks.

In the second year of the study, the study was implemented in seven schools in two districts. Both districts were in the fringe of urban areas, and the average district had 95 schools and 68,000 students. Fourteen fourth grade teachers volunteered to participate in the second year. Seven teachers were assigned randomly to use the product and seven were assigned not to use the product, with at least a pair of treatment and control teachers in each school. Teachers had 18.5 years of experience on average and 33 percent had master's degrees. A fall and spring SAT-10 reading test was administered to 282 students. The scores indicated that the average student had reading skills at the 48th percentile in fall 2005 and 38 percent were below the 33rd percentile. The average age of the students was 9.8 years and 49 percent were female. The average student was logged on for 951 minutes in the second year and used the product during 16 weeks.

## Findings

The estimated treatment effect (in normal curve equivalent units) is –0.16 ($p$-value = 0.88). The estimated treatment effect is not statistically significant at the 0.05 level of significance. Using only second-year data, the estimated treatment effect (in normal curve equivalent units) is 1.86 ($p$-value = 0.54).

## F. Fourth Grade Reading: LeapTrack

*LeapTrack,* published by LeapFrog SchoolHouse, is a supplemental reading product to improve phonemic awareness, phonics, vocabulary, and reading comprehension in addition to other reading skills (the program also includes math, which was not studied here). Teachers use LeapTrack Assessments to identify the skills students need to develop and, based on the assessments, the program provides a "Learning Path" for each student, a list of skill cards, and books for the student to complete to learn the skill. Students work on these activities at their own pace using the LeapPad, LeapTrack skill cards, and LeapFrog

SchoolHouse books. The program is recommended to be used in classrooms for at least 15 minutes three to five days a week. Teachers receive a day of pre-implementation training and up to four days of follow-up training. Ongoing support during the school year is available by phone and through a website. The study estimated the annualized cost per student to be $154. Of that amount, 47 percent is the license fee and the remaining 53 percent is for teacher training and support, technical support, and printed materials and supplies.

More information about the product, its technical requirements, and contact information can be found at http://www.leapfrogschoolhouse.com.

### Study Design and Context

Across the two years of the study, the study was implemented in 19 schools in 4 districts. Two districts were located in an urban fringe area and the other two in urban areas. The average district had 87 schools and 38,050 students. Fifty-five fourth grade teachers volunteered to participate in the first or second year. Twenty-nine teachers were assigned randomly to use the product and 26 were assigned not to use the product, with at least a pair of treatment and control teachers in each school.

Across the two years of the study, fall and spring reading test scores were obtained for 1,274 students. The scores indicated that the average student in the study had reading skills at the 38th percentile in the fall test, and 50 percent had reading skills below the 33rd percentile. The average age of the students was 9.7 years and 51 percent were female. Teachers had, on average, 11 years of teaching experience and 35 percent had a master's degree.

LeapTrack includes a database that tracks time students were logged on. During the two years of the study, the average student was logged on to the product for 520 minutes a year (s.d. 352 minutes). The product does not track weeks of usage.

In the second year, the study was implemented in four schools in two districts. One district was located in an urban fringe area and the other in an urban area. The average district had 14 schools and 6,500 students. Eight fourth grade teachers volunteered to participate in the second year. Four teachers were assigned randomly to use the product and four were assigned not to use the product, with at least a pair of treatment and control teachers in each school. Teachers had on average 20 years of teaching experience and 65 percent had a master's degree. A fall and spring SAT-10 reading test was administered to 181 students. The scores indicated that the average student in the study had reading skills at the 56th percentile in fall 2005 and 32 percent were below the 33th percentile. The average age of the students was 9.6 years and 48 percent were female. Average student product usage was 883 minutes during the second year.

### Findings

The estimated treatment effect (in normal curve equivalent units) is 1.97 (*p*-value = 0.01). The estimated treatment effect is statistically significant at the 0.05 level of significance. Using only second-year data, the estimated treatment effect (in normal curve equivalent units) is 2.88 (*p*-value = 0.20).

## G. Sixth Grade Math: PLATO Achieve Now

PLATO Achieve Now Mathematics Series 3, published by PLATO Learning Inc., is a supplemental math program. for teaching pre-algebraic topics that include rational numbers in related organizational patterns, proportion and percent, integers, probability, statistics, problem solving, geometry, measurement, and the foundational concepts of algebra I. Students use the product for independent practice and reinforcement of math skills. The courseware contains an assessment component that helps place the students within it. Based on the assessments, students work at their own pace on activities identified by the teacher. Recommended usage is 30 minutes per day, four days a week, for at least 10 weeks. Teachers receive training through web-based meetings and on-line self-tutorials. Ongoing support during the school year also is provided. The study estimated the annualized cost per student to be $36. Of that amount, 42 percent is the license fee and the remaining 58 percent is for teacher training and support, technical support, and printed materials and supplies.

More information about the product, its technical requirements, and contact information can be found at http://www.plato.com/Elementary-Solutions/Elementary-Mathematics/PLATO-Achieve-Now-Mathematics.aspx.

### Study Design and Context

Across the two years of the study, the study was implemented in 13 schools in 3 districts. Two of these districts were located in an urban fringe area and one in a small town. The average district had 63 schools and about 41,000 students. Thirty-nine sixth grade teachers volunteered to participate in the first or second year. Twenty-one teachers were assigned randomly to use the product and 18 were assigned not to use the product, with at least a pair of treatment and control teachers in each school.

Across the two years of the study, fall and spring math test scores were obtained for 1,037 students. The scores indicated that students in the study had math skills at the 41st percentile in the fall test and 40 percent were below the 33rd percentile. The average age of the students was 11.7 years and 53 percent were female. Teachers had, on average, 11 years of teaching experience and 33 percent had a master's degree. The product does not provide the usage time by student.

In the second year, the study was implemented in eight schools in three districts. Two of these districts were located in an urban fringe area and one in a town. The average district had 63 schools and 41,083 students. Eighteen sixth grade teachers volunteered to participate in the second year. Nine teachers were assigned randomly to use the product and nine were assigned not to use the product, with at least a pair of treatment and control teachers in each school. Teachers had on average 12 years of teaching experience and 35 percent had a master's degree. A fall and spring SAT-10 mathematics test was administered to 313 students. The scores indicated that the average student in the study had math skills at the

40th percentile in fall 2005 and 49 percent were below the 33rd percentile. The average age of the students was 11.7 years and 50 percent were female.

### Findings

The estimated treatment effect (in normal curve equivalent units) is –0.58 (*p*-value = 0.69). The estimated treatment effect is not statistically significant at the 0.05 level of significance. Using only second-year data, the estimated treatment effect (in normal curve equivalent units) is –1.59 (*p*-value = 0.72). The second-year treatment effect is not statistically significant at the 0.05 level of significance.

## H. Sixth Grade Math: Larson Pre-Algebra

*Larson Pre-Algebra,* published by Houghton-Mifflin, is designed to supplement the curriculum with extra instruction, practice, and assessments. This product is the same as *Larson Algebra I* but starts at different points within the sequence. It covers whole numbers, fractions, decimals, percents, rational numbers, probability and statistics, coordinate geometry, pre-algebra, and algebra I. The program addresses both skill building and problem solving, and allows the teacher to track student progress. Teachers can choose the quantity and order in which the topics are presented to the students. Recommended time of usage varies according to the number of topics and the number of weeks in the course; however, the developer recommends at least once-weekly usage. The program is designed to be used in computer labs. Teachers receive two hours to one day of pre-implementation training. Ongoing support during the school year is available by phone, e-mail, and through a website. The study estimated the annualized cost per student to be $15. Of that amount, 60 percent is the license fee and the remaining 40 percent is for teacher training and support, technical support, and printed materials and supplies).

More information about the product, its technical requirements, and contact information can be found at http://www.larsonmath.com/lmc_prea/prealgebra.htm.

### Study Design and Context

Across the two years of the study, the study was implemented in 13 schools in 5 districts. Three districts were in an urban fringe area and two were in an urban area. The average district had 225 schools and 186,975 students. Thirty-nine sixth grade teachers volunteered to participate in the first or second year. Twenty-four teachers were assigned randomly to use the product and 15 were assigned not to use the product, with at least a pair of treatment and control teachers in each school.

Across the two years of the study, fall and spring math test scores were obtained for 2,588 students. The scores indicated that the average student in the study had math skills at the 55th percentile in the fall test, and 35 percent had math skills below the 33rd percentile. The average age of the students was 11.6 years and 51 percent were female. Teachers had, on average, 11 years of teaching experience and 32 percent had a master's degree.

The product includes a database that tracks time students were logged on. During the two years of the study, the average student was logged on for 817 minutes a year (s.d. 502 minutes) and usage occurred during 19 weeks.

In the second year, the study was implemented in eight schools in three districts. Two of these districts were located in an urban fringe area and one in an urban area. The average district had 335 schools and 285,887 students. Eighteen sixth grade teachers volunteered to participate in the second year. Ten teachers were assigned randomly to use the product and eight were assigned not to use the product, with at least a pair of treatment and control teachers in each school. Teachers had on average 12 years of teaching experience and 11 percent had a master's degree. A fall and spring SAT-10 math test was administered to 386 students. The scores indicated that students in the study had math skills at the 56th percentile in fall 2005 and 30 percent were below the 33rd percentile. The average age of the students was 11.6 years and 50 percent were female. The average student was logged on for 642 minutes in the second year and used the product during 13 weeks.

### Findings

The estimated treatment effect (in normal curve equivalent units) is 2.37 (*p*-value = 0.14). The estimated treatment effect is not statistically significant at the 0.05 level of significance. Using only second-year data, the estimated treatment effect (in normal curve equivalent units) is –0.44 (*p*-value = 0.87). Neither effect is statistically significant at the 0.05 level.

## I. Algebra I: Larson Algebra I

*Larson Algebra I,* published by Houghton-Mifflin, is designed to supplement the curriculum with extra instruction, practice, and assessments. This product is the same as *Larson Pre-Algebra* but starts at different points within the sequence. It covers whole numbers, fractions, decimals, percents, rational numbers, probability and statistics, coordinate geometry, pre-algebra, and algebra I. The program addresses both skill building and problem solving, and allows the teacher to track student progress. Teachers can choose the quantity and order in which the topics are presented to the students. Recommended time of usage varies according to the number of topics and the number of weeks in the course; however, the developer recommends at least once-weekly usage. The program is designed to be used in computer labs. Teachers receive two hours to one day of pre-implementation training. Ongoing support during the school year is available by phone, e-mail, and through a website. The study estimated the annualized cost per student to be $13. Of that amount, 62 percent is the license fee and the remaining 38 percent is for teacher training and support, technical support, and printed materials and supplies.

More information about the product, its technical requirements, and contact information can be found at http://www.larsonmath.com/lmc_prea/prealgebra.htm.

### Study Design and Context

Across the two years of the study, the study was implemented in 12 schools in 5 districts. All of these districts were located in urban fringe areas. The average district had 91 schools and 68,000 students. Forty-three teachers volunteered to participate in the first or second year. Twenty-four teachers were assigned randomly to use the product and 19 were assigned not to use the product, with at least a pair of treatment and control teachers in each school. Teachers had, on average, 10 years of teaching experience and 63 percent had a master's degree.

Across the two years of the study, fall and spring algebra I test scores were obtained for 1,204 students. The students in the study had on average 35 percent correct answers on the fall test. The average age of the students was 15 years and 51 percent were female. Seven percent were in the eighth grade, 87 percent were in the ninth grade, and 6 percent were in higher grades.

The product includes a database that tracks time students were logged on. During the two years of the study, the average student was logged on for 313 minutes a year (s.d. 380 minutes) and usage occurred during six weeks a year.

In the second year, the study was implemented in eight schools in three districts. All of these districts were located in urban fringe areas. The average district had 90 schools and 63,635 students. Eighteen algebra I teachers volunteered to participate in the second year. Ten teachers were assigned randomly to use the product and eight were assigned not to use the product, with at least a pair of treatment and control teachers in each school. Teachers had on average 13 years of teaching experience and 65 percent had a master's degree.

In the second year, fall and spring algebra I test scores were obtained for 471 students. The students in the study had on average 41 percent correct answers in fall 2005. The average age of the students was 15 years and 51 percent were female, and 9 percent were in eighth grade and 91 percent were in ninth grade. Average student product usage was 297 minutes during the second year and usage occurred during six weeks of the year.

### Findings

The estimated treatment effect (in terms of the percent correct on the exam) is –0.10 (*p*-value = 0.93). Using only second-year data, the estimated treatment effect (in terms of the percent correct on the exam) is 2.59 (*p*-value = 0.15). Neither effect is statistically significant at the 0.05 level.

## J. Algebra I: Cognitive Tutor

Cognitive Tutor Algebra I, published by Carnegie Learning, Inc., is a full curriculum that includes proportional reasoning, solving linear equations and inequalities, solving systems of linear equations, analyzing data, and using polynomial functions, powers, and exponents. The product presents problems in scenarios, asks students to use graphs to represent problems related to the scenarios, and asks the students to use a solver to answer questions related to the scenarios. It also evaluates students' skill levels based on their answers. A textbook accompanies the software. The product provides teachers with reports on student progress and performance. Students use the product in a computer lab two days a week and use the textbook three days a week (the software also can supplement another textbook). Teachers receive four days of initial training on using the product, conducted by a qualified trainer at a school or district location. Support is also provided by phone and e-mail. The study estimates the annualized cost per student to be $69. Of that amount, 43 percent is the license fee and the remaining 57 percent is for teacher training and support, technical support, and printed materials and supplies.

More information about the product and its technical requirements is available at http://www.carnegielearning.com/products_algebraI.cfm.

### Study Design and Context

Across the two years of the study, the study was implemented in 11 schools in 4 districts. Districts were in urban and urban fringe areas, and the average district had 230 schools and 133,000 students. Twenty-nine teachers participated in the study. Fifteen teachers were assigned randomly to use the product and 14 were assigned not to use the product, with at least a pair of treatment and control teachers in each school. Teachers averaged 13 years of teaching experience and 41 percent had a master's degree.

Across the two study years, fall and spring algebra I test scores were obtained for 755 students. The students in the study got 28 percent of questions correct on the fall test. The average age of the students was 15 years and 49 percent were female, and 14 percent were in eighth grade and 86 percent were in ninth grade.

Cognitive Tutor includes a database that tracks time students were logged on. During the two years of the study, the average student was logged on an average of 2,149 minutes a year (s.d. 1087 minutes), and the product was used during 24 weeks on average. Minutes of usage do not include time using the product's textbook for lectures.

In the second year, the study was implemented in nine schools in four districts. Districts were in urban and urban fringe areas and the average district had 230 schools and 133,000 students. Eighteen algebra I teachers participated in the study. Nine teachers were assigned randomly to use the product and nine were assigned not to use the product, with at least a pair of treatment and control teachers in each school. Teachers had on average 16 years of teaching experience and 47 percent had a master's degree. The fall and spring algebra I test was administered to 276 students. The students in the study had on average 28 percent correct answers in fall 2005. The average age of the students was 14 years and 51 percent were female. The average student was logged on for 1,840 minutes in the second year and used the product during 18 weeks. Eighteen percent of students were in the eighth grade and 82 percent were in ninth grade.

### Findings

The estimated treatment effect (in terms of the percent correct on the exam) is –1.28 (*p*-value = 0.26). The estimated treatment effect is not statistically significant at the 0.05 level of significance. Using only second-year data, the estimated treatment effect (in terms of the percent correct on the exam) is –2.10 (*p*-value = 0.30).

## K. Summary of Findings for the 10 Software Products

Tables III.1 and III.2 summarize the context and findings for each of the 10 software products. The tables highlight the main features from the text descriptions and findings separately for the six reading products and the four math products.

The study's main objective was to assess the effects that using software products may have had on reading or math scores on standardized achievement tests. Nine of the 10 products had statistically insignificant effects on test scores for the full sample (two years of student data) and the second-year sample. One product had a positive and statistically significant effect for the full sample. The magnitude of this effect is equivalent to moving the average student from the 50th percentile to the 54th percentile (an effect size of 0.09).

The limitations of the study preclude direct comparisons of product effects in the columns. Because districts and schools volunteered to implement particular products, their characteristics differ and these differences may relate to effectiveness. The study design does not rule out the possibility that a product the study finds to be ineffective could be effective if implemented by other districts or schools. Also, the limited data collection in the second year precluded the study from exploring how teachers may have used products differently in the second year compared to the first, and from exploring how classroom practices and experiences may have differed between products.

**Table III.1. Estimates for Reading Products**

| Destination Reading | Headsprout | Plato Focus | Waterford Early Reading Program | Academy of Reading | LeapTrack |
|---|---|---|---|---|---|
| First Grade | First Grade | First Grade | First Grade | Fourth Grade | Fourth Grade |
| **Product Description** | | | | | |
| Supplemental program for decoding, reading comprehension, and other reading skills. The product studied is course 1, which covers material for students in kindergarten and first grade. | Supplemental reading program for phonemic awareness, phonics, fluency, vocabulary, and comprehension. 80 episodes; the first 40 episodes focus on decoding, segmenting, and blending, the next 40 on vocabulary, reading fluency and comprehension. | Curriculum for phonemic awareness, phonics, fluency, vocabulary, and reading comprehension. | Supplemental reading program; level 2, which was used in the study, includes letter sounds, word recognition, and comprehension. | Supplemental reading product for phonemic awareness and sound-symbol association, phonics and decoding skills, fluency and comprehension, and reading proficiency. | Supplemental reading product for phonemic awareness, phonics, vocabulary, and reading comprehension in addition to other reading skills. The product also includes math, which was not studied here. |
| | Self-paced. | | Self-paced, with student books sent home weekly with directions for parents. | Self-paced. | Self-paced. |
| Teacher training takes two to three days on site, with ongoing support during the school year by phone and e-mail. | Teachers receive one day of initial training and ongoing support during the school year by phone and e-mail. | Teachers receive three to six days of training, a day or more of training during the school year and in-class consultation, in addition to ongoing support by phone and through the product website. | Teacher training is two days on site, with ongoing support available by phone, e-mail, and through a website. | Teachers receive one day of initial training and one day of training four to six weeks later, with ongoing support by phone, e-mail, and webinars. | Teachers receive a day of initial training and up to four days of follow-up training, with ongoing support available by phone and through a website. |
| Annualized cost is estimated to be $78 per student. | Annualized cost estimated to be $146 per student. | Annualized cost estimated to be $351 per student. | Annualized cost estimated to be $223 per student. | Annualized cost estimated to be $217 per student. | Annualized cost estimated to be $154 per student. |
| **Districts, Schools, Teachers, and Students in the Study** | | | | | |
| Overall sample: 2 districts, 12 schools, 35 teachers (21 in the treatment group and 14 in the control group), and 742 students. Teachers averaged 16 years of teaching experience and 43 percent had a master's degree. The average student had reading skills at the 43rd percentile on the fall test. | Overall sample: 3 districts, 12 schools, 63 teachers (32 in the treatment group and 31 in the control group), and 1,079 students. Teachers averaged 11 years of teaching experience and 58 percent had a master's degree. The average student had reading skills in the 65th percentile. | Overall sample: 3 districts, 8 schools, 29 teachers, (15 in the treatment group and 14 in the control group), and 618 students. Teachers averaged 17 years of teaching experience and 55 percent had a master's degree. The average student was reading at the 40th percentile on the fall test. | Overall sample: 3 districts, 13 schools, 46 teachers (28 in the treatment group and 22 in the control group), and 1,155 students. Teachers averaged 11 years of teaching experience and 35 percent had a master's degree. The average student was reading at the 52nd percentile on the fall test. | Overall sample: 4 districts, 15 schools, 41 teachers (22 in the treatment group and 19 in the control group), and 899 students. Teachers averaged 9 years of experience and 32 percent had master's degrees. The average student was reading at the 34th percentile on the fall test. | Overall sample: 4 districts, 19 schools, 55 teachers (29 in the treatment group and 26 in the control group), and 1,274 students. Teachers averaged 11 years of teaching experience and 35 percent had a master's degree. The average student was reading at the 38th percentile on the fall test. |
| Second-year sample: 2 districts, 9 schools, 25 teachers (15 in the treatment group and 10 in the control group), and 453 students. Teachers averaged 13.3 years of teaching experience and 35 percent had a master's degree. The average student had reading skills at the 45th percentile on the fall test. | Second-year sample: 3 districts, 7 schools, 18 teachers (9 in the treatment group and 9 in the control group), and 268 students. Teachers averaged 13.0 years of teaching experience and 71 percent had a master's degree. The average student had reading skills at the 65th percentile on the fall test. | Second-year sample: 3 districts, 8 schools, 18 teachers (9 in the treatment group and 9 in the control group), and 319 students. Teachers averaged 19.5 years of teaching experience and 65 percent had a master's degree. The average student had reading skills at the 42nd percentile on the fall test. | Second-year sample: 3 districts, 9 schools, 20 teachers (11 in the treatment group and 9 in the control group), and 331 students. Teachers averaged 9.2 years of teaching experience and 26 percent had a master's degree. The average student had reading skills at the 54th percentile on the fall test. | Second-year sample: 2 districts, 7 schools, 14 teachers (7 in the treatment group and 7 in the control group), and 282 students. Teachers averaged 18.5 years of teaching experience and 33 percent had a master's degree. The average student had reading skills at the 48th percentile on the fall test. | Second-year sample: 2 districts, 4 schools, 8 teachers (4 in the treatment group and 4 in the control group), and 181 students. Teachers averaged 19.5 years of teaching experience and 65 percent had a master's degree. The average student had reading skills at the 56th percentile on the fall test. |

Table III.1 (*continued*)

| Destination Reading | Headsprout | Plato Focus | Waterford Early Reading Program | Academy of Reading | LeapTrack |
|---|---|---|---|---|---|
| First Grade | First Grade | First Grade | First Grade | Fourth Grade | Fourth Grade |
| **Product Usage** | | | | | |
| During the two years of the study, the average student was logged on to the product 615 minutes a year, and used a product during 25 weeks. | During the two years of the study, the average student was logged on to the product 857 minutes a year, and used the product during 26 weeks a year. | The software did not provide data on usage by student. | During the two years of the study, the average student was logged on for 3,643 minutes a year, and usage occurred during 34 weeks. | During the two years of the study, the average student was logged on 624 minutes a year and usage occurred during 13 weeks. | During the two years of the study, the average student was logged on to the product for 520 minutes a year. The product does not track weeks of usage. |
| In the second year, the average student was logged on to the product 693 minutes and used a product during 26.6 weeks. | In the second year, the average student was logged on to the product 772 minutes and used a product during 22.9 weeks. | | In the second year, the average student was logged on for 2,794 minutes a year, and usage occurred during 35 weeks. | In the second year, the average student was logged on for 951 minutes and usage occurred during 16 weeks. | In the second year, the average student was logged on for 883 minutes. |
| **Product Effects on Test Scores** | | | | | |
| The estimated treatment effect for the full sample (in normal curve equivalent units) is 1.91 (*p*-value = 0.27). | The estimated treatment effect for the full sample (in normal curve equivalent units) is 0.29 (*p*-value = 0.79). | The estimated treatment effect for the full sample (in normal curve equivalent units) is 0.50 (*p*-value = 0.72). | The estimated treatment effect for the full sample (in normal curve equivalent units) is 0.42 (*p*-value = 0.77). | The estimated treatment effect for the full sample (in normal curve equivalent units) is –0.16 (*p*-value = 0.88). | The estimated treatment effect for the full sample (in normal curve equivalent units) is 1.97 (*p*-value = 0.01). |
| Using only second-year data, the estimated treatment effect (in normal curve equivalent units) is 2.19 (*p*-value = 0.31). | Using only second-year data, the estimated treatment effect (in normal curve equivalent units) is –4.13 (*p*-value = 0.06). | Using only second-year data, the estimated treatment effect (in normal curve equivalent units) is –.10 (*p*-value = 0.95). | Using only second-year data, the estimated treatment effect (in normal curve equivalent units) is -1.76 (*p*-value = 0.41). | Using only second-year data, the estimated treatment effect (in normal curve equivalent units) is 1.86 (*p*-value = 0.54). | Using only second-year data, the estimated treatment effect (in normal curve equivalent units) is 2.88 (*p*-value = 0.20). |
| Neither effect is statistically significant at the 0.05 level. | Neither effect is statistically significant at the 0.05 level. | Neither effect is statistically significant at the 0.05 level. | Neither effect is statistically significant at the 0.05 level. | Neither effect is statistically significant at the 0.05 level. | The estimated treatment effect for the full sample is statistically significant at the 0.05 level of significance. |

**Table III.2. Estimates for Math Products**

| Larson Pre-Algebra | Achieve Now | Larson Algebra I | Cognitive Tutor |
|---|---|---|---|
| Sixth Grade | Sixth Grade | Algebra I | Algebra I |
| **Product Description** | | | |
| Supplemental program for skills in whole numbers, fractions, decimals, percents, rational numbers, probability and statistics, coordinate geometry, pre-algebra, and algebra. Same as Larson Algebra I but starts at different points within the sequence. | Supplemental math program covering pre-algebraic topics including rational numbers in related organizational patterns, proportion and percent, integers, probability, statistics, problem solving, geometry, measurement, and the foundational concepts of algebra I. | Supplemental program for skills in whole numbers, fractions, decimals, percents, rational numbers, probability and statistics, coordinate geometry, pre-algebra, and algebra I. Same as Larson Pre-Algebra but starts at different points within the sequence. | Full curriculum that includes proportional reasoning, solving linear equations and inequalities, solving systems of linear equations, analyzing data, and using polynomial functions, powers, and exponents. A textbook accompanies the software. |
| Designed to be used in computer labs. | Self-paced. | Designed to be used in computer labs. | Students use the product in a computer lab two days a week and use the textbook three days a week. |
| Teachers receive two hours to one day of training and ongoing support during the school year by phone, e-mail, and through a website. | Teacher training is done through web-based meetings and on-line self-tutorials, with ongoing support. | Teachers receive two hours to one day of training and ongoing support during the school year by phone, e-mail, and through a website. | Teachers receive four days of initial training on using the product, conducted by a qualified trainer at a school or district location. Support is also provided by phone and e-mail. |
| Annualized cost estimated to be $15 per student. | Annualized cost estimated to be $36 per student. | Annualized cost estimated to be $15 per student. | Annualized cost estimated to be $69 per student. |
| **Districts, Schools, Teachers, and Students in the Study** | | | |
| Overall sample: 5 districts, 13 schools, 39 teachers (24 in the treatment group and 15 in the control group), and 2,588 students. Teachers averaged 11 years of teaching experience and 32 percent had a master's degree. The average student had math skills at the 55th percentile on the fall test. | Overall sample: 3 districts, 13 schools, 39 teachers (21 in the treatment group and 18 in the control group), and 1,037 students. Teachers averaged 11 years of teaching experience and 33 percent had a master's degree. The average student had math skills at the 41st percentile. | Overall sample: 5 districts, 12 schools, 43 teachers (24 in the treatment group, 19 in the control group), and 1,204 students. Teachers averaged 10 years of teaching experience and 63 percent had a master's degree. The average student scored 35 percent correct on the fall test. | Overall sample: 4 districts, 11 schools, 29 teachers (15 in the treatment group and 14 in the control group), and 755 students. Teachers averaged 13 years of teaching experience and 41 percent had a master's degree. The average student scored 28 percent correct on the fall test. |
| Second-year sample: 3 districts, 8 schools, 18 teachers (10 in the treatment group and 8 in the control group), and 386 students. Teachers averaged 12.3 years of teaching experience and 11 percent had a master's degree. The average student had math skills at the 56th percentile on the fall test. | Second-year sample: 3 districts, 8 schools, 18 teachers (10 in the treatment group and 8 in the control group), and 386 students. Teachers averaged 12.3 years of teaching experience and 11 percent had a master's degree. The average student had math skills at the 56th percentile on the fall test. | Second-year sample: 3 districts, 8 schools, 17 teachers (10 in the treatment group and 7 in the control group), and 471 students. Teachers averaged 13.3 years of teaching experience and 65 percent had a master's degree. The average student scored 41 percent correct on the fall algebra I test. | Second-year sample: 4 districts, 9 schools, 18 teachers (9 in the treatment group and 79 in the control group), and 276 students. Teachers averaged 15.6 years of teaching experience and 47 percent had a master's degree. The average student scored 41 percent correct on the fall algebra I test. |
| **Product Usage** | | | |
| During the two years of the study, the average student was logged on to the product for 817 minutes a year and usage occurred during 19 weeks a year. | The product does not provide the usage time by student. | During the two years of the study, the average student was logged on to the product for 313 minutes a year and usage occurred during 6 weeks a year. | During the two years of the study, the average student was logged on to the product 2,149 minutes a year and usage occurred during 24 weeks a year. |
| In the second year, the average student was logged on to the product for 642 minutes a year and usage occurred during 13 weeks a year. | | In the second year, the average student was logged on to the product for 297 minutes a year and usage occurred during 6 weeks a year. | In the second year, the average student was logged on to the product 1,840 minutes a year and usage occurred during 18 weeks a year. |

Table III.2 (*continued*)

| Larson Pre-Algebra | Achieve Now | Larson Algebra | Cognitive Tutor |
|---|---|---|---|
| Sixth Grade | Sixth Grade | Algebra I | Algebra I |
| **Product Effects on Test Scores** | | | |
| The estimated treatment effect for the full sample (in normal curve equivalent units) is 2.37 (*p*-value = 0.14).<br>Using only second-year data, the estimated treatment effect (in normal curve equivalent units) is –0.44 (*p*-value = 0.87).<br>Neither effect is statistically significant at the 0.05 level. | The estimated treatment effect for the full sample (in normal curve equivalent units) is –0.58 (*p*-value = 0.69).<br>Using only second-year data, the estimated treatment effect (in normal curve equivalent units) is –1.59 (*p*-value = 0.72).<br>Neither effect is statistically significant at the 0.05 level. | The estimated treatment effect for the full sample (in terms of the percent correct on the exam) is –0.10 (*p*-value = 0.93).<br>Using only second-year data, the estimated treatment effect (in terms of the percent correct on the exam) is 2.59 (*p*-value = 0.15).<br>Neither effect is statistically significant at the 0.05 level. | The estimated treatment effect for the full sample (in terms of the percent correct on the exam) is –1.28 (*p*-value = 0.26).<br>Using only second year data, the estimated treatment effect (in terms of the percent correct on the exam) is –2.10 (*p*-value = 0.30).<br>Neither effect is statistically significant at the 0.05 level. |

# References

Agodini, R., M. Dynarski, M. Honey, and D. Levin. "The Effectiveness of Educational Software: Issues and Recommendations for the National Study." Report prepared for Institute of Education Sciences, U.S. Department of Education. Princeton, NJ: Mathematica Policy Research, Inc., May 2003.

Autoskill International, Inc. *Academy of Reading 3.2, copyright 2003*. Product information accessed on November 24, 2008 at http://www.autoskill.com/products/reading/index.php.

Blok, H., R. Oostdam, M.E. Otter, and M. Overmaat. "Computer-Assisted Instruction in Support of Beginning Reading Instruction: A Review." *Review of Educational Research,* vol. 72, 2002, pp. 101-30.

Buros Institute of Mental Measurements, *Test Reviews Online.* Lincoln, NE: University of Nebraska-Lincoln, 1998.

Carnegie Learning Corporation, Cognitive Tutor Algebra I, http://www.carnegielearning.com/software_features.cfm (accessed November 20, 2008).

CTB/McGraw Hill. *California Achievement Tests, Edition 6(CAT/6): Norms Book*. Monterey, CA: Author, 2001a.

CTB/McGraw Hill. *California Achievement Tests, Edition 6 (CAT/6): Test Directions for Teachers.* Monterey, CA: Author, 2001b.

Dynarski, M., R. Agodini, S. Heaviside, T. Novak, N. Carey, L. Campuzano. "Effectiveness of Reading and Mathematics Software Products: Findings from the First Student Cohort." Report to Congress. Publication NCEE 2007-4005. Washington, DC: U.S. Department of Education, March 2007.

Educational Testing Service. *End-of-Course Algebra Assessment Administrator's Manual.* Princeton, NJ: ETS, 1997.

Headsprout. *Headsprout Early Reading.* Product information accessed on November 24, 2008 at http://www.headsprout.com.

Houghton-Mifflin. *Larson's Pre-Algebra Version 2.1, copyright 2004.* Product information accessed on November 24, 2008 at http://www.larsonmath.com.

Houghton-Mifflin. *Larson's Algebra1 Version 1.1, copyright 2004.* Product information accessed on November 24, 2008 at http://www.larsonmath.com.

Kulik, C.C., and J.A. Kulik. "Effectiveness of Computer-Based Instruction: An Updated Analysis." *Computers in Human Behavior,* vol. 7, 1991, pp. 75-94.

Kulik, J.A. *Effects of Using Instructional Technology in Elementary and Secondary Schools: What Controlled Evaluation Studies Say.* Arlington, VA: SRI International, 2003.

Kulik, J.A. "Meta-analytic Studies of Findings on Computer-Based Instruction." In *Technology Assessment in Education and Training*, edited by E.L. Baker and H.F. O'Neil Jr. Hillsdale, NJ: Lawrence Erlbaum, 1994, pp. 9-33.

LeapFrog SchoolHouse, *LeapTrack, Version 3,* Product information accessed on November 24, 2008 at http://www.leapfrogschoolhouse.com.

Murphy, R., W. Penuel, B. Means, C. Korbak, and A. Whaley. *E-DESK: A Review of Recent Evidence on the Effectiveness of Discrete Educational Software.* Menlo Park, CA: SRI International, 2001.

Pearson, P.D., R.E. Ferdig, R.L. Blomeyer Jr., and J. Moran. *The Effects of Technology on Reading Performance in the Middle-School Grades: A Meta-analysis with Recommendations for Policy.* Naperville, IL: Learning Point Associates, 2005.

Pearson Education, Inc. *New Mexico Standards-Based Assessment (NMSBA): Direction for Administration (DFA).* San Antonio, TX: Author, 2006.

Pearson Education, Inc. *Stanford Early School Achievement Test (SESAT), Level 2: Directions for Administering (DFA).* San Antonio, TX: Author, 1996a

Pearson Education, Inc. *Stanford Achievement Test, Ninth Edition (SAT-9), Primary 1: Directions for Administering (DFA).* San Antonio, TX: Author, 1996b.

Pearson Education, Inc. *Stanford Early School Achievement Test (SESAT), Level 2: Multilevel Norms Booklet.* San Antonio, TX: Author, 1996c.

Pearson Education, Inc. *Stanford Achievement Test, Ninth Edition (SAT-9), Primary 1: Multilevel Norms Booklet.* San Antonio, TX: Author, 1996d.

Pearson Education, Inc. *Stanford Achievement Test, Tenth Edition (SAT-10), Primary 3: Directions for Administering (DFA).* San Antonio, TX: Author, 2003a.

Pearson Education, Inc. *Stanford Achievement Test, Tenth Edition (SAT-10), Intermediate 1: Directions for Administering (DFA).* San Antonio, TX: Author, 2003b.

Pearson Education, Inc. *Stanford Achievement Test, Tenth Edition (SAT-10), Intermediate 2: Directions for Administering (DFA).* San Antonio, TX: Author, 2003c.

Pearson Education, Inc. *Stanford Achievement Test, Tenth Edition (SAT-10), Intermediate 3: Directions for Administering (DFA).* San Antonio, TX: Author, 2003d.

Pearson Education, Inc. *Stanford Achievement Test, Tenth Edition (SAT-10), Primary 3: Multilevel Norms Booklet.* San Antonio, TX: Author, 2003e.

Pearson Education, Inc. *Stanford Achievement Test, Tenth Edition (SAT-10), Intermediate 1: Multilevel Norms Booklet.* San Antonio, TX: Author, 2003f.

Pearson Education, Inc. *Stanford Achievement Test, Tenth Edition (SAT-10), Intermediate 2: Multilevel Norms Booklet.* San Antonio, TX: Author, 2003g.

Pearson Education, Inc. *Stanford Achievement Test, Tenth Edition (SAT-10), Intermediate 3: Multilevel Norms Booklet.* San Antonio, TX: Author, 2003h.

Pearson Education, Inc. *Waterford Early Reading Program* Version 3.x, copyright 2005. Product information accessed on November 24, 2008 at http://www.pearondigital.com/waterford/.

PLATO Learning Corporation. *PLATO Focus Reading and Language Program Version 1.3,* copyright 2003 Product information accessed on November 24, 2008 at http://www.plato.com/Products/PLATO-Focus-Reading-and-Language-Program.aspx.

PLATO Learning Corporation., PLATO Achieve Now Mathematics Series 3, Product information accessed on November 24, 2008 at http://www.plato.com/Elementary-Solutions/Elementary-Mathematics/PLATO-Achieve-Now-Mathematics.aspx.

Riverdeep. *Destination Reading Course 1.* Product information accessed on November 24, 2008 at http://hmlt.hmco.com/DR-EL.php.

Riverside. *Iowa Tests of Basic Skills (ITBS): Directions for Administration.* Chicago: Author, 2001.

Schacter, J. The Impact of Educational Technology on Student Achievement: What the Most Current Research Has to Say. Santa Monica, CA: Milken Exchange on Education Technology, 2001.

Sivin-Kachala, J. *Report on the Effectiveness of Technology in Schools, 1990-1997.* Washington, DC: Software Publishers Association, 1998.

Torgesen, J., R. Wagner, and C. Rashotte. *Test of Word Reading Efficiency: Examiner's Manual.* Austin, TX: Pro-Ed, Inc., 1999.

Waxman, H.C., M-F Lin, and G.M. Michko. *A Meta-analysis of the Effectiveness of Teaching and Learning with Technology on Student Outcomes*. Naperville, IL: Learning Point Associates, 2003.

# Appendix A

# Second-Year Data Collection and Response Rates

This appendix describes the study's data collection approach in the second year and provides more detail about response rates.

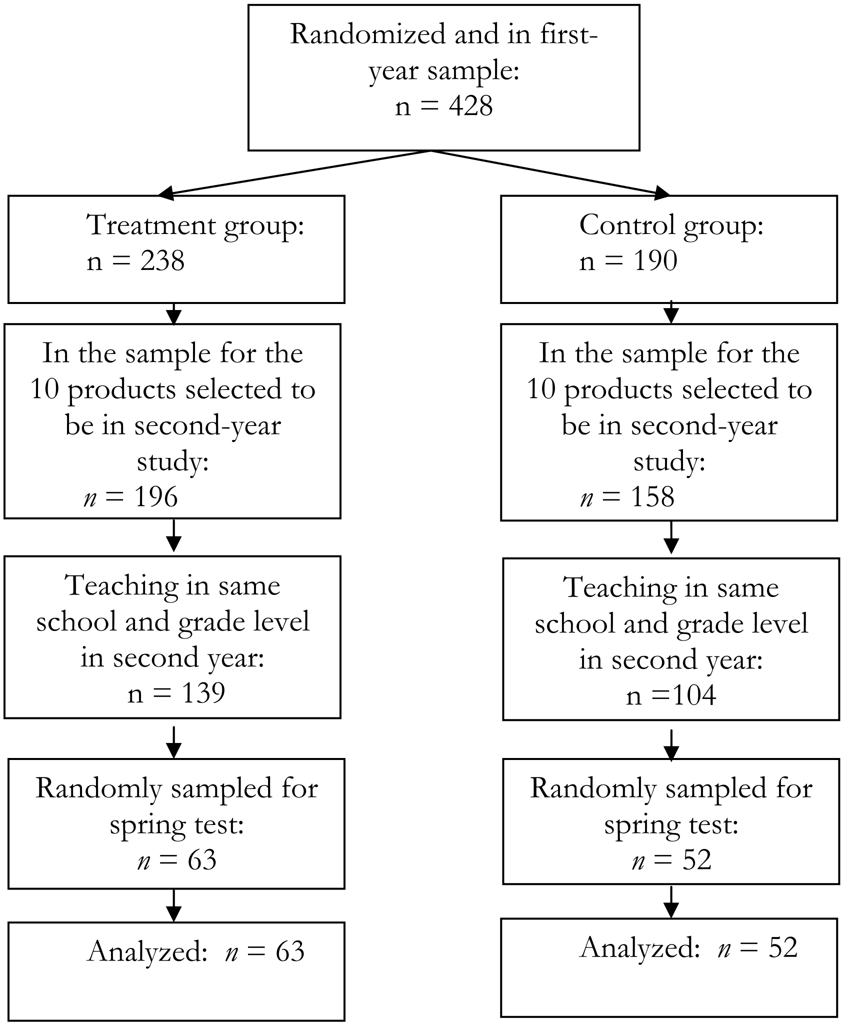
The study's data collection is based on the framework established in the study's first year. During this time, teachers who volunteered to participate in the study were randomly assigned to treatment or control groups. However, not all teachers who had participated in the first year were part of the second year study, due to attrition and mobility. Moreover, products that had been implemented only in a few schools and for which detecting a product effect was unlikely because of low statistical power were not included in the second year. The study team also added some schools and teachers to increase sample sizes for some products that were on the margin of adequate statistical power. Teachers new to the study were randomly assigned to the treatment or control groups as was done in the first year.

To reduce costs, the study tested fewer classrooms in spring 2006 than in fall 2005. Schools that had one treatment and one control teacher were tested. For schools that had more than one treatment or control teacher, one treatment teacher and one control teacher were randomly sampled from the groups. For example, if a school had three treatment and two control teachers, one of the three treatment teachers was sampled and one of the two control teachers was sampled. The sampling probability was set such that one teacher was sampled from the treatment or control groups. For example, if three teachers were in the treatment group, the sampling probability for a treatment teacher was 33 percent. An additional cost modification in the second year was that for some districts that administered their own nationally normed test, the study collected scores for that test from district records rather than conduct its own test.

## A. Teacher Samples

Chapter II examined product effects after teachers had a year of experience using products. Figure A.1 shows the components of the teacher sample that were used in that analysis.

**Figure A.1. Teacher Sample for Experience Effects (Chapter II)**


Randomized and in first-year sample:
n = 428
Treatment group:
n = 238
Control group:
n = 190
In the sample for the 10 products selected to be in second-year study:
*n* = 196
In the sample for the 10 products selected to be in second-year study:
*n* = 158
Teaching in same school and grade level in second year:
n = 139
Teaching in same school and grade level in second year:
n =104
Randomly sampled for spring test:
*n* = 63
Randomly sampled for spring test:
*n* = 52
Analyzed: *n* = 63
Analyzed: *n* = 52

Three aspects of the design determined the teacher sample for the analysis in Chapter II. First, of the 428 teachers in the first year of the study, selecting the 10 products for the second year left 354 teachers. Of that number, mobility to other schools and grade levels left 243 teachers. Randomly sampling teachers left 63 treatment group teachers and 52 control group teachers, which is the analysis sample used to study the effects of a second year of teaching experience using software products on student test scores presented in Chapter II. For the sample of teachers used for the analysis of individual products presented in Chapter III, see Appendix B.

For the study of individual product effects in Chapter III, the flow of teachers consists of teachers who were in the sample only the first year, only in the second year, and in both years. Figure A.2 shows the treatment and control group samples for the three components of the teacher sample. The largest of the three components, almost 60 percent of the total, is the sample of teachers who were only in the first year.

**Figure A.2. Teacher Sample for Individual Product Effects (Chapter III)**

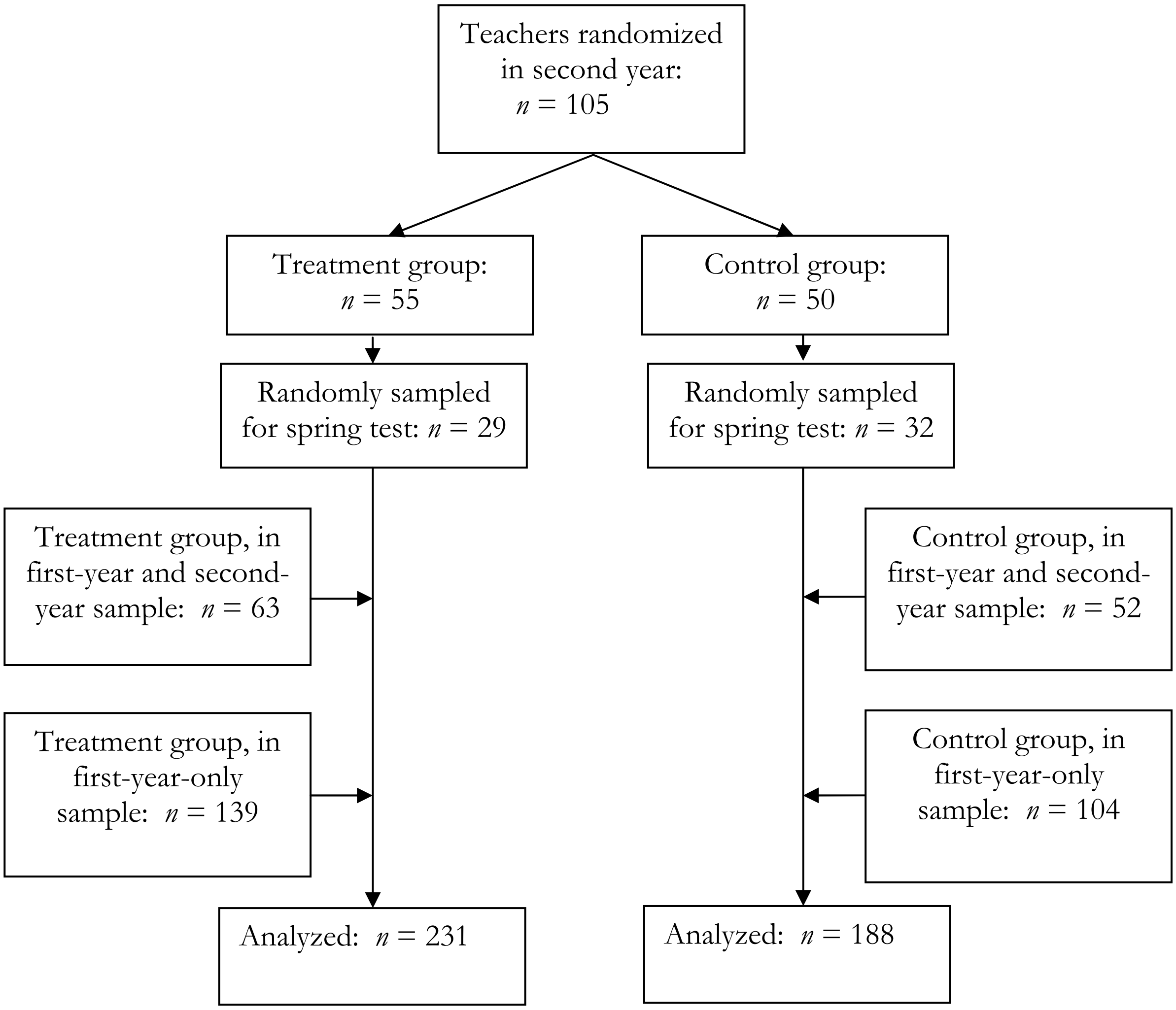

Teachers randomized in second year: *n* = 105
Treatment group: *n* = 55
Control group: *n* = 50
Randomly sampled for spring test: *n* = 29
Randomly sampled for spring test: *n* = 32
Treatment group, in first-year and second-year sample: *n* = 63
Control group, in first-year and second-year sample: *n* = 52
Treatment group, in first-year-only sample: *n* = 139
Control group, in first-year-only sample: *n* = 104
Analyzed: *n* = 231
Analyzed: *n* = 188

Table A.1 shows the breakdown of teachers in the second year by product and by whether the teachers also were included in the first year.

## B. Teacher Survey

In November 2005, teacher questionnaires were mailed to schools for those teachers new to the study in the second year and teachers who had not completed a questionnaire in the first year. Ultimately, 97 percent of teachers completed a questionnaire. Completion rates ranged from 91 percent of fourth grade teachers to 100 percent of sixth grade teachers.

**Table A.1. Teacher Sample Sizes, by Product**

| | All | | | Treatment | | | Control | | |
|---|---|---|---|---|---|---|---|---|---|
| | Total | Year 2 and Year 1 | Year 2 Only | Total | Year 2 and Year 1 | Year 2 Only | Total | Year 2 and Year 1 | Year 2 Only |
| Total | 176 | 115 | 61 | 92 | 63 | 29 | 84 | 52 | 32 |
| First Grade: Destination Reading | 25 | 8 | 17 | 15 | 5 | 10 | 10 | 3 | 7 |
| First Grade: Headsprout | 18 | 9 | 9 | 9 | 5 | 4 | 9 | 4 | 5 |
| First Grade: Plato Focus | 18 | 6 | 12 | 9 | 3 | 6 | 9 | 3 | 6 |
| First Grade: Waterford Early Reading | 20 | 20 | 0 | 11 | 11 | 0 | 9 | 9 | 0 |
| Fourth Grade: Academy of Reading | 14 | 5 | 9 | 7 | 3 | 4 | 7 | 2 | 5 |
| Fourth Grade: LeapTrack | 8 | 8 | 0 | 4 | 4 | 0 | 4 | 4 | 0 |
| Sixth Grade: Achieve Now | 20 | 18 | 2 | 9 | 8 | 1 | 11 | 10 | 1 |
| Sixth Grade: Larson Pre-Algebra | 18 | 17 | 1 | 10 | 10 | 0 | 8 | 7 | 1 |
| Algebra I: Cognitive Tutor | 18 | 12 | 6 | 9 | 8 | 1 | 9 | 4 | 5 |
| Algebra I: Larson Algebra I | 17 | 12 | 5 | 9 | 6 | 3 | 8 | 6 | 2 |

**Table A.2 Teachers Completing the Teacher Survey, Second Year**

| | Teachers | | |
|---|---|---|---|
| | Total | Number Completing Survey | Percentage |
| Total | 264 | 255 | 97 |
| First Grade | 112 | 109 | 97 |
| Fourth Grade | 57 | 52 | 91 |
| Sixth Grade | 47 | 47 | 100 |
| Algebra I | 48 | 47 | 98 |

## C. Student Data Collection

The two criteria for testing students in the fall were: (1) parental consent was received, and (2) students did not have barriers to testing (disability or language issues). For the spring test, classrooms randomly selected for testing included students who had been tested in the fall as well as students who had entered study classrooms after the baseline test was administered. To reduce costs, the study team did not test students in districts that could provide nationally normed standardized test score data.

### Student Sample in the Second Year

Table A.3 shows students by classroom assignment status, as well as the breakdown of treatment and control groups by product. The table corresponds to the sample of students who participated in the study in the second year.

**Table A.3. Eligible Student Sample by Assignment and Grade, Second Year**

| | Eligible Sample | | In Treatment Classrooms | | In Control Classrooms | |
|---|---|---|---|---|---|---|
| | Students | Teachers | Students | Teachers | Students | Teachers |
| Total | 3,884 | 176 | 2,111 | 92 | 1,773 | 84 |
| **First Grade** | 1,460 | 81 | 804 | 44 | 656 | 37 |
| Destination Reading | 465 | 25 | 277 | 15 | 188 | 10 |
| Headsprout | 284 | 18 | 150 | 9 | 134 | 9 |
| Plato Focus | 329 | 18 | 164 | 9 | 165 | 9 |
| Waterford Early Reading Program | 382 | 20 | 213 | 11 | 169 | 9 |
| **Fourth Grade** | 581 | 22 | 305 | 11 | 276 | 11 |
| Academy of Reading | 319 | 14 | 159 | 7 | 160 | 7 |
| LeapTrack | 262 | 8 | 146 | 4 | 116 | 4 |
| **Sixth Grade** | 899 | 38 | 490 | 19 | 409 | 19 |
| Achieve Now | 400 | 20 | 186 | 9 | 214 | 11 |
| Larson Pre-Algebra | 499 | 18 | 304 | 10 | 195 | 8 |
| **Algebra I** | 944 | 35 | 512 | 18 | 432 | 17 |
| Cognitive Tutor | 381 | 18 | 203 | 9 | 178 | 9 |
| Larson Algebra I | 563 | 17 | 309 | 9 | 254 | 8 |

## Student Tests

To conserve resources, in the second year the study only administered tests in districts where the district did not administer a standardized normed test as part of their assessments. In districts where standardized tests were available, those scores were used as fall or spring scores by the study team. For first grade, one district provided scores on the Iowa Tests of Basic Skills administered in October of 2005, which were used as fall scores. Another district provided scores on the Stanford Achievement Test, tenth edition, administered in March of 2006, which were used as spring test scores. For fourth grade, one district provided scores on the Iowa Tests of Basic Skills administered in October of 2005 and another provided scores on the California Achievement Test, sixth edition, administered in March of 2005 in the previous grade and school year. Scores from both districts were used as fall test scores. For sixth grade, one district provided scores on the Iowa Tests of Basic Skills administered in October of 2005 and another provided scores on the New Mexico Standards Based Assessment administered in March of 2005 in the previous grade and school year. Scores from both districts were used as fall test scores. Furthermore, one district provided scores on the New Mexico Standards Based Assessment administered in March of 2006, which

**Figure A.3. Achievement Tests Administered by the Study or Provided by Districts**

| | **Fall 2005 Test** | **Spring 2006 Test** |
|---|---|---|
| First Grade | Stanford Early School Achievement Test (SESAT 2, Form S)<br>One district provided Iowa Tests of Basic Skills (ITBS) scores | Stanford Achievement Test, Abbreviated Primary 1, Ninth Edition, Form S (SAT-9)<br>One district provided Stanford Achievement Test, Tenth Edition (SAT-10) scores |
| Fourth Grade | Stanford Achievement Test Abbreviated Battery Primary 3, Tenth Edition (SAT-10)<br>One district provided Iowa Tests of Basic Skills (ITBS) scores<br>One district provided California Achievement Test, Sixth Edition (CAT/6) scores | Stanford Achievement Test Abbreviated Battery Intermediate 1, Tenth Edition (SAT-10) |
| Sixth Grade | Stanford Achievement Test Abbreviated Battery Intermediate 2, Tenth Edition (SAT-10)<br>One district provided Iowa Tests of Basic Skills (ITBS) scores<br>One district provided New Mexico Standards Based Assessment (NMSBA) scores | Stanford Achievement Test Abbreviated Battery Intermediate 3, Tenth Edition (SAT-10)<br>One district provided New Mexico Standards Based Assessment (NMSBA) scores |
| Algebra 1 | Educational Testing Service End-of-Course Algebra Test (ETS)<br>One district provided Iowa Tests of Basic Skills (ITBS) scores | Educational Testing Service End-of-Course Algebra Test (ETS) |

were used as spring scores. For algebra I, one district provided scores on the Iowa Tests of Basic Skills administered in October of 2005, which were used as fall scores.

The study team administered tests during regular class periods in the fall and spring. Tests were normally administered two to three weeks after the start of the school year and four to six weeks before the end of the school year. In the fall, the testing response rate averaged 88 percent for treatment classrooms and ranged from 75 percent in algebra I to 98 percent in first grade. In the spring, the testing response rate averaged 83 percent for treatment classrooms and ranged from 75 percent in sixth grade to 94 percent in first grade. In the spring, the study tested 1,760 students and districts provided scores for 484 students (see bottom of Table A.4). Figure A.3 lists the tests the study administered and tests that districts provided.

**Table A.4. Number of Students and Percentage Tested in Fall and Spring, 2005-2006 School Year**

| | Eligible Students in Treatment Classrooms | Eligible Students in Control Classrooms | Eligible Students in Treatment Classrooms Tested by Study | Eligible Students in Treatment Classrooms Tested by District | Eligible Students in Control Classrooms Tested by Study | Eligible Students in Control Classrooms Tested by District | Response Rate, Treatment Classrooms | Response Rate, Control Classrooms |
|---|---|---|---|---|---|---|---|---|
| *First Grade* | | | | | | | | |
| Fall | 804 | 656 | 753 | 38 | 600 | 33 | 98% | 96% |
| Spring | 804 | 656 | 531 | 223 | 461 | 156 | 94% | 94% |
| *Fourth Grade* | | | | | | | | |
| Fall | 305 | 276 | 145 | 98 | 138 | 104 | 80% | 88% |
| Spring | 305 | 276 | 232 | 0 | 231 | 0 | 76% | 84% |
| *Sixth Grade* | | | | | | | | |
| Fall | 490 | 409 | 356 | 94 | 245 | 119 | 92% | 89% |
| Spring | 490 | 409 | 325 | 42 | 269 | 63 | 75% | 81% |
| *Algebra I* | | | | | | | | |
| Fall | 512 | 432 | 345 | 39 | 302 | 19 | 75% | 79% |
| Spring | 512 | 432 | 407 | 0 | 340 | 0 | 79% | 79% |
| *Total* | | | | | | | | |
| Fall | 2,111 | 1,773 | 1,599 | 269 | 1,285 | 275 | 88% | 88% |
| Spring | 2,111 | 1,773 | 1,495 | 265 | 1,301 | 219 | 83% | 88% |

Table A.5 presents sample sizes by product. Student attrition rates reported in the table are calculated by dividing students with a spring 2006 test score by the number of eligible students for whom test scores could have been provided. The first grade sample has the lowest attrition rate, at 6.1 percent, and sixth grade had the highest attrition rate, at 22.2 percent.

**Table A.5. Student Attrition Rates in the Second Year**

| | All | | | Students in Treatment Group Classrooms | | | Students in Control Group Classrooms | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | N | Percentage of Eligible Students | Attrition Rate | N | Percentage of Eligible Students | Attrition Rate | N | Percentage of Eligible Students | Attrition Rate | Differential Attrition Rate |
| **First Grade** | 1,371 | 93.9 | 6.1 | 754 | 93.8 | 6.2 | 617 | 94.1 | 6.0 | 0.3 |
| Destination Reading | 453 | 97.4 | 2.6 | 269 | 97.1 | 2.9 | 184 | 97.9 | 2.1 | 0.8 |
| Headsprout | 268 | 94.4 | 5.6 | 145 | 96.7 | 3.3 | 123 | 91.8 | 8.2 | -4.9 |
| Plato Focus | 319 | 97.0 | 3.0 | 159 | 97.0 | 3.1 | 160 | 97.0 | 3.0 | 0.1 |
| Waterford | 331 | 86.7 | 13.4 | 181 | 85.0 | 15.0 | 150 | 88.8 | 11.2 | 3.8 |
| **Fourth Grade** | 463 | 79.7 | 20.3 | 232 | 76.1 | 23.9 | 231 | 83.7 | 16.3 | 7.6 |
| Academy of Reading | 282 | 88.4 | 11.6 | 136 | 85.5 | 14.5 | 146 | 91.3 | 8.8 | 5.7 |
| LeapTrack | 181 | 69.1 | 30.9 | 96 | 65.8 | 34.2 | 85 | 73.3 | 26.7 | 7.5 |
| **Sixth Grade** | 699 | 77.8 | 22.2 | 367 | 74.9 | 25.1 | 332 | 81.2 | 18.8 | 6.3 |
| Achieve Now | 313 | 78.3 | 21.8 | 145 | 78.0 | 22.0 | 168 | 78.5 | 21.5 | 0.5 |
| Larson Pre-Algebra | 386 | 77.4 | 22.6 | 222 | 73.0 | 27.0 | 164 | 84.1 | 15.9 | 11.1 |
| **Algebra I** | 747 | 79.1 | 20.9 | 407 | 79.5 | 20.5 | 340 | 78.7 | 21.3 | -0.8 |
| Cognitive Tutor | 276 | 72.4 | 27.6 | 145 | 71.4 | 28.6 | 131 | 73.6 | 26.4 | 2.2 |
| Larson Algebra I | 471 | 83.7 | 16.3 | 262 | 84.8 | 15.2 | 209 | 82.3 | 17.7 | -2.5 |

## Imputing Missing Data

Some students did not take all tests or subtests and some districts did not provide test scores or other data. The largest number of missing tests occurred for the algebra I pre-test. The study imputed about 30 percent of fall 2005 scores. In first grade, approximately 5 percent of test scores were imputed. In fourth and sixth grades, one percent of spring test scores and 3 to 4 percent of fall test scores were imputed. Components of the test scores and student age and gender were imputed using the Markov Chain Monte Carlo (MCMC) method in SAS 9. The imputation was done five times separately for students in treatment and control classrooms. The HLM estimation procedure used by the study used the five imputed data sets and calculated variances of the estimates that incorporated the added variance from the imputation. As noted in the first year report (Dynarski et al. 2007, p. 88), the imputation method was tested in the first year by setting random samples of data to "missing," and calculating correlations between imputed scores and actual scores. The correlations were high, in the range of 90 percent to 95 percent for different samples, indicating that the MCMC method successfully imputed scores that were close to the actual scores.

# Appendix B

# Description of Sample for the 10 Products

For the analysis of individual product effects, the study focused on the set of products for which data were collected in the second year of the study. The analysis sample includes all students, teachers, and schools that participated in the study in the first or second year of the study, restricting to those schools that used one of the 10 products for which data were collected in the second year.

The final sample includes 127 schools in 29 school districts that participated in the first or second year of the study and that used any of the 10 products for which data were collected in the second year. The sample includes 419 teachers, 231 assigned to the treatment group and 188 assigned to the control group. Table B.1 shows final counts of teachers in the sample by assignment status, by year of participation, and by product.

Table B.2 shows final counts of students by classroom assignment status, as well as the breakdown of treatment and control groups by product. The table corresponds to the full sample of students used for estimations of individual product effects on test scores.

Tables B.3a-d show means and standard deviations for all data items used in the estimation models. Some data items are defined only for treatment classrooms, and school characteristics are the same for treatment and control classrooms.

**Table B.1. Sample of Teachers, by Product**

| | Number of Teachers Participating | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | All | | | | Treatment | | | | Control | | | |
| | Total | Only Year 1 | Year 2 and Year 1 | Only Year 2 | Total | Only Year 1 | Year 2 and Year 1 | Only Year 2 | Total | Only Year 1 | Year 2 and Year 1 | Only Year 2 |
| Total | 419 | 243 | 115 | 61 | 231 | 139 | 63 | 29 | 188 | 104 | 52 | 32 |
| First Grade: Destination Reading | 35 | 10 | 8 | 17 | 21 | 6 | 5 | 10 | 14 | 4 | 3 | 7 |
| First Grade: Headsprout | 63 | 45 | 9 | 9 | 32 | 23 | 5 | 4 | 31 | 22 | 4 | 5 |
| First Grade: Plato Focus | 29 | 11 | 6 | 12 | 15 | 6 | 3 | 6 | 14 | 5 | 3 | 6 |
| First Grade: Waterford Early Reading Program | 46 | 26 | 20 | 0 | 28 | 17 | 11 | 0 | 18 | 9 | 9 | 0 |
| Fourth Grade: Academy of Reading | 41 | 27 | 5 | 9 | 22 | 15 | 3 | 4 | 19 | 12 | 2 | 5 |
| Fourth Grade: LeapTrack | 55 | 47 | 8 | 0 | 29 | 25 | 4 | 0 | 26 | 22 | 4 | 0 |
| Sixth Grade: Achieve Now | 39 | 19 | 18 | 2 | 21 | 12 | 8 | 1 | 18 | 7 | 10 | 1 |
| Sixth Grade: Larson Pre-Algebra | 39 | 21 | 17 | 1 | 24 | 14 | 10 | 0 | 15 | 7 | 7 | 1 |
| Algebra I: Cognitive Tutor | 29 | 11 | 12 | 6 | 15 | 6 | 8 | 1 | 14 | 5 | 4 | 5 |
| Algebra I: Larson Algebra | 43 | 26 | 12 | 5 | 24 | 15 | 6 | 3 | 19 | 11 | 6 | 2 |

**Table B.2. Sample of Students, by Product**

| | Number of Students Participating | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | All | | | Treatment | | | Control | | |
| | Total | Year 1 | Year 2 | Total | Year 1 | Year 2 | Total | Year 1 | Year 2 |
| Total | 11,351 | 8,071 | 3,280 | 6,423 | 4,663 | 1,760 | 4,928 | 3,408 | 1,520 |
| First Grade: Destination Reading | 742 | 289 | 453 | 448 | 179 | 269 | 294 | 110 | 184 |
| First Grade: Headsprout | 1,079 | 811 | 268 | 574 | 429 | 145 | 505 | 382 | 123 |
| First Grade: Plato Focus | 618 | 299 | 319 | 327 | 168 | 159 | 291 | 131 | 160 |
| First Grade: Waterford Early Reading Program | 1,155 | 824 | 331 | 689 | 508 | 181 | 466 | 316 | 150 |
| Fourth Grade: Academy of Reading | 899 | 617 | 282 | 495 | 359 | 136 | 404 | 258 | 146 |
| Fourth Grade: LeapTrack | 1,274 | 1,093 | 181 | 665 | 569 | 96 | 609 | 524 | 85 |
| Sixth Grade: Achieve Now | 1,037 | 724 | 313 | 547 | 402 | 145 | 490 | 322 | 168 |
| Sixth Grade: Larson Pre-Algebra | 2,588 | 2,202 | 386 | 1,590 | 1,368 | 222 | 998 | 834 | 164 |
| Algebra I: Cognitive Tutor | 755 | 479 | 276 | 440 | 295 | 145 | 315 | 184 | 131 |
| Algebra I: Larson Algebra I | 1,204 | 733 | 471 | 648 | 386 | 262 | 556 | 347 | 209 |

**Table B.3a. First Grade, Descriptive Statistics (means with standard deviations in parentheses)**

| | First Grade—All Products | | | First Grade—Destination Reading | | | First Grade—Headsprout | | | First Grade—Plato Focus | | | First Grade—Waterford Reading | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | All | Treatment | Control | All | Treatment | Control | All | Treatment | Control | All | Treatment | Control | All | Treatment | Control |
| **Student** | | | | | | | | | | | | | | | |
| Student is female | 49.08 | 48.87 | 49.36 | 48.38 | 47.54 | 49.66 | 48.47 | 48.43 | 48.51 | 52.43 | 52.6 | 52.23 | 48.31 | 48.33 | 48.28 |
| | (50.00) | (50.00) | (50.01) | (50.01) | (50.00) | (50.08) | (50.00) | (50.02) | (50.03) | (49.98) | (50.01) | (50.04) | (49.99) | (50.01) | (50.02) |
| Student's age | 6.64 | 6.63 | 6.65 | 6.68 | 6.68 | 6.67 | 6.67 | 6.64 | 6.69 | 6.63 | 6.61 | 6.66 | 6.60 | 6.60 | 6.59 |
| | (0.41) | (0.39) | (0.42) | (0.4) | (0.41) | (0.39) | (0.45) | (0.41) | (0.49) | (0.38) | (0.36) | (0.39) | (0.37) | (0.37) | (0.37) |
| Fall test total NCE | 50.99 | 50.67 | 51.39 | 46.22 | 46.82 | 45.31 | 58.15 | 56.99 | 59.47 | 44.46 | 44.44 | 44.48 | 50.85 | 50.88 | 50.8 |
| | (20.53) | (20.9) | (20.03) | (18.66) | (19.18) | (17.84) | (20.47) | (20.85) | (19.96) | (20.13) | (20.66) | (19.56) | (19.87) | (20.7) | (18.59) |
| Spring test total NCE | 51.87 | 51.78 | 51.98 | 50.15 | 50.82 | 49.13 | 56.11 | 55.24 | 57.10 | 50.8 | 51.15 | 50.40 | 49.58 | 49.83 | 49.21 |
| | (19.11) | (19.36) | (18.78) | (17.94) | (17.88) | (18.01) | (20.10) | (20.63) | (19.45) | (18.38) | (18.74) | (17.98) | (18.66) | (19.16) | (17.92) |
| **Sample Size** | **3,594** | **2,038** | **1,556** | **742** | **448** | **294** | **1,079** | **574** | **505** | **618** | **327** | **291** | **1,155** | **689** | **466** |
| **Teacher** | | | | | | | | | | | | | | | |
| Teacher is female | 0.99 | 0.99 | 1.00 | 0.97 | 0.95 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | (0.08) | (0.10) | (0.00) | (0.17) | (0.22) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| Teaching experience | 12.86 | 12.89 | 12.81 | 16.21 | 17.59 | 14.14 | 10.52 | 9.56 | 11.51 | 16.57 | 16.02 | 17.17 | 11.17 | 11.51 | 10.62 |
| | (9.74) | (9.79) | (9.74) | (11.14) | (10.24) | (12.47) | (8.04) | (8.04) | (8.05) | (10.29) | (11.83) | (8.74) | (9.26) | (8.67) | (10.34) |
| Teacher has a master's degree | 48.36 | 42.19 | 56.06 | 42.86 | 35.71 | 53.57 | 58.20 | 50.00 | 66.67 | 55.17 | 46.67 | 64.29 | 34.78 | 35.71 | 33.33 |
| | (49.70) | (49.38) | (49.33) | (48.72) | (47.81) | (49.86) | (49.36) | (50.8) | (47.14) | (50.61) | (51.64) | (49.72) | (48.15) | (48.80) | (48.51) |
| **Sample Size** | **173** | **96** | **77** | **35** | **21** | **14** | **63** | **32** | **31** | **29** | **15** | **14** | **46** | **28** | **18** |
| **School** | | | | | | | | | | | | | | | |
| Percentage scoring below fall test 33rd percentile | 33.29 | | | 34.71 | | | 22.73 | | | 49.92 | | | 31.47 | | |
| | (18.80) | | | (19.84) | | | (19.1) | | | (16.36) | | | (11.99) | | |
| Percentage scoring below spring test 33rd percentile | 29.16 | | | 28.26 | | | 23.47 | | | 31.30 | | | 33.93 | | |
| | (13.17) | | | (17.3) | | | (13.98) | | | (10.92) | | | (7.18) | | |
| Percentage receiving | 50.29 | | | 71.06 | | | 34.46 | | | 47.64 | | | 47.35 | | |

Table B.3a *(continued)*

| | First Grade—All Products | | | First Grade—Destination Reading | | | First Grade—Headsprout | | | First Grade—Plato Focus | | | First Grade—Waterford Reading | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | All | Treatment | Control | All | Treatment | Control | All | Treatment | Control | All | Treatment | Control | All | Treatment | Control |
| free/reduced-price lunch | | | | | | | | | | | | | | | |
| | (27.79) | | | (14.48) | | | (21.99) | | | (19.92) | | | (35.62) | | |
| Student/teacher ratio | 16.2 | | | 18.95 | | | 14.53 | | | 15.79 | | | 15.44 | | |
| | (2.75) | | | (2.75) | | | (1.40) | | | (2.24) | | | (2.25) | | |
| Percentage of Hispanic students | 20.07 | | | 34.35 | | | 5.83 | | | 27.29 | | | 15.60 | | |
| | (20.32) | | | (25.26) | | | (9.28) | | | (12.78) | | | (17.10) | | |
| Percentage of black students | 23.49 | | | 31.49 | | | 13.47 | | | 5.32 | | | 36.55 | | |
| | (26.58) | | | (18.67) | | | (12.96) | | | (3.75) | | | (39.25) | | |
| Urban | 53.33 | | | 83.33 | | | 50.00 | | | 75.00 | | | 15.38 | | |
| | (50.45) | | | (38.92) | | | (52.22) | | | (46.29) | | | (37.55) | | |
| **Sample Size** | **45** | | | **12** | | | **12** | | | **8** | | | **13** | | |

**Table B.3b. Fourth Grade, Descriptive Statistics (means with standard deviations in parentheses)**

| | Fourth Grade—Total | | | Fourth Grade—Academy of Reading | | | Fourth Grade—LeapTrack | | |
|---|---|---|---|---|---|---|---|---|---|
| | All | Treatment | Control | All | Treatment | Control | All | Treatment | Control |
| **Student** | | | | | | | | | |
| Student is female | 49.95 | 47.99 | 52.16 | 49.72 | 47.68 | 52.23 | 50.55 | 48.87 | 52.38 |
| | (50.01) | (49.98) | (49.98) | (50.03) | (50.00) | (50.01) | (50.02) | (50.02) | (49.98) |
| Student's age | 9.74 | 9.75 | 9.72 | 9.74 | 9.74 | 9.75 | 9.72 | 9.74 | 9.70 |
| | (0.60) | (0.63) | (0.57) | (0.55) | (0.56) | (0.53) | (0.64) | (0.68) | (0.59) |
| Fall test total NCE | 42.65 | 41.65 | 43.78 | 41.20 | 39.42 | 43.38 | 43.68 | 43.66 | 43.71 |
| | (18.58) | (19.15) | (17.88) | (17.65) | (17.59) | (17.51) | (19.07) | (19.94) | (18.09) |
| Spring test total NCE | 44.01 | 43.76 | 44.29 | 39.90 | 38.63 | 41.45 | 45.62 | 45.31 | 45.95 |
| | (19.87) | (20.62) | (19.01) | (18.18) | (18.31) | (17.92) | (21.26) | (21.70) | (20.78) |
| **Sample Size** | **2,173** | **1,160** | **1,013** | **899** | **495** | **404** | **1,274** | **665** | **609** |
| **Teacher** | | | | | | | | | |
| Teacher is female | 84.38 | 80.39 | 88.89 | 80.49 | 72.73 | 89.47 | 87.27 | 86.21 | 88.46 |
| | (36.50) | (40.10) | (31.78) | (40.12) | (45.58) | (31.53) | (33.63) | (35.09) | (32.58) |
| Teaching experience | 10.44 | 9.33 | 11.70 | 9.28 | 7.04 | 11.87 | 11.31 | 11.07 | 11.58 |
| | (9.17) | (8.17) | (10.14) | (8.03) | (5.09) | (9.99) | (9.93) | (9.62) | (10.44) |
| Teacher has a master's degree | 33.33 | 29.41 | 37.78 | 31.71 | 27.27 | 36.84 | 34.55 | 31.03 | 38.46 |
| | (47.39) | (46.02) | (49.03) | (47.11) | (45.58) | (49.56) | (47.99) | (47.08) | (49.61) |
| **Sample Size** | **96** | **51** | **45** | **41** | **22** | **19** | **55** | **29** | **26** |
| **School** | | | | | | | | | |
| Percentage scoring below fall test 33rd percentile | 51.68 | | | 53.38 | | | 50.11 | | |
| | (23.03) | | | (24.45) | | | (21.76) | | |
| Percentage scoring below spring test 33rd percentile | 51.81 | | | 58.41 | | | 46.28 | | |
| | (26.22) | | | (25.43) | | | (25.55) | | |
| Percentage receiving free/reduced-price lunch | 62.66 | | | 64.49 | | | 61.22 | | |
| | (22.24) | | | (20.49) | | | (23.98) | | |
| Student/teacher ratio | 16.58 | | | 15.48 | | | 17.44 | | |
| | (2.54) | | | (1.56) | | | (2.86) | | |
| Percentage of Hispanic students | 18.44 | | | 28.76 | | | 10.30 | | |
| | (24.30) | | | (25.57) | | | (20.39) | | |
| Percentage of black students | 55.86 | | | 54.42 | | | 57.00 | | |
| | (39.15) | | | (31.45) | | | (45.14) | | |
| Urban | 52.94 | | | 53.33 | | | 52.63 | | |
| | (50.66) | | | (51.64) | | | (51.30) | | |
| **Sample Size** | **34** | | | **15** | | | **19** | | |

**Table B.3c. Sixth Grade, Descriptive Statistics (means with standard deviations in parentheses)**

| | Sixth Grade—Total | | | Sixth Grade—Achieve Now | | | Sixth Grade—Larson Pre-Algebra | | |
|---|---|---|---|---|---|---|---|---|---|
| | All | Treatment | Control | All | Treatment | Control | All | Treatment | Control |
| **Student** | | | | | | | | | |
| Student is female | 51.60 | 51.54 | 51.68 | 53.52 | 52.29 | 54.90 | 50.81 | 51.45 | 50.10 |
| | (49.98) | (49.99) | (49.99) | (49.90) | (49.99) | (49.81) | (50.00) | (49.99) | (50.02) |
| Student's age | 11.63 | 11.61 | 11.66 | 11.66 | 11.64 | 11.69 | 11.62 | 11.60 | 11.65 |
| | (0.52) | (0.50) | (0.55) | (0.56) | (0.55) | (0.56) | (0.51) | (0.48) | (0.54) |
| Fall test total NCE | 50.29 | 49.53 | 51.37 | 45.16 | 43.40 | 47.13 | 52.39 | 50.82 | 53.45 |
| | (20.90) | (20.35) | (21.61) | (17.42) | (17.07) | (17.61) | (21.83) | (21.31) | (23.04) |
| Spring test total NCE | 51.82 | 51.72 | 51.96 | 48.24 | 46.06 | 50.67 | 53.28 | 53.42 | 52.59 |
| | (20.30) | (20.15) | (20.51) | (19.02) | (18.44) | (19.38) | (20.63) | (20.30) | (21.03) |
| **Sample Size** | **3,625** | **2,137** | **1,488** | **1,037** | **547** | **490** | **2,588** | **1,590** | **998** |
| **Teacher** | | | | | | | | | |
| Teacher is female | 67.95 | 62.22 | 75.76 | 79.49 | 80.95 | 77.78 | 56.41 | 45.83 | 73.33 |
| | (46.97) | (49.03) | (43.52) | (40.91) | (40.24) | (42.78) | (50.24) | (50.90) | (45.77) |
| Teaching experience | 10.54 | 10.17 | 11.05 | 10.49 | 8.56 | 12.74 | 10.59 | 11.58 | 9.02 |
| | (9.22) | (8.79) | (9.90) | (9.19) | (8.53) | (9.65) | (9.38) | (8.96) | (10.14) |
| Teacher has a master's degree | 32.05 | 28.89 | 36.36 | 33.33 | 23.81 | 44.44 | 30.77 | 33.33 | 26.67 |
| | (46.97) | (45.84) | (48.85) | (47.76) | (43.64) | (51.13) | (46.76) | (48.15) | (45.77) |
| **Sample Size** | **78** | **45** | **33** | **39** | **21** | **18** | **39** | **24** | **15** |
| **School** | | | | | | | | | |
| Percentage scoring below fall test 33rd percentile | 37.35 | | | 40.18 | | | 34.51 | | |
| | (19.69) | | | (21.74) | | | (17.82) | | |
| Percentage scoring below spring test 33rd percentile | 32.86 | | | 34.93 | | | 30.78 | | |
| | (18.16) | | | (21.46) | | | (14.73) | | |
| Percentage receiving free/reduced-price lunch | 64.36 | | | 74.04 | | | 54.69 | | |
| | (22.03) | | | (14.21) | | | (24.63) | | |
| Student/teacher ratio | 17.28 | | | 14.82 | | | 19.75 | | |
| | (4.03) | | | (2.26) | | | (3.95) | | |
| Percentage of Hispanic students | 39.67 | | | 42.44 | | | 36.90 | | |
| | (36.49) | | | (35.85) | | | (38.38) | | |

Table B.3c *(continued)*

| | Sixth Grade—Total | | | Sixth Grade—Achieve Now | | | Sixth Grade—Larson Pre-Algebra | | |
|---|---|---|---|---|---|---|---|---|---|
| | All | Treatment | Control | All | Treatment | Control | All | Treatment | Control |
| Percentage of black students | 27.94 | | | 40.19 | | | 15.69 | | |
| | (35.32) | | | (44.50) | | | (17.14) | | |
| Urban | 34.62 | | | 0.00 | | | 69.23 | | |
| | (48.52) | | | (0.00) | | | (48.04) | | |
| **Sample Size** | **26** | | | **13** | | | **13** | | |

**Table B.3d. Algebra I, Descriptive Statistics (means with standard deviations in parentheses)**

| | Total | | | Cognitive Tutor | | | Larson Algebra I | | |
|---|---|---|---|---|---|---|---|---|---|
| | All | Treatment | Control | All | Treatment | Control | All | Treatment | Control |
| **Student** | | | | | | | | | |
| Student is female | 49.90 | 51.32 | 48.14 | 48.87 | 51.14 | 45.71 | 50.83 | 51.85 | 49.64 |
| | (50.01) | (50.01) | (49.99) | (50.02) | (50.04) | (49.90) | (50.01) | (50.00) | (50.04) |
| Student's age | 14.85 | 14.83 | 14.87 | 14.93 | 14.93 | 14.93 | 14.84 | 14.82 | 14.86 |
| | (1.03) | (0.98) | (1.08) | (0.97) | (0.89) | (1.07) | (1.08) | (1.07) | (1.09) |
| Fall test (percent correct) | 32.22 | 31.96 | 32.55 | 28.26 | 27.67 | 29.07 | 34.83 | 35.04 | 34.58 |
| | (11.82) | (11.83) | (11.81) | (10.35) | (9.81) | (11.02) | (11.95) | (12.09) | (11.79) |
| Spring test (percent correct) | 35.71 | 35.28 | 36.25 | 31.47 | 30.55 | 32.76 | 38.51 | 38.64 | 38.37 |
| | (13.30) | (13.23) | (13.36) | (11.60) | (10.54) | (12.86) | (13.56) | (13.84) | (13.23) |
| **Sample Size** | **1,959** | **1,088** | **871** | **755** | **440** | **315** | **1,204** | **648** | **556** |
| **Teacher** | | | | | | | | | |
| Teacher is female | 62.50 | 56.41 | 69.70 | 58.62 | 60.00 | 57.14 | 65.12 | 54.17 | 78.95 |
| | (48.75) | (50.04) | (46.67) | (50.12) | (50.71) | (51.36) | (48.22) | (50.90) | (41.89) |
| Teaching experience | 11.21 | 11.48 | 10.90 | 12.77 | 14.18 | 11.25 | 10.17 | 9.80 | 10.64 |
| | (9.50) | (9.09) | (10.10) | (8.66) | (7.88) | (9.48) | (9.99) | (9.54) | (10.78) |
| Teacher has a master's degree | 54.17 | 53.85 | 54.55 | 41.38 | 40.00 | 42.86 | 62.79 | 62.50 | 63.16 |
| | (50.18) | (50.50) | (50.57) | (50.12) | (50.71) | (51.36) | (48.91) | (49.45) | (49.56) |
| **Sample Size** | **72** | **39** | **33** | **29** | **15** | **14** | **43** | **24** | **19** |
| **School** | | | | | | | | | |
| Percentage receiving free/reduced-price | 52.34 | | | 63.17 | | | 42.42 | | |
| | (25.79) | | | (18.12) | | | (28.40) | | |
| Student/teacher ratio | 16.20 | | | 15.08 | | | 17.22 | | |
| | (3.60) | | | (4.65) | | | (1.99) | | |
| Percentage of Hispanic students | 14.53 | | | 20.54 | | | 9.02 | | |
| | (22.31) | | | (26.47) | | | (17.01) | | |
| Percentage of black students | 44.64 | | | 53.69 | | | 36.35 | | |
| | (34.83) | | | (29.63) | | | (38.35) | | |
| Urban | 47.83 | | | 63.64 | | | 33.33 | | |
| | (51.08) | | | (50.45) | | | (49.24) | | |
| **Sample Size** | **23** | | | **11** | | | **12** | | |

# Appendix C

# Details of Estimation Methods

The first part of the study tests whether teachers' experience using software products for a second year had larger effects on student test scores than in the first year. The question is addressed by restricting the sample of teachers to those that participated in both years of the study. The method used for estimating product effects on student test scores is a two-level hierarchical linear model with students nested within teachers and student and teacher characteristics as predictors of student test scores. The models allow for product effects on student achievement to differ in the first year and in the second year, supporting a test of the hypothesis that teacher experience is related to product effects.

A two-level model is used to estimate experience effects. The model's key component is an interaction between the treatment indicator and a year indicator, as shown in the following equations:

*(C.1 Student)*

$$Y_{ij} = \alpha_{0j} + \alpha_{1j} Y2_{ij} + \pi X_{ij} + \varepsilon_{ij}$$

*(C.2 Teachers)*

$$\alpha_{oj} = \beta_{00} + \beta_{01} T_j + \varphi W_j + \mu_{0j}$$
$$\alpha_{1j} = \beta_{10} + \beta_{11} T_j$$

where the dependent variable Y is the student spring test score. The predictors in the first-level equation (the X variables) are student age, gender, and fall test score[29], and Y2, which is an indicator variable of whether the student participated in the second year of the study. (which is 1 if the student was in the second year and 0 if the student was in the first year). The predictors in the second-level equation are T, an indicator variable of whether the teacher is in the treatment or control group, and W, which are teacher characteristics (years of teaching experience, whether the teacher has a master's degree). Schools are modeled as second-level fixed effects (for each school, the model includes an indicator variable equal to 1 for teachers belonging to a school and 0 for teachers not belonging to the school).

[29] District test scores were used for some students in the second year and the models also include an indicator variable for whether students have a district test score instead of the study administered test score, which is interacted with the fall test score (for example, interaction variables such as ITBS*fall test score or CAT6*fall test score in Tables C.1, C.2, and C.3).

Combining the equations and collecting terms yields a mixed-model estimating equation in which the product effect is related to student and teachers characteristics:

*(C.3 Mixed model with interactions)*

$$Y_{ij} = \beta_{00} + \beta_{01}T_j + \beta_{10}Y2_{ij} + \beta_{11}T_j * Y2_{ij} + \pi X_{ij} + \varphi W_j + \xi_{ij}$$

and the error term has the structure:

$$\xi_{ij} = \mu_{0j} + \varepsilon_{ij} .$$

To simplify the presentation, equation C.3 does not include terms for the school-level indicator variables and for the test interactions (discussed in footnote 30).

The treatment-effect estimator in (C.3) has two components, $\beta_{01}$ and $\beta_{11}$. The first is the product effect in the first year of the study, $\beta_{01}$, the coefficient of the treatment indicator. The second is the difference of the product effect between the first year and the second year, $\beta_{11}$, the coefficient of the interaction of the treatment indicator with the year indicator. The total product effect in the second year is $\beta_{01} + \beta_{11}$. Statistically significant estimates of $\beta_{11}$ are evidence of differences in product effects between the first and second years.

Table C.1 shows complete estimation results and the variables used in the models, (except for coefficients of school indicator variables). Positive coefficients indicate a variable is correlated with an increase in the spring test score and negative coefficients indicate a variable is correlated with a decrease. The units of the coefficient are the same as the units of the test scores, which is normal curve equivalents for first, fourth, and sixth grades, and percent correct for algebra I. The table also shows residual variances at the student and teacher levels, at the bottom of the table.

Treatment effects on year 1 spring test scores reported in the text refer to the estimated coefficients of the "treatment classroom" indicator variable at the teacher level. For example, the treatment effect on first grade spring scores in year 1 shown in Table C.1 as 0.86 corresponds to the estimated coefficient of the treatment classroom indicator. The *p*-value shown in Table II.3 in the main text above is the *p*-value of the estimated treatment coefficient.

The treatment effects on year 2 spring scores reported in the text are the sum of the estimated coefficients of the "treatment classroom" indicator variable at the teacher level and the "Year 2 * Treatment (interaction)" estimate. For example, the second-year treatment effect of –1.28 reported in Table II.3 corresponds to the sum of 0.86, the estimated treatment effect of year 1, and –2.14, the interaction of year 2 with the treatment indicator, which is the amount by which the first-year effect is shifted to become the second-year effect. Finally, the difference in effects reported in Table II.3 of -2.14 corresponds to the interaction of year 2 with the treatment indicator, which is what we interpret as the experience effect using software products for a second year on student test scores.

### Models for Individual Product Effects

The model used to estimate individual product effects is similar to the model presented above. The difference is that product effects are constrained to be equal in both years, which is done by setting $\beta_{11} = 0$. The constraint forces the treatment effect to have one component, $\beta_{01}$.

Table C.2 presents estimates of individual product effects based on teachers, students, and schools that participated in the study either in the first or in the second year. The effects are referred to as product effects for the full sample because they are based on samples that include teachers who participated in the study either in one year of the study (first or second) and teachers who participated in both years. Table C.3 presents product effects using only the sample of teachers, students, and schools that participated in the second year of the study. In the tables, the estimated coefficients for the variable "treatment classroom" are the treatment effects of interest.

**Table C.1. Product Effects in Year 2 Compared to Product Effects in Year 1 Hierarchical Linear Model Estimates: Outcome Is Spring Test Score (standard errors in parentheses)**

| | First | Fourth | Sixth | |
|---|---|---|---|---|
| Variable Name | Grade | Grade | Grade | Algebra I |
| **Student Level** | | | | |
| Intercept | 49.11*** | 50.31*** | 52.96*** | 35.34*** |
| | (1.22) | (1.07) | (1.24) | (0.82) |
| Student age | -3.44*** | -3.82*** | | |
| | (0.83) | (1.08) | | |
| Student is female | 1.37** | 1.37 | 0.46 | -1.70** |
| | (0.61) | (0.95) | (0.49) | (0.71) |
| Fall test score | 0.70*** | 0.74*** | 0.72*** | 0.36*** |
| | (0.01) | (0.02) | (0.01) | (0.03) |
| Year 2 | 3.61*** | -1.30 | -1.39 | -1.16 |
| | (0.97) | (1.55) | (0.90) | (1.08) |
| ITBS*Fall test score | -0.03 | 0.02 | 0.03 | -0.13 |
| | (0.04) | (0.04) | (0.04) | (0.10) |
| NMSBA*Fall test score | | | 0.12*** | |
| | | | (0.03) | |
| SAT10*Fall test score | 0.01 | | | |
| | (0.03) | | | |
| CAT6*Fall test score | | 0.31*** | | |
| | | (0.06) | | |
| **Classroom Level** | | | | |
| Treatment classroom | 0.86 | 2.65 | -0.44 | -0.34 |
| | (1.67) | (1.54) | (1.87) | (1.13) |
| Year 2* treatment classroom | -2.14* | 2.02 | -2.80** | 2.90** |
| | (1.22) | (1.89) | (1.14) | (1.44) |
| Teacher has a master's degree | -3.75 | 2.78 | -3.26 | -0.07 |
| | (2.33) | (2.08) | (2.82) | (1.21) |
| Years of teaching experience | -0.06 | 0.19* | 0.03 | -0.02 |
| | (0.13) | (0.06) | (0.11) | (0.05) |
| **Residual Variance** | | | | |
| Student level | 125.74 | 129.47 | 138.60 | 125.75 |
| Classroom level | 17.67*** | 0.03 | 16.86*** | 0.27 |

Note: School indicators were also included as covariates in the models but are not presented in the tables.

*Statistically significant at the .10 level, two-tailed test.
**Statistically significant at the .05 level, two-tailed test.
***Statistically significant at the .01 level, two-tailed test.

**Table C.2** **Product Effects for the Full Sample (First and Second Years)**
**Hierarchical Linear Model Estimates: Outcome Is Spring Test Score**
**(standard errors in parentheses)**

| | First Grade | | | | Fourth Grade | | Sixth Grade | | Algebra I | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Destination Reading | Headsprout | Plato Focus | Waterford Early Reading Program | Academy of Reading | LeapTrack | Achieve Now | Larson Pre-Algebra | Cognitive Tutor | Larson Algebra I |
| **Student Level** | | | | | | | | | | |
| Intercept | 50.23*** | 55.97*** | 50.77*** | 49.11*** | 39.82*** | 45.54*** | 38.13 | 52.73*** | 32.19*** | 37.84*** |
| | (0.77) | (0.52) | (0.65) | (0.67) | (0.45) | (0.39) | (23.35) | (0.73) | (0.53) | (0.55) |
| Student is female | 1.33 | -0.47 | 0.26 | 1.48** | 1.67** | 0.79 | 0.11 | 0.12 | -0.89 | -0.55 |
| | (0.8) | (0.81) | (0.95) | (0.67) | (0.7) | (0.61) | (0.64) | (0.49) | (0.72) | (0.69) |
| Student age | -1.61 | -3.33*** | -5.46*** | -2.45** | -0.47 | -2.73*** | -0.49 | -1.42*** | | |
| | (1.01) | (0.84) | (1.39) | (0.98) | (0.69) | (0.52) | (0.67) | (0.51) | | |
| Fall test score | 0.68*** | 0.77*** | 0.71*** | 0.74*** | 0.79*** | 0.74*** | 0.6 | 0.7*** | 0.28*** | 0.43*** |
| | (0.02) | (0.02) | (0.02) | (0.01) | (0.02) | (0.01) | (0.36) | (0.01) | (0.03) | (0.03) |
| ITBS*Fall test score | | | | 0.02 | -0.04 | | -0.04 | | | -0.17** |
| | | | | (0.03) | (0.02) | | (0.04) | | | (0.07) |
| SAT10*Fall test score | 0.01 | | | | | | | | | |
| | (0.02) | | | | | | | | | |
| CAT6*Fall test score | | | | | | 0.29*** | | | | |
| | | | | | | (0.05) | | | | |
| NMSBA*Fall test score | | | | | | | 0.05 | | | |
| | | | | | | | (0.03) | | | |

Table C.2 *(continued)*

| | First Grade | | | | Fourth Grade | | Sixth Grade | | Algebra I | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Destination Reading | Headsprout | Plato Focus | Waterford Early Reading Program | Academy of Reading | LeapTrack | Achieve Now | Larson Pre-Algebra | Cognitive Tutor | Larson Algebra I |
| **Classroom Level** | | | | | | | | | | |
| Treatment classroom | 1.91 | 0.29 | 0.50 | 0.42 | -0.16 | 1.97** | -0.58 | 2.37 | -1.28 | -0.1 |
| | (1.67) | (1.09) | (1.39) | (1.41) | (1.01) | (0.73) | (1.45) | (1.56) | (1.1) | (1.08) |
| Teacher has a master's degree | -1.05 | 0.15 | -0.42 | -2.02 | -0.14 | 1.52 | -1.60 | 1.23 | 0.96 | 0.77 |
| | (2.09) | (1.33) | (1.95) | (1.65) | (1.31) | (1.01) | (2.26) | (1.96) | (1.75) | (1.54) |
| Years of teaching experience | -0.28** | -0.05 | 0.07 | -0.05 | 0.03 | 0.10** | 0.08 | 0.02 | -0.08 | 0.10 |
| | (0.12) | (0.07) | (0.09) | (0.11) | (0.09) | (0.04) | (0.1) | (0.12) | (0.1) | (0.07) |
| **Residual Variance** | | | | | | | | | | |
| Student level | 113.64 | 143.27 | 129.75 | 124.25 | 103.28 | 111.58 | 97.51 | 147.64 | 92.81 | 135.26 |
| Classroom level | 15.11 | 8.32 | 5.92 | 15.21 | 3.24 | 1.81 | 11.81 | 17.64 | 3.45 | 5.34 |

Note: School indicators were also included as covariates in the models but are not presented in the tables.

*Statistically significant at the .10 level, two-tailed test.
**Statistically significant at the .05 level, two-tailed test.
***Statistically significant at the .01 level, two-tailed test.

**Table C.3 Product Effects for the Second-Year Sample**
**Hierarchical Linear Model Estimates: Outcome Is Spring Test Score**
**(standard errors in parentheses)**

| | First Grade | | | |
|---|---|---|---|---|
| Variable Name | Destination Reading | Headsprout | Plato Focus | Waterford Early Reading |
| **Student Level** | | | | |
| Intercept | 53.84*** | 57.42*** | 52.51*** | 51.20*** |
| | (1.00) | (0.85) | (0.71) | (1.02) |
| Student is female | 1.65* | 1.10 | -0.21 | 1.40 |
| | (0.91) | (1.33) | (1.22) | (1.16) |
| Student age | -0.65 | -4.73*** | -6.6*** | -2.52 |
| | (1.27) | (1.53) | (1.72) | (1.67) |
| Fall test score | 0.62*** | 0.64*** | 0.62*** | 0.66*** |
| | (0.06) | (0.04) | (0.03) | (0.03) |
| ITBS*Fall test score | | | | -0.09 |
| | | | | (0.08) |
| SAT10*Fall test score | 0.05 | | | |
| | (0.07) | | | |
| **Classroom Level** | | | | |
| Treatment classroom | 2.19 | -4.13* | -0.10 | -1.76 |
| | (2.08) | (1.92) | (1.45) | (2.02) |
| Teacher has a master's degree | -2.19 | -3.97 | -3.27 | -4.13 |
| | (2.76) | (3.26) | (2.21) | (3.06) |
| Years of teaching experience | -0.32 | -0.19 | 0.01 | -0.24 |
| | (0.20) | (0.13) | (0.09) | (0.20) |
| **Residual Variance** | | | | |
| Student level | 88.68 | 115.16 | 105.67 | 104.18 |
| Classroom level | 19.87 | 4.26 | 2.98 | 11.11 |

Table C.3 *(continued)*

| | Fourth Grade | | Sixth Grade | | Algebra I | |
|---|---|---|---|---|---|---|
| Variable Name | Academy of Reading | LeapTrack | Achieve Now | Larson Pre-Algebra | Cognitive Tutor | Larson Algebra I |
| **Student Level** | | | | | | |
| Intercept | 46.21*** | 59.95*** | 47.54*** | 51.41*** | 31.88*** | 40.19*** |
| | (1.16) | (1.05) | (1.65) | (1.22) | (0.93) | (0.67) |
| Student is female | 2.00 | 2.27 | 0.53 | 0.32 | 0.39 | -2.24* |
| | (1.34) | (1.92) | (1.36) | (1.32) | (1.23) | (1.21) |
| Student age | -0.90 | -7.08*** | -0.73 | -1.61 | | |
| | (1.61) | (2.55) | (1.45) | (1.43) | | |
| Fall test score | 0.86*** | 0.63*** | 0.77*** | 0.68*** | 0.34*** | 0.53*** |
| | (0.06) | (0.06) | (0.08) | (0.04) | (0.06) | (0.05) |
| ITBS*Fall test score | -0.08 | | 0.06 | | | -0.14 |
| | (0.08) | | (0.12) | | | (0.21) |
| SAT10*Fall test score | | | | | | |
| CAT6*Fall test score | | 0.19 | | | | |
| | | (0.13) | | | | |
| NMSBA*Fall test score | | | 0.01 | | | |
| | | | (0.10) | | | |
| **Classroom Level** | | | | | | |
| Treatment classroom | 1.86 | 2.88 | -1.59 | -0.44 | -2.10 | 2.59 |
| | (2.78) | (1.94) | (4.32) | (2.53) | (1.87) | (1.57) |
| Teacher has a master's degree | 1.74 | | -6.32 | -3.51 | 4.56 | -0.41 |
| | (3.74) | | (4.69) | (6.75) | (2.62) | (2.70) |
| Years of teaching experience | 0.10 | | 0.19 | 0.16 | -0.03*** | 0.07 |
| | (0.23) | | (0.26) | (0.21) | (0.18) | (0.08) |
| **Residual Variance** | | | | | | |
| Student level | 121.61 | 157.01 | 132.60 | 154.06 | 98.37 | 147.31 |
| Classroom level | 12.82 | 0.19 | 43.59 | 18.99 | 7.90 | 0.39 |

Note: School indicators were also included as covariates in the models but are not presented in the tables.

*Statistically significant at the .10 level, two-tailed test.
**Statistically significant at the .05 level, two-tailed test.
***Statistically significant at the .01 level, two-tailed test.

www.ed.gov ies.ed.gov