A Multisite Cluster Randomized Trial of the Effects of CompassLearning Odyssey® Math on the Math Achievement of Selected Grade 4 Students in the Mid-Atlantic Region

Final Report

ies **NATIONAL CENTER** FOR **EDUCATION EVALUATION** AND **REGIONAL ASSISTANCE**

Institute of Education Sciences

U.S. Department of Education

REL **MID-ATLANTIC**

Regional Educational Laboratory
At Pennsylvania State University

A Multisite Cluster Randomized Trial of the Effects of CompassLearning Odyssey® Math on the Math Achievement of Selected Grade 4 Students in the Mid-Atlantic Region

November 2009

Authors:

Kay Wijekumar
Pennsylvania State University

John Hitchcock
ICF International and Ohio University

Herb Turner
ANALYTICA and University of Pennsylvania

PuiWa Lei
Pennsylvania State University

Kyle Peck
Pennsylvania State University

Project Officer:

Ok-Choon Park
Institute of Education Sciences

NATIONAL CENTER FOR
EDUCATION EVALUATION
AND REGIONAL ASSISTANCE

Institute of Education Sciences

U.S. Department of Education

Disclosure of potential conflict of interest

None of the authors or other staff involved in the study from ANALYTICA, ICF International, Ohio University, Pennsylvania State University, or the University of Pennsylvania have financial interests that could be affected by the content of this report.*

CONTENTS

FIGURES

TABLES

Contents   v

EXHIBITS

SUMMARY

A major goal of U.S. education policymakers during the past two decades has been to improve math achievement (Faulkner et al. 2008). Toward this end, policymakers have passed legislation, formulated policies, raised standards, and redesigned assessments (MacCaffrey et al. 2001; Business Coalition for Education Reform 1998). The No Child Left Behind Act of 2001 emphasizes the importance of mathematics, among other areas, by requiring that all U.S. students be proficient in math by 2014, as measured by annual state-level assessments (NCLB 2009). Because the Regional Educational Laboratory (REL) Mid-Atlantic, in discussions with stakeholders, had identified the need to find innovative and effective approaches to improve math achievement as a priority and because Gonzalez et al. (2004) have shown that grade 4 is a critical point in the elementary school curriculum at which the United States is losing ground to other countries, REL Mid-Atlantic proposed to study promising approaches to mathematics instruction at the grade 4 level.

In an effort to identify instructional methods that might improve mathematics learning at this level when used in a variety of educational settings under typical conditions, the research team looked for promising, replicable practices that were being used broadly by teachers in U.S. schools, for which research showed promising results but had not been conducted using methodologies that can establish causal relationships.

CompassLearning's Odyssey® Math product met all of these criteria. Odyssey Math is marketed as a comprehensive mathematics instructional software product that can help math educators improve their instruction as either a core math curriculum or a partial substitute. Compass Learning's Odyssey®, which includes Odyssey Math, is used with 3 million students in 5,000 schools throughout the United States. Since the software was released, more than 11 million students have used it. The developer also reports that 693 schools in the Mid-Atlantic Region were using the Odyssey software in 2005.

Despite this widespread use, the effect of Odyssey Math software on math achievement has not been rigorously studied in a randomized trial of effectiveness. An effectiveness trial would study the effect of Odyssey Math on student learning in the instructional environment that would typically occur had the school district purchased Odyssey Math and associated professional development and implemented it naturally. Previous research on Odyssey Math lacked the appropriate control groups to generate evidence from which to draw conclusions about the effects of the software (CompassLearning 2005, 2006, 2007, 2008a, 2008b). This, coupled with educators' growing desire to use better quality evidence when making curriculum decisions, prompted this effectiveness study, which addresses the following confirmatory research question:

- Do grade 4 classrooms using Odyssey Math as a partial substitute for the standard math curriculum outperform control classrooms on the math subtest of the TerraNova CTBS Basic Battery in a typical school setting?

- What is the effect of Odyssey Math on the math performance differential between male and female students in a typical school setting?
- What is the effect of Odyssey Math on the math performance differential between low- and medium/high-scoring students on a math pretest in a typical school setting?

Consistent with the purpose of an effectiveness study, REL Mid-Atlantic defined "use of Odyssey Math" as classrooms having access to Odyssey Math and students using the software modules as a partial substitute for the core math curriculum under the supervision of teachers who had received five "days" of CompassLearning's professional development. Teachers were advised and regularly encouraged to deliver Odyssey Math to their students for 60 minutes each week. However, the study team did not intervene with teachers whose curriculum delivery resulted in students using Odyssey Math less than 60 minutes per week. During monthly conference calls, the study team received confirmation from the Odyssey Math team that the implementation within schools was typical. Variation in teacher delivery and student use of Odyssey Math was consistent with the research questions addressed in an effectiveness study. Actual student use of the curriculum was monitored and recorded through a tracking system built into the Odyssey software.

RECRUITMENT, STATISTICAL POWER, AND STUDY CONDITIONS

The study was designed as a randomized controlled trial to obtain statistically unbiased estimates of the effect of Odyssey Math on the math achievement of grade 4 students. A statistical power analysis, which assumed a minimum detectable effect size of 0.20, showed that at least 28 elementary schools would be needed for the study. To provide a buffer against attrition, 32 elementary schools (including intermediate and charter schools) were recruited from the Mid-Atlantic Region (Delaware, District of Columbia, Maryland, New Jersey, and Pennsylvania). All schools volunteered to participate in the study and were not randomly sampled from the universe of eligible schools in the region. The final sample included 32 schools in Delaware, New Jersey, and Pennsylvania.

Within each participating school, all grade 4 teachers' classrooms were randomly assigned to intervention or control groups. The control group in each school used the same mathematics curriculum as the intervention group in that school. The random assignment produced two groups of classrooms that did not differ significantly on a pre-intervention measure of math achievement or other characteristics, including socioeconomic status, percentage of English language learner students, racial/ethnic minority students, gender, and teacher participation in professional development.

Teachers in the intervention condition were advised and regularly reminded to use Odyssey Math for 60 minutes each week as a partial substitute for the regular math curriculum by the CompassLearning professional development team during professional development sessions and by the REL study team in letters. Total time for daily and weekly math instruction was to be identical for both the intervention and control classrooms. The Odyssey Math usage statistics showed that intervention classrooms devoted an average of 38 minutes each week to the software. The time spent on Odyssey Math was expected to be

Odyssey Math usage statistics showed that intervention classrooms devoted an average of 38 minutes each week to the software. The time spent on Odyssey Math was expected to be integrated into the overall math instructional time to avoid confounding the amount of instructional time with the use of Odyssey Math.

ANALYSIS AND RESULTS

At posttest the sample included 32 schools, 122 teachers, and 2,456 students, approximately balanced across intervention and control conditions. The analyses tested the mean difference of student achievement between intervention and control conditions at the classroom level while accounting for students clustered by classrooms, which were clustered by schools.

This study found no statistically significant difference between classrooms that used Odyssey Math and those that did not on an end-of-school-year math achievement test, the math subtest of the TerraNova Basic Battery (CTB/McGraw-Hill 2000).

CONCLUSIONS

This study was the first randomized controlled trial to assess the impact of Odyssey Math on student achievement. The study had the statistical power needed to detect a 0.20 effect size and was well designed in that comparable groups were created at baseline and maintained through posttesting. Implementation during the school year was documented and shown to be consistent with typical implementation of the Odyssey Math software. The results from the multilevel model with pretest covariates also indicate that Odyssey Math did not yield a statistically significant impact on end-of-year student achievement. This study generated a statistically unbiased estimate of the effect of Odyssey Math on student achievement when implemented in typical school settings with typical teacher and student use. However, the findings apply only to participating schools, teachers, and students because the study used a volunteer sample.

1. STUDY BACKGROUND

Mathematics is an integral part of science, technology, and many other aspects of modern life, from managing household accounts to modeling complex systems and competing for a high-skilled, high-wage job in the global economy (National Council of Teachers of Mathematics 2008). Improving math achievement has been a major goal of U.S. education policymakers during the past two decades (Faulkner et al. 2008). Policymakers have formulated policies, passed legislation, raised standards, and redesigned assessments (MacCaffrey et al. 2001; Business Coalition for Education Reform 1998). Much of this intensified concern came in response to the 1983 National Commission on Excellence in Education's *A Nation at Risk*, which argued that raising U.S. students' math achievement to world-class levels was essential to their success in a global economy and in life (National Commission on Excellence in Education 1983). Through the No Child Left Behind Act of 2001*,* improving math achievement is now a legislative mandate for state and district education policymakers (Elledge et al. 2009). Emphasizing the importance of math, the act requires that all students be proficient in math by 2014, as measured by annual state-level assessments.

NEED FOR THE STUDY

In needs identification conversations with the Regional Educational Laboratory (REL) Mid-Atlantic, state and local education stakeholders in Delaware, the District of Columbia, Maryland, New Jersey, and Pennsylvania all identified improving math achievement as a priority and expressed a need for effective and innovative approaches to enhance math achievement. To address this need, REL Mid-Atlantic proposed an investigation into the use of a computer-based math curriculum as a partial substitute for regular math instruction.

Computer-based math curricula have been reported to assist teachers with varying levels of subject expertise, provide individualized instruction, motivate students, and provide continual feedback and assessment (Faulkner et al. 2008).

REL Mid-Atlantic further proposed to study a computer-based math curriculum that targets grade 4 students. In a report on the 2003 Trends in International Mathematics and Science Study (TIMSS), Gonzales et al. (2004) show that grade 4 is a critical point in the elementary school curriculum. They further reveal that U.S. student achievement in math at the grade 4 level was declining relative to the achievement of students in 14 other tested countries, from ranking 6th among 15 countries in 1995 to 8th among 15 in 2003. The National Assessment of Educational Progress also showed that 18 percent of U.S. grade 4 students performed below basic in their math achievement test (NAEP 2007).

Odyssey® Math (CompassLearning 2005) was selected as the program to be studied because it met the criteria set for the study: it was widely used, was replicable if some evidence of effectiveness were found, offered professional development and support

throughout the school year, and showed promise of effectiveness through prior research, though that research was not methodologically sufficient to establish a causal relationship.

A BRIEF DESCRIPTION OF ODYSSEY MATH

Odyssey Math is a computer-based math curriculum developed by CompassLearning, Inc., to improve math learning for K–12 students. The software consists of a web-accessed series of learning activities, assessments, and math tools. These components constitute the basic framework of the software. CompassLearning professional development trainers presented the learning activities, math tools, and assessments as available options to intervention teachers during the summer professional development session.

The Odyssey Math software includes learning activities with narrative descriptions of how to solve problems, practice tasks that allow learners to apply their knowledge in different contexts, quizzes, assessments, and feedback for students. Teachers can select practice tasks for all students or allow the software to assess each student's skill level and place individual students in appropriate learning activities. Teachers can also preselect a series of lessons through which students progress during the year. The software is intended to be used as the main curriculum in a school or as a partial substitute for the main curriculum. The second mode was chosen for this study. (Chapter 3 provides further details about the software and its use in this study.)

Professional development

The Odyssey package includes teacher professional development, offered in large group sessions during the summer and in individual in-class coaching sessions throughout the school year. Several professional development packages are offered, varying by number of "days" and content.[1] For this study five days[2] of professional development were purchased for each teacher, consisting of two large group presentations and three in-class coaching sessions. This level of professional development was selected because it represented what the vendor agreed was a typical implementation. The large group sessions covered introduction to the software and guidance on selecting learning activities, running reports, and choosing assessments. The individual coaching sessions covered these areas in more depth and were customized to each teacher's needs. Teachers learned to identify math learning objectives and to assess student progress in meeting these objectives using on-screen manipulatives and guided feedback embedded in the software. (See chapter 3 for complete information about the professional development packages available, rationale for the choice, and descriptions of the contents.)

---

[1] The developer uses the term "day" for financial accounting purposes and not to describe actual instructional contact time between CompassLearning staff and teachers. A "day" is roughly the amount of time the developer needs to prepare and deliver the intended curriculum. Summer training "days" average 5–6 hours of training time. Coaching "days" average 1–2 hours of instruction for an individual teacher.

[2] The original contract was to include six days, but the last of those days was scheduled to occur after the posttest and was about planning for the following year.

Intended implementation

The study design called for the software to be delivered for approximately 60 minutes each week by teachers who participated in five "days" of professional development on the software. Key intervention features for students were built-in individualized assessments for each learning objective, multimedia-based interactive learning activities, and practice tasks with feedback. The students would use the software's assessments (quizzes), learning activities, and feedback in place of a teacher-led learning activity during this 60 minutes. The student to computer ratio was expected to be 1:1.

According to the developer and its professional development model (see appendix A), these features of the program combine to allow trained teachers to apply principles of differentiated instruction for learners with different prior knowledge and mathematics skills. Use of assessments generates data that can be used to develop specialized instructional plans using modules built into the package. Furthermore, the developer believes that the software's immediate feedback coupled with graphics and sound can help teachers better deliver math content and thus improve student performance.

Current and prospective use in Mid-Atlantic Region

As of September 2005 Odyssey Math was used in all the Mid-Atlantic jurisdictions (table 1). In all, 693 schools in the Mid-Atlantic Region used Odyssey Math, and 145 schools planned to purchase it. According to the developer, nationwide the Odyssey suite of products (Math, Language Arts, and others) is used with 3 million students in grades K–12 in 5,000 schools.[3]

Table 1. Current and prospective use of Odyssey Math in the Mid-Atlantic Region, 2004/05 (number of schools)

Jurisdiction	Current use	Planned purchase	Total
Delaware	6	10	16
District of Columbia	4	0	4
Maryland	30	20	50
New Jersey	252	40	292
Pennsylvania	401	75	476
Total	693	145	838

Source: U.S. Department of Education 2008.

PREVIOUS RESEARCH ON ODYSSEY MATH

A literature search was conducted to review research on the effects of Odyssey Math on grade 4 students in the Mid-Atlantic Region and across the country. The search identified 15 reports describing 14 studies. No studies were published in peer-reviewed journals. Thirteen reports were published by the software developer, CompassLearning. Another report was published as a CompassLearning report, but it was a reanalysis of a previous

---

[3] Since Odyssey's release, more than 11 million students have used it.

study reported by CompassLearning (Brandt and Hutchinson 2006). One was an unpublished dissertation (Martin 2005).

Of the 14 studies reviewed, 2 were conducted in high schools, 4 in middle schools, and 8 in elementary schools. Seven studies reported results for grade 4 students (table 2). Among the findings:

- Of the five studies that reported weekly use, use ranged from 30 to 135 minutes.
- All studies reported positive gain scores or effect sizes for grade 4 math achievement but did not report whether these gains were statistically significant. For example, CompassLearning (2008b) reported an average increase of 11.1 points (compared with the Northwest Evaluation Association increase of 8.8 points in the norm sample) and Clariana (2007) reported effect sizes as high as 0.33 and 0.49 standard deviations.
- All the studies evaluated the effect on math achievement based on changes in outcome scores between the start and end of the school year.
- None of the studies used a randomized controlled trial design.
- None of the studies used a valid control group as a counterfactual.
- Of the two studies that used a comparison group, only one controlled for pretest differences between the comparison group and the group using Odyssey Math.

Table 2. Odyssey Math studies reporting results for grade 4 students, 2005–08

Study	Target population	Weekly use (minutes)	Design and analysis	Math outcome measure
CompassLearning (2005)	Grade 4	60–90	Trends	District test
CompassLearning (2006)	Grades 2–6	75	Trends	Mississippi Curriculum Test
Bailey and Majors (2007)	Grades 4 and 5	135	Nonequivalent control group	Ohio Achievement Test
Clariana (2007)	Grades 3 and 4	30–60	Trends and correlations	NJ Assessment of Skills–Math
CompassLearning (2007)	Grades 4–6	Not reported	Trends	Measure of Academic Progress–Math
CompassLearning (2008a)[a]	Grades 3–6	30	Trends	Michigan Educational Assessment Program–Math
CompassLearning (2008b)	Grades K–8	Not reported	Trends	Measure of Academic Progress–Math

a. Study does not separate outcomes for grade 4.
Source: Authors' compilation.

Based on the gains in scores shown in these studies using nonexperimental research designs, Odyssey Math showed that it might generate a positive effect on student achievement. However, without a randomized controlled trial design and a valid control group, the many alternative factors that could explain the observed gains could not be ruled out (Bloom 2005; Boruch 1997; Wiersma and Jurs 2005).

In interpreting the observed achievement gains, there are also other concerns about the statistical validity of the conclusions. None of the score gains was reported with its

standard error, which measures the variability in the score gain due to sampling (Moore, McCabe, and Craig 2009; Lipsey and Wilson 2001). Thus, some of the positive gain in scores could be due to chance, attributable to study sample selection (sampling variability). None of the studies reports levels of statistical significance.

Thus, all the studies show positive growth in math achievement but lack valid randomly assigned control groups that would enable the achievement gains to be causally attributed to Odyssey Math.

NEED FOR EXPERIMENTAL EVIDENCE

A compelling case therefore exists for conducting a randomized controlled trial on Odyssey Math at grade 4 in the Mid-Atlantic Region, based on the following factors:

- There is a strong interest in raising math achievement in the Mid-Atlantic Region.
- The use of Odyssey Math is broad and growing in the Mid-Atlantic Region.
- No experimental evidence rules out alternative explanations for the observed effects of Odyssey Math.
- The No Child Left Behind Act of 2001 requires that education decision makers base instructional practices and programs on scientifically valid research.
- Only a randomized controlled trial—that has sufficient statistical power, is well designed (creating comparable groups at baseline and maintaining their comparability to the end of the study), and is implemented with high fidelity—can generate statistically unbiased estimates of the effects of Odyssey Math on outcomes of interest, such as student achievement (Boruch 1997).

RESEARCH QUESTIONS

This study sought to answer one confirmatory question and two exploratory questions. While the answer to the first question can be used to inform curriculum decisions, the answers to the other two questions can be used only to inform future research—as the exploratory analyses are not designed to determine whether the observed effects of Odyssey Math are real or due to chance.

The confirmatory question:

- Do grade 4 classrooms using Odyssey Math as a partial substitute for the standard math curriculum outperform control classrooms on the math subtest of the TerraNova CTBS Basic Battery (CTB/McGraw-Hill 2000) in a typical school setting?

The study also posed two exploratory questions. One is on gender differences in math achievement, which have concerned educators and researchers over the last several decades

(Campbell and Clewell 1999; Liu and Wilson 2009; Neuschmidt, Barth, and Hastedt 2008). The other considers whether Odyssey Math has a differential impact on low scorers and high scorers, as interventions often do (Caraisco-Alloggiamento 2008). The two exploratory questions:

- What is the effect of Odyssey Math on the math performance differential between male and female students in a typical school setting?
- What is the effect of Odyssey Math on the math performance differential between low- and medium/high-scoring students on a math pretest in a typical school setting?[4]

Consistent with the purpose of an effectiveness study, the study team defined "use of Odyssey Math" as classrooms having access to Odyssey Math and students using the software modules as a partial substitute for the core math curriculum under the supervision of teachers who had received five "days" of CompassLearning's professional development. As is typical for such use of Odyssey Math, teachers were able to decide whether to substitute Odyssey Math for classroom learning activities, teacher-led instruction, quizzes, tests, or some combination. Teachers were advised and encouraged by CompassLearning trainers and subsequently by the REL Mid-Atlantic study team to use Odyssey Math as a partial substitute for the core curriculum for 60 minutes a week throughout the school year.

---

[4] Low-scoring students are defined as those who score below the grade 4 level on a TerraNova CTBS Basic Battery pretest. Medium/high-scoring students are those who score at or above grade 4 level.

2. STUDY DESIGN AND METHODOLOGY

This chapter presents the study design and methodology. It describes the research design, sample recruitment and incentives to participate, random assignment, baseline equivalence, outcome measures, and data collection and analysis methods. It also discusses missing data, alternative models, and sensitivity analyses.

A MULTISITE CLUSTER RANDOMIZED TRIAL

The study used a multisite cluster randomized trial to assess the effects of Odyssey Math on the math achievement of grade 4 students in the Mid-Atlantic Region. A volunteer sample of teachers and their classrooms were randomly assigned to intervention and control conditions within schools. Teachers in the intervention condition agreed to integrate Odyssey Math into the standard math curriculum by substituting Odyssey Math for 60 minutes a week of regular math instruction. This weekly use was based on the software developer's definition of "typical use" of Odyssey Math. During the rest of the math instructional time the intervention teachers provided math instruction using their school's standard curriculum. The control teachers used the school's standard mathematics curriculum for the total math instructional time. Schools signed a memorandum of understanding agreeing to keep total math instructional time at the standard length for all classrooms during the academic year.

JUSTIFICATION OF THE STUDY DESIGN

A multisite cluster randomized trial design that uses teacher random assignment within each school was selected over other designs that use school- or student-level random assignment. A design based on student-level random assignment was considered but rejected because of the expectation that school officials, teachers, and parents would object to leaving student placement in classrooms to chance, creating challenges to school recruitment. Furthermore, random assignment of teachers rather than students reflects the software's typical implementation, in addition to offering the other advantages described. A brief description of additional justifications for choosing the multisite cluster randomized trial design is presented below.

Statistical power

The statistical power analyses showed the within-school random assignment design to be more efficient than the school-level random assignment design. Holding constant other assumptions used in a statistical power analysis, the within-school design required approximately half as many schools as the school-level design to detect the same effect.

Curricular consistency between intervention and control

A within-school random assignment, which randomly assigned classrooms within schools to either the intervention or the control group, ensured that the same curriculum was used in both study conditions in each school.

Access to Odyssey Math as a study recruitment tool

This design offered all teachers professional development and the opportunity to eventually use the Odyssey Math software. The intervention teachers received professional development to deliver the instruction in 2007/08, while the control teachers were offered the same professional development for the following year once the study was completed, along with the option to use Odyssey Math.

Delivery of Odyssey Math and intervention diffusion

Intervention teachers delivered the Odyssey Math software-based instruction in their classrooms or in a computer lab in the school. To limit the risk of intervention diffusion (the use of Odyssey Math in control classrooms), the intervention teachers were instructed not to share their software access passwords or professional development materials with other teachers in the school. The expectation of no diffusion of the Odyssey Math intervention to control teachers and their classrooms was reasonable, because control teachers did not receive professional development and could not view the lesson contents or use Odyssey Math in their classrooms without a password. The risks and consequences of such contamination were explained to teachers and administrators during recruitment and training, and classroom observers who documented instructional activities in intervention and control schools were asked to note any apparent use of Odyssey Math in control classrooms.

STUDY TIMELINE

Table 3 presents a timeline for key activities of the study.

TARGET POPULATION AND RECRUITMENT

Statistical power analysis was conducted in August 2006 using a random effects model to determine the number of schools, teachers, and students needed to detect a minimum effect size for the intervention (see appendix B). Because it seemed likely that teachers would vary in their implementation of Odyssey Math and that the effect sizes would also vary, teacher-level effects were assumed to vary across schools in the hierarchical linear models used in the study.

The statistical power analysis indicated that a minimum of 28 schools and 108 teachers (assumed average of 4 per school) were required (table B1 in appendix B details the complete power analysis). To provide a buffer against potential attrition-related problems,

the study planned to recruit 33 schools,[6] 132 teachers, and 3,100 students (assumed average of 25 per classroom) to detect a 0.2 standard deviation difference between intervention and control classrooms on post-intervention mathematics achievement.

Table 3. Timeline of the Odyssey Math effectiveness study, June 2007–May 2008

Date	Task
June 2007	Participation agreement (memorandum of understanding)
June–July 2007	Assignment of students to classrooms by schools
July 2007	Random assignment of teachers
August 2007	Class rosters emailed from schools in response to study requests
	Notification to schools of teacher random assignment and invitation to intervention teachers for professional development
	Intervention teacher professional development (large group, two-day session)
	Notification of parents for consent forms
September–October 2007	Pretests and submission of student consent
October 2007	Intervention begins
	First in-class coaching session (intervention teacher professional development)
December 2007–January 2008	Classroom observations conducted by study team (intervention and control classrooms)
January 2008	Intervention teacher professional development (large group, one-day)
February–March 2008	Second in-class coaching session (intervention teacher professional development)
April–May 2008	Posttest

Source: Authors' compilation.

Phased recruitment for the study began in January 2007 with outreach and awareness and concluded with schools signing a memorandum of understanding during the summer of 2007. In January 2007 the study team built awareness about the study among schools, districts, and intermediate units across the Mid-Atlantic Region covering Delaware, the District of Columbia, Maryland, New Jersey, and Pennsylvania.

The Common Core of Data was used to develop a list of all elementary schools in these five jurisdictions (U.S. Department of Education 2008). Information from CompassLearning was used to identify and remove from the list schools that were already using Odyssey Math or that had used it within two years of the start date for this study (September 2007).

Later in January 2007 schools were invited to participate in the study. Letters were sent to 1,702 eligible districts with 2,286 elementary schools in the five Mid-Atlantic Region jurisdictions (table 4). Laboratory Extension Specialists followed up with phone calls to the 933 districts closest to REL Mid-Atlantic partner sites (because of the condensed recruiting timeline) to gauge their interest in participating in the study. Additional forums were held for school superintendents and principals at regional locations to broaden the outreach beyond the districts that were called. These activities resulted in 122 informal expressions of interest from districts.

---

[6] Access to participate was open to all schools that met the eligibility criteria, including charter schools.

Prequalification screening was based on the following factors:

- *Number of classrooms available.* Schools had to have a minimum of two grade 4 classrooms so that each school could have at least one intervention classroom and one control classroom. No school was disqualified for having too many available classrooms.
- *The schools' education practices.* Schools were ineligible to participate if they used any of the following practices, which would undermine a multisite cluster randomized trial:
  - o Tracked students into classrooms based on academic performance.
  - o Used different curricula within grade 4 classrooms.
  - o Departmentalized instruction, so that there was only one grade 4 math teacher.
- *Adequate technology.* Schools had to have available at least one computer per student. Students could use central computer laboratories, laptops dedicated to the class during the Odyssey Math use, or laptops assigned to students.
- *No evidence of present or recent (within the last two years) Odyssey Math use in grades 3 or 4.*

Also considered were perceived motivation by principals and teachers to participate in the study and geographic proximity of the school to other study-eligible schools (because of budgetary implications for professional development and data collection).

After prequalification screening and requests for formal expressions of interest between February and May 2007, 64 schools qualified for site visits to solidify interest in the study and assess their readiness to participate, including a technology assessment of school computers and Internet connections.

In June 2007, after receiving approval from the U.S. Office of Management and Budget and the Pennsylvania State University Office of Research Protections, 62 schools were invited to sign memoranda of understanding detailing the conditions for participating, including professional development, random assignment, notification of any students moving into or out of the school district, and use of Odyssey Math for 60 minutes each week. (Two schools were excluded because they did not have the required student to computer ratio of 1:1 that they had reported during initial recruitment.) All classrooms and teachers in the 62 schools were invited to participate in the study. Thirty-two schools signed and returned the memorandum of understanding by the deadline.[7]

Although the recruitment campaign reached out to districts and schools in all the jurisdictions of the Mid-Atlantic Region, in the end all schools meeting the eligibility criteria were in Delaware, New Jersey, and Pennsylvania.

---

[7] Thirty-three schools originally signed and returned the memorandum of understanding, but one school was discovered to be ineligible to participate in the study because of current use of Odyssey Math. This school was dropped following random assignment. Dropping the school did not compromise the study's internal validity, because a multisite cluster trial can be conceived of as a series of miniexperiments that are then aggregated for analysis. Dropping the school meant that both the intervention and the control classrooms were excluded.

Table 4. Sample sizes at different stages of recruitment for the Odyssey Math study

Recruitment activity	Number of districts	Number of schools	Percentage of original sample schools	Percentage of previous sample schools
Invitations mailed (includes charter schools)	1,702	2,286	100	na
Contacted with two follow-up calls	933	na	na	na
Interested in prequalifying	122	na	na	na
Participated in prequalification	94	120	7	na
Submitted an expression of interest	49	79[a]	4	62
Participated in a site visit observation	44	64[a]	3	53
Placed in the memorandum of understanding review pool	42	62[b]	3	97
Placed in the random assignment pool	24	32[c]	1	53

na is not applicable.

a. The drop from 79 schools to 64 schools was a result of scheduling conflicts and the recruitment timeline.

b. Two schools did not qualify for the review pool because they did not have the necessary student to computer ratio.

c. Although 33 schools were randomized, 1 school was determined to be ineligible because of previous use of Odyssey Math and was dropped from the pool.

Source: Authors' analysis.

Table 5 presents the demographic characteristics of the 32 participating elementary, intermediate, and charter schools. Participating schools had an average rate of 78 percent proficiency on state grade 4 math assessment tests, 14.9 students per teacher, and an education expenditure rate of $8,058 per student. The student population was 19 percent racial/ethnic minorities and 36 percent socioeconomically disadvantaged. Half (16) the schools were in rural areas, 19 percent (6) in the urban fringe of a large city, 19 percent (6) in the urban fringe of a mid-size city, 6 percent (2) in a small town, and 3 percent (1 each) in a large city and mid-size city.

INCENTIVES TO PARTICIPATE IN THE STUDY

The study included several incentives for schools to participate. One incentive was access to the Odyssey Math software in intervention teachers' classrooms during the 2007/08 school year (the study year) at no cost and in control teachers' classrooms in 2008/09 (after the study was completed).[8] REL Mid-Atlantic paid the developer $18 per student for use of the software each year.

---

[8] The student subscription cost of $18 per student was based on use of Odyssey Math only rather than the full set of curriculum modules in other subject areas that the developer offers. The developer does not usually separate the costs for the different subjects supported in Odyssey but did so to accommodate this study.

Table 5. Mean characteristics of the 32 participating schools and 122 teachers

Characteristics	Sample mean	Standard deviation	Weighted mean[a]
School characteristics[b]			
Proficiency in state grade 4 math assessment (percent)	77.8	15.8	46.1
Students per teacher	14.9	2.1	14.1
Proportion of racial/ethnic minority students (percent)	18.7	25.8	38.8
Proportion of students eligible for free or reduced-price lunch (percent)	36.3	21.5	35.9
Student education expenditure rate (dollars)[c]	8,058	1,436	na
Teacher characteristics[d]			
Years in current school	10.9	9.8	na
Years of teaching experience	15.4	11.5	na
Proportion with master's degree (percent)	37.8	48.7	na
Previous professional development (past two years)			
Hours of university math courses	6.6	15.7	na
Hours of conferences or workshops on math			
Long training (more than half day)	11.9	17.6	na
Short training (half day or less)	11.5	16.7	na
Hours of math coaching received	6.9	14.1	na

na is not applicable.

a. The number of total reporting schools in each state is used as the weight.

b. Data were obtained from School Data Direct (www.schooldatadirect.org) on January 14, 2009.

c. Defined broadly as expenditures per student for the academic component of their schooling (excluding costs like transportation). An example of the calculation of this rate is available at www.pde.state.pa.us/school_acct/cwp/view.asp?a=182&q=54624.

d. Compiled from the teacher survey developed for this study.

Source: Authors' analysis based on data described in the text.

A second incentive was professional development for all participating teachers at no cost to the school. Intervention teachers received the professional development in 2007/08 and control teachers in 2008/09. The five-day professional development was offered by CompassLearning at a reduced rate based on the large number of "days" purchased for the study, a standard practice. REL Mid-Atlantic purchased 75 "days" of professional development services (both the large group instruction and individual coaching sessions) each year at a per day cost of $1,350.

Finally, REL Mid-Atlantic paid teachers $150 a day for two "days" of summer professional development (to the intervention teachers in 2007/08 and the control teachers in 2008/09). School districts were also reimbursed for the cost of substitute teachers while regular teachers attended professional development sessions.

RANDOM ASSIGNMENT OF TEACHERS

All grade 4 teachers in the participating schools were invited to participate, and none declined. All grade 4 teachers were randomly assigned to the intervention and control conditions after students had been assigned to teachers and before the August 2007 professional development and September 2007 student pretesting. Parent consent forms were mailed before the school year began and did not contain information on student classroom assignment.

Figure 1. Reduction of sample size and explanations from baseline to the final analytical sample
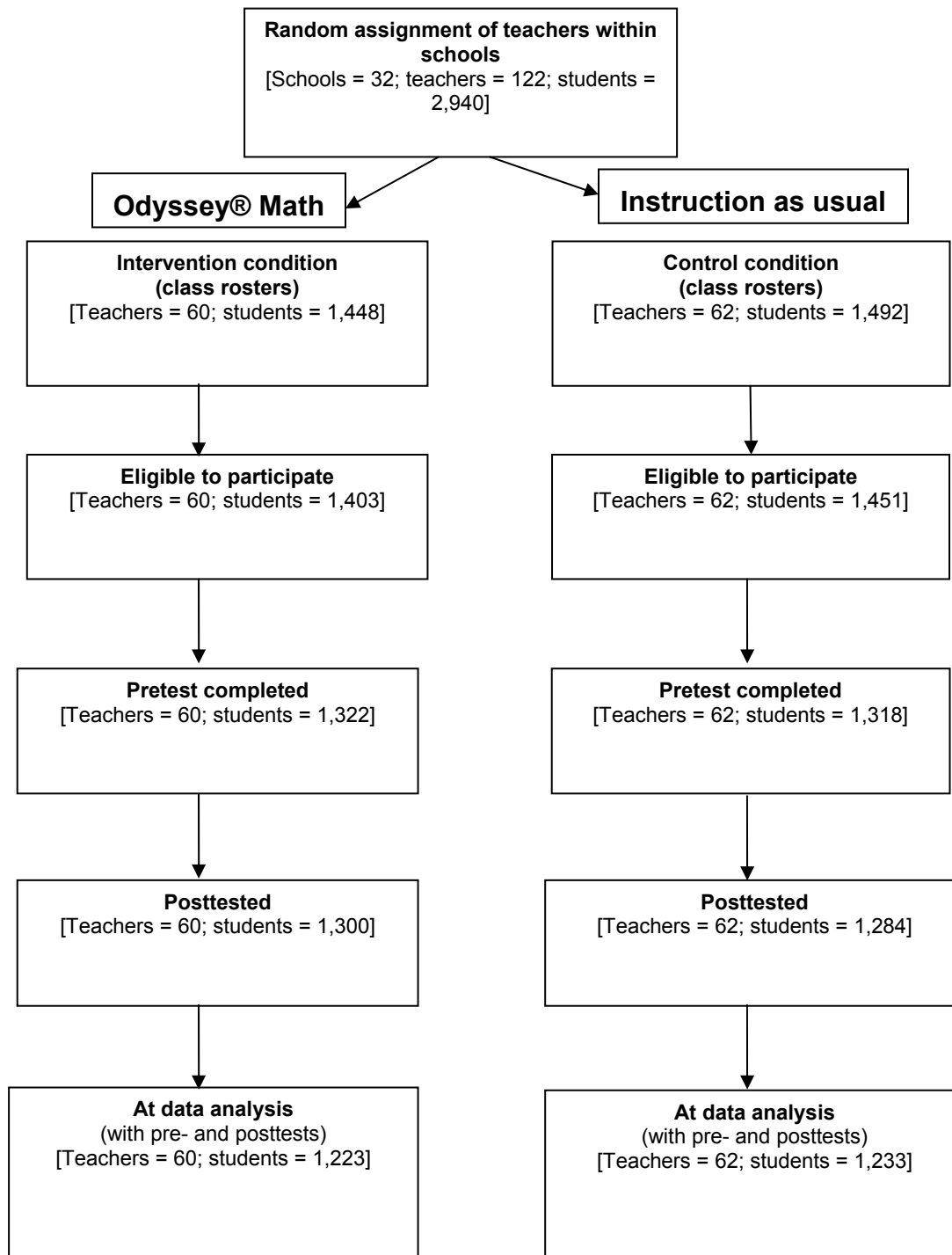
Source: Adapted from the Consolidated Standards on Reporting Trials CONSORT statement (www.consort-statement.org).

In all, 122 teachers were randomly assigned to conditions within schools using Microsoft Excel™ (figure 1 and table 6). The probability of assignment to each condition was 50 percent for schools with an even or odd number of classrooms. An example of how

the random assignment was implemented in all schools, for schools with even and odd numbers of teachers, is in appendix C.

Table 6. Number of schools and grade 4 teachers in random assignment pool

Number of grade 4 teachers in a school	Number of schools	Total number of grade 4 teachers	Percentage of school sample	Cumulative percentage of school sample
2	6	12	19	19
3	5	15	16	35
4	13	52	41	76
5	5	25	15	91
6	3	18	9	100
Total	32	122	100	100

Source: Authors' analysis based on data described in text.

RANDOM ASSIGNMENT, STUDY PARTICIPANTS, AND PARTICIPANT LOSS

To assess whether the integrity of random assignment was maintained throughout the study, the numbers of schools, teachers, and students were tracked through all phases of the study. Figure 1 summarizes the accounting from random assignment to the final analytic sample using a flowchart adapted from the Consolidated Standards on Reporting Trials (CONSORT) statement. The CONSORT statement is required for reporting the results of trials in the *British Medical Journal*. Full documentation of tracking results is in appendix D.

Random assignment phase

Sixty teachers (with 1,448 students) were randomly assigned to the intervention condition, and 62 teachers (with 1,492 students) were randomly assigned to the control condition.

Participation of special education and English language learner students

The schools provided rosters with codes indicating students' special education or English language learner status.[9] These students were classified as ineligible for the pretest when the schools identified them as not having access to the regular math curriculum or not eligible for typical testing conditions because of a specific testing requirement (such as the presence of a translator). Students in these categories were not counted as attrition.[10] Eligibility was determined by school staff. Allowing the schools to make this decision was consistent with typical implementation of Odyssey Math. School staff followed predefined individualized education programs for the students.

---

[9] The schools also notified the study team when a student's status changed.

[10] There were 38 students in this group (29 in the intervention condition and 9 in the control condition). An additional 48 students (16 in the intervention condition and 32 in the control condition) were pretest ineligible because they were either Title I math students or in the dropped school (see table D1 in appendix D).

Eligible to participate in study phase

The pretest eligible sample comprised 32 schools, 122 teachers, and 2,854 students. In this sample 60 teachers and 1,403 students were in the intervention condition, and 62 teachers and 1,451 students were in the control condition. All teachers invited to participate in the study agreed to do so.

Ineligible for pretest stage

Before pretesting, one teacher in the intervention group declined to use the software but agreed to allow students to participate in pre- and posttesting. This teacher was labeled in the sample as an intent-to-treat teacher and was not counted as a reduction in the number of teachers at pretesting (figure 1 lists 60 intervention teachers rather than 59 in the eligible to participate box). Although not shown in figure 1 (but documented in table D1 in appendix D), 15 students in the intervention condition and 16 students in the control condition did not have parental permission to participate and were excluded from testing. Additionally, 27 students in the intervention condition and 84 students in the control condition did not take the pretest for other reasons not reported to the study team. Finally, 39 students in the intervention condition and 33 students in the control condition were not available on the dates established for pretesting.

Eligible to participate

Of the 1,403 students in the intervention condition eligible to participate, 1,322 were pretested. Of the 1,451 students in the control condition eligible to participate, 1,318 were pretested.

Between pretest and posttest phases

Between pre- and posttesting there was a net loss of 22 students in the intervention group and 34 students in the control group. These losses included transient students (those who moved in or out of study classrooms) and students whose special education status prevented them from participating. (See appendix D for an accounting of the loss of these students.) There were no teacher-level crossovers and no change in the number of participating teachers. There were, however, nine student-level crossovers (four students from intervention to control and five from control to intervention) who moved within the school district classrooms. The study received verification from each school principal that student crossovers were based on scheduling or other needs and did not switch classrooms in order to have access to Odyssey Math. Thus, decisions that created crossovers were independent of the random assignment of the teacher to the intervention or control condition. The nine student crossovers were included in the analysis in their originally assigned research condition.

Posttest phase

At the posttest stage of the study, there were 1,300 students in the intervention group and 1,284 in the control group. These numbers include students who had moved into the schools during the academic year (with parental consent). Thus, the analytic sample includes

students who moved to classrooms after random assignment, a group that was not pretested. (Additional details on handling this group are provided below.) Some students' special education status changed, but they remained in the study. The figures exclude students who were absent on the day of posttests and did not complete makeup tests.

Data analysis phase

At the data analysis stage the sample consisted of 60 teachers and 1,223 students in the intervention condition and 62 teachers and 1,233 students in the control condition (nested in 32 schools). The analytic sample had fewer students than the posttest sample because it included only students who completed both a pretest and posttest. Thus, at the teacher-classroom level (the level of random assignment) there was no attrition from pretesting to the final data analysis stage.

ATTRITION RATES

At study completion the overall student attrition rate was approximately 14 percent, and the differential attrition rate (between intervention and control classrooms) was approximately 2 percent (table 7). The overall and differential attrition rates were below the threshold planned for during the power analyses for this study, which was 20 percent. Again, there was no attrition at the level of random assignment (teacher-classroom level).[11]

More important, the overall attrition rates for schools, teachers, and students did not reduce statistical power to unacceptable levels because five more schools and 10 more teachers were recruited than required by the power analysis. The 2 percent differential attrition rate for the study is important because differential attrition has the potential to compromise the baseline equivalence established by random assignment and, as a result, to bias impact estimates.

Table 7. Attrition rates for intervention and control groups at teacher and student level

	Teachers			Students			
Data collection	Intervention group	Control group	Difference	Intervention group	Control group	Difference	Total
Random assignment; enrollment from rosters	60	62	na	1,448	1,492	na	2,940
Eligible sample	60	62	na	1,403	1,451	na	2,854
Pretest completed	60	62	na	1,322	1,318	na	2,640
Total analytic sample[a]	60	62	na	1,223	1,233	na	2,456
Attrition from eligible sample to analytic sample (percent)	0	0	0	12.8	15	2.2	13.9

a. Consisted of students who completed both the pre- and posttests.
Source: Authors' analysis based on data described in text.

---

[11] The attrition rates for the study do not include the school dropped from the study because it failed to report that it was already using Odyssey Math at the target grade. Had school personnel reported this fact, the school would have been ineligible to participate and its classrooms would not have been randomized to study conditions.

BASELINE EQUIVALENCE OF INTERVENTION AND CONTROL GROUPS

To evaluate whether random assignment resulted in statistically equivalent groups, the intervention and control groups were compared on important teacher and classroom baseline characteristics prior to intervention. These characteristics were hypothesized to be correlated with student achievement.

Baseline characteristics for 122 teachers and their 124 classrooms with 2,637 students that completed the pretest are displayed in table 8. Comparisons were made at the teacher level because that was the level of random assignment, and at this level random assignment is expected to equate groups on measured and unmeasured characteristics.[12] A t-test or chi-square test was used for the comparisons depending on the scale of the baseline characteristic (nominal or interval).

None of the 14 baseline characteristics compared was statistically different from zero at the $p < .05$ level. However, the number of long and short workshops was included as a covariate in the models as a sensitivity test because these variables were significant at $p < .10$.

Table 8. Mean baseline characteristics for intervention and control group teachers and classrooms

Baseline characteristics	Intervention group	Control group	Difference	Test statistic[a]	p-value
Teacher characteristics					
Years in current school	12.02 (sd = 10.56 n = 59)	9.79 (sd = 8.93 n = 58)	2.22	t = 1.23 (1.81)	.22
Years of teaching experience	16.95 (sd = 12.53 n = 59)	13.79 (sd = 10.26 n = 58)	3.16	t = 1.49 (2.12)	.14
Proportion with master's degree (percent)[b]	38.98 (sd = 49.19 n = 59)	36.67 (sd = 48.60 n = 60)	2.31	χ^2 = .07	.79
Previous professional development (past two years)					
Hours of university math course	5.98 (sd = 16.74 n = 58)	7.32 (sd = 14.56 n = 56)	−1.34	t = 0.45 (2.94)	.65
Hours of conferences or workshops on math					
Long training (more than half day)	8.68 (sd = 11.97 n = 56)	15.11 (sd = 21.52 n = 56)	−6.43	t = 1.95 (3.29)	.053
Short training (half day or less)	8.63 (sd = 13.57 n = 56)	14.32 (sd = 19.03 n = 57)	−5.69	t = 1.83 (3.11)	.07
Hours of math coaching received	4.72 (sd = 10.72 n = 58)	9.09 (sd = 16.67 n = 56)	−4.37	t = 1.67 (2.62)	.10

---

[12] The baseline data met standard statistical assumptions for t-tests: normally distributed with equal variances and no influential outliers.

Student characteristics					
Proportion of girls (percent)	50.60 (sd = 9.65 n = 60)	48.54 (sd =7.80 n = 62)	2.06	*t* = 1.36 (1.51)	.18
Proportion of racial/ethnic minority students (percent)[c]	25.37 (sd = 32.96 n = 43)	23.82 (sd = 31.65 n = 43)	1.55	*t* = 0.22 (6.97)	.82
Proportion of English language learner students (percent)	6.24 (sd = 18.79 n = 60)	6.74 (sd = 21.63 n = 62)	–0.50	*t* = 0.14 (3.67)	.89
Proportion of students eligible for free or reduced-price lunch (percent)	19.05 (sd = 21.78 n = 60)	16.90 (sd = 19.34 n = 62)	2.15	*t* = 0.58 (3.73)	.57
Student age (months)	115.63 (sd = 2.14 n = 60)	116.02 (sd = 2.86 n = 62)	–0.39	*t* = 0.85 (.46)	.40
Classroom average test score					
TerraNova Basic Battery math subtest	620.67 (sd = 15.49 n = 60)	621.19 (sd = 14.83 n = 62)	–0.52	*t* = 0.19 (2.75)	.85
TerraNova Basic Battery math subtest for students that completed the posttest	621.90 (sd = 14.40 n = 60)	622.44 (sd = 14.36 n = 62)	–0.54	*t* = 0.21 (2.60)	.84

Note: Although not displayed in the table, the number of students for the teacher classroom comparisons varied slightly depending on whether a characteristic was reported for a particular student. All statistics, including *p*-values, were rounded to two decimal places. Two of the 122 teachers taught two classrooms each, and for this table their classrooms were aggregated and reported as one classroom for each.

a. Numbers in parentheses are standard errors (for *t*-statistics) or degrees of freedom (for chi-square).

b. All teachers had a bachelor's degree, but no teacher had a Ph.D.

c. Students in some participating schools did not complete their racial/ethnic code during the pretest. Both the control and intervention classrooms within the school did not complete the information, so the report includes statistics for only 86 classrooms.

Source: Authors' analysis based on data described in text.

DATA COLLECTION INSTRUMENTS

This section discusses the study data collection instruments: student classroom rosters, TerraNova Basic Battery math subtest, test accommodations and scoring, teacher background survey, and classroom observation protocol.

Student classroom rosters

Student classroom rosters were the primary source of student and teacher data. Each roster included the name of the school district, school name, student name, student Odyssey Math username, and access status (active or inactive).

Math subtest of the TerraNova Basic Battery

The TerraNova Basic Battery was the only student outcome measure for this study. The Basic Battery edition consists of the reading/language arts subtest and the math subtest. According to the developer, each subset can be administered separately, and therefore only the math subtest was administered (CTB/McGraw-Hill 2000).

The math subtest's objectives reflect the National Council of Teachers of Mathematics standards (National Council of Teachers of Mathematics 2008) as well as state and local curriculum documents and the conceptual framework of the National Assessment of Educational Progress (National Assessment of Educational Progress 2008). The grade 4 math subtest consists of 57 selected-response items and takes 1 hour and 10 minutes to administer. Form A of the Basic Battery was administered as the pre- and posttest measures of math achievement, in accordance with the test developer's recommendation.[13] The internal consistency of the math subtest, as measured by the Kuder-Richardson formula 20 (KR20) coefficient, is .93 with a standard error of measurement of 3.13. This information is based on a standardized national sample reported by CTB/McGraw Hill (2000). The Cronbach coefficient alpha reported for the sample at pre- and posttest is .91.

Test accommodations and scoring

According to the publisher, a series of test accommodations are designed to assist test users with administration and explain the implications of these accommodations for interpreting test results. However, no special accommodations were required in this study except extra time for special education students (fewer than three students for each participating school). Norms, updated in 2005, are representative of the K–12 student population and include students with disabilities and English language learner students. These norms were used to interpret the test scores.[14] To ensure accuracy, the CTB/McGraw-Hill scoring service (which considers test accommodations) was used to score the grade 4 math subtest. Complete test score data files were returned in ASCII format and included selected student demographic information such as gender, date of birth, and student ID numbers.

Teacher background survey

Designed by the REL study team, the teacher survey consisted of five questions used to collect data about teachers' experiences, degrees, professional development, and experience with computer software (see appendix E for the survey).

---

[13] When using the same form for pre- and posttest the test developer recommended that there be at least six months between a pretest and a posttest administration. Additional documentation is available from the developer.

[14] The 2005 norms are an update of the published 2000 norms using a combination of the 2000 standardization data and customer data from 2001 and 2005 to adjust for two factors: the changing demographic composition of the public school student population and instructional intervention programs, which have altered student performance since they were observed in 2000.

Classroom observation protocol

Observations were conducted using a modified version of the standards observation form (Stonewater 1996). The protocols were designed to document how consistent classroom instruction was with National Council of Teachers of Mathematics (NCTM) standards. Math content experts at Pennsylvania State University updated the protocols to address NCTM standards revisions since the original standards observation form was developed 10 years earlier.

Two versions of the protocol were created, one to document observations in intervention classrooms and one to document interventions in control classrooms (see appendix F). Both protocols had three sections. The first section in both protocols documented the classroom environment with short answers from the observer on such matters as number of students, number of students with access to computers, and whether the class period was dedicated to math instruction or included other activity.

The second section in both protocols contained questions on teacher–student interactions rated on a scale of 1–5 (1 being least favorable, 5 being exceptional) and with short answers from the observer. This section focused on the types of questions students were asking and on teacher responses.

The third section focused on the math content and instructional practices observed. The focus in the control group observation protocol was on the learning objectives and the instructional practices observed. The observer noted the name of any software used and how it was used in the classroom. In the intervention observation protocol, the focus was on the learning objects within Odyssey Math. Again, the observer noted what learning activities and assessments were used and how they were used.

DATA COLLECTION PROCEDURES

This section discusses the study data collection procedures for classroom rosters, teacher and school characteristics, site visits to test software, classroom observation, and student data.

Student classroom assignments and rosters

After random assignment, invitations were mailed to intervention teachers for one of five regional summer 2007 professional development sessions led by CompassLearning. Attendance was confirmed through follow-up telephone calls.

Classroom rosters were collected in August 2007 before notification of random assignment. The rosters and student classroom assignments were verified during the pretesting session and served as the primary source of student and teacher data for the analytical sample.

Teacher and school characteristics

Intervention classroom teachers completed the teacher demographics survey during the professional development sessions conducted in the summer of 2007 after completing the consent forms. The surveys were mailed to the control classroom teachers and collected during the pretesting sessions in the schools in September–October 2007. The survey completion rate was 97.5 percent (3 of the 122 participating teachers did not complete the survey). School characteristic data were collected from the School Data Direct web site (School Data Direct 2009).

Site visits to test software and student software use

Members of CompassLearning's technical group conducted site visits at each school selected for this study to test schools' computer laboratories with the Odyssey Math software, which runs from a central server (A. Manilla, CompassLearning educational consultant, personal communication, August 2, 2007). Tests were conducted for bandwidth and availability of necessary software and hardware. The 32 participating schools were all found to have the hardware and software needed for typical implementation of Odyssey Math (CompassLearning 2008b).

All students in the intervention condition were assigned a username and password for the Odyssey Math software. The software logged each student's activity on the system, and the study team downloaded access reports monthly.

Classroom observations

Observations were conducted using a modified version of the Standards Observation Form (Stonewater 1996). The protocols were designed to document how consistent classroom instruction was with National Council of Teachers of Mathematics (NCTM) standards. Math content experts at Pennsylvania State University updated the protocols to address NCTM standards revisions because the original standards observation form was developed 10 years earlier.

Observing intervention implementation

The study team observed implementation of the intervention during one full class period in each intervention classroom at approximately the midpoint of the school year (December 2007–February 2008). Classroom observations were conducted during the same timeframe in control classrooms to better understand the counterfactual and to describe the curriculum and practices used. Separate observation protocols were used for the intervention and control classrooms, as described above.

Collecting student achievement data

The TerraNova Basic Battery math subtest was administered during September–October 2007 (pretest) and April–May 2008 (posttest) under similar settings for intervention and control conditions within each school (such as a quiet auditorium or cafeteria). Two

trained study team members administered the student informed consent forms and tests in the presence of teachers, following written guidelines prepared by the principal investigators. Written test-taking instructions were read to the students.

If more than two students were absent at the pretest in a school, the test administrators conducted makeup sessions in some schools. Because of budget considerations, pretest makeup sessions were not held at all schools. However, posttest makeup sessions were held in all schools with more than two student absences.

DATA ANALYSIS METHODS

The primary focus of this report is an intent-to-treat analysis of a single confirmatory question that included all originally assigned teachers. The confirmatory question was addressed using the following approaches:

- Unadjusted mean differences between intervention and control classrooms.
- Application of multilevel models (hierarchical linear models), with and without pretest covariates.
- Two sensitivity analyses that handle missing data.

To empirically address the confirmatory research question for this study, a multilevel model was used to estimate the intervention's effects and test the statistical hypotheses. Model parameters were estimated for empirical and statistical reasons (Luke 2004). Because students were nested within teachers, and teachers were nested within schools, students in the same teacher's classroom were more likely to have similar math achievement scores than were students in different teachers' classrooms. For the same reason, student math achievement scores aggregated to the teacher level were more likely to be similar within schools than between schools. Statistically, unlike conventional least squares or ordinary least squares regression analysis, multilevel models take the nested structure of the data into account by allowing error structures to be correlated (whereas ordinary least squares assumes that these errors are independent), thus generating more accurate standard errors for impact estimates.

Multilevel models also allow for impact estimates at the teacher level to vary randomly across schools. A significant variation in impact estimates across schools would suggest a differential effect of Odyssey Math depending on the school. The power analysis presented earlier was conducted for a random intervention effects model to ensure sufficient power to detect a minimum effect size of 0.20 (see appendix B).

The multilevel model

This section describes the multilevel model that was estimated to answer the confirmatory question:

- Do grade 4 classrooms using Odyssey Math as a partial substitute for the standard math curriculum outperform control classrooms on the math subtest of the TerraNova Basic Battery in a typical school setting?

First, simple differences were calculated, without adjusting for covariates, between the intervention and control classrooms on average pretest and posttest scores. These differences were tested for statistical significance with standard errors that took into account the nested data structure. The mean difference between the intervention and control classrooms on the posttest scores gave an initial impact estimate prior to estimating impact using the full multilevel model with covariates and random coefficients.

Second, the full three-level model was estimated with students at level 1, teachers at level 2, and schools at level 3. The model was specified using Raudenbush and Bryk (2002) nomenclature.

Level 1 (student level)

$$Y_{ijk} = \pi_{0jk} + e_{ijk}$$

where Y_{ijk} is the outcome for student i in teacher j's class in school k, π_{0jk} is the average outcome of students in teacher j's class in school k, e_{ijk} is a random error associated with student i in teacher j's class in school k, and $e_{ijk} \sim N(0, \sigma^2)$.

The classroom average outcome in a school estimated by the level 1 intercept π_{0jk} was modeled as varying randomly across teachers and as a function of the intervention (partial substitution of Odyssey Math software for regular math instruction) at level 2, the teacher level, controlling for the classroom average pretest scores on the TerraNova Basic Battery subtest.[15]

Even though intervention and control groups were formed using random assignment, there is always a chance that a particular sample may have a statistically significant difference on some measured characteristic at baseline. To control for this possibility, related covariates (a baseline imbalance covariate) were included at the teacher level. However, no statistically significant imbalance was found between intervention conditions on any baseline characteristic (see table 8). Thus, level 2 was specified as shown below.

Level 2 (teacher level)

$$\Pi_{0jk} = \beta_{00k} + \beta_{01k} (Odyssey)_{jk} + \beta_{02k} (Pretest)_{jk} + r_{0jk}$$

where β_{00k} is the adjusted average student outcome across all control teachers' classrooms in school k, β_{01k} is the adjusted difference in student outcome between the

---

[15] The inclusion of a pretest covariate typically yields improved statistical precision of the parameter estimates (Bloom, Richburg-Hays, and Black 2007; Raudenbush, Martinez, and Spybrook 2005).

intervention teachers' classrooms and the control teachers' classrooms (intervention effect) in school k, *Odyssey* is an effect indicator variable for the intervention that takes a value of 1 for an intervention teacher's classroom and 0 for a control teacher's classroom, B_{02k} is the effect of the mean classroom pretest score on classroom average student outcome in school k, r_{0jk} is a random error associated with teacher j's classroom in school k on classroom average student outcome $r_{0jk} \sim N(0, \tau_{\pi00})$, and *Pretest* is the classroom grand mean–centered average pretest score.

Level 3 (school level)

In the level 3 model both the school average outcome (β_{00k}) and the intervention impact in each school (β_{01k}), estimated from the teacher-level model, were modeled as random effects. There are two analytic benefits to modeling the intervention effect as random. One is that the intervention could have a positive effect on some schools but not on others. Treating the intervention effect as random would reveal any such variation across schools, whereas in a fixed effects model positive and negative effects on individual schools might cancel each other out and show no overall significant intervention effect. A second benefit is that if the random effects model reveals no significant variation in intervention effect across schools, the treatment effect could be interpreted as being consistent across schools and so more likely to generalize to schools with characteristics similar to those in the analytic sample.

Assuming that the coefficients for classroom average pretest were homogeneous across schools, the effect of *Pretest* was fixed at the school level, as shown in the following specification:[16]

$$\beta_{00k} = \gamma_{000} + u_{00k}$$

$$\beta_{01k} = \gamma_{010} + u_{01k}$$

$$\beta_{02k} = \gamma_{020}$$

where γ_{000} is the adjusted average student outcome in the control condition across all schools, u_{00k} is a random error associated with school k on adjusted school average student outcome $u_{00k} \sim N(0, \tau_{\beta00})$, γ_{010} is the average intervention effect across all schools after controlling for differences in pretest scores, u_{01k} is a random error associated with school k on the intervention impact $u_{01k} \sim N(0, \tau_{\beta11})$, and γ_{020} is the average effect of *Pretest* on student outcome across all schools.

---

[16] Because no imbalances between intervention and control groups were found on baseline characteristics, only *Pretest*, which was supposed to be highly correlated with the outcome measure and hence would increase statistical power, was retained as a covariate. An alternative model with *Pretest* included as a level 1 covariate was also analyzed, but as is shown in the results section, this did not increase statistical precision nor did it alter the interpretation of the estimate of the effect of Odyssey Math on student achievement.

Of primary interest among the level 3 coefficients was γ_{010}, which represents the intervention's main effect on the outcome across all schools. A statistically significant positive value of γ_{010} would be reason to reject the null hypothesis of no difference between intervention and control groups in favor of the alternative hypothesis that students in the intervention teachers' classrooms demonstrate higher levels of math achievement than do their counterparts in the control teachers' classrooms. A multilevel model 6 (Raudenbush, Bryk, and Congdon 2008) was used to analyze all the multilevel models with the default maximum likelihood estimator for three-level models.

In addition to the statistical significance of the effect of the Odyssey Math intervention, the magnitude of the effect was also expressed in standard deviation units. Specifically, the effect size was computed as a standardized mean difference (Hedge's g) by dividing the adjusted group mean difference (γ_{010}) by the pooled within-intervention and control group standard deviation of the student-level outcome score. Glass's delta was computed by dividing the adjusted group mean difference by the control group standard deviation of the student-level outcome score. Large differences between the two effect size measures would indicate an intervention effect on the variability of the student outcome because both measures simply divide the same numerator (γ_{010}) by different standard deviations (either for the pooled across intervention and control groups or for the control group).

Sensitivity analyses

Random and fixed effects models. To evaluate how sensitive the impact estimate (or treatment effect) and standard error are to the decision to model school effects as random in the core analysis, a sensitivity analysis was conducted by estimating a series of fixed effect models:

- A two-level model with students at level 1 and classrooms at level 2, as specified previously, but with the impact estimate (or treatment effect), β_{01k}, modeled as fixed across schools (a two-level model estimated without the school level); however, clustering due to schools was disregarded.
- A two-level model with students at level 1 and classrooms at level 2, as specified previously, but with the impact estimate (or treatment effect), β_{01k}, modeled as fixed and school effects modeled as fixed by including $Z - 1$ dummy variables (where Z is the total number of schools in the sample) at the classroom level.

Pretest covariate at different levels of model. Achievement pretest scores were a student-level variable aggregated to the teacher-classroom level as a grand mean–centered covariate in the model for the core analysis to address the confirmatory question. These scores can be used as a level 1 covariate instead of using the classroom mean score as a level 2 covariate. This alternative model with the grand mean–centered student achievement pretest score entered at level 1 and the classroom study condition (1 = intervention and 0 = control) entered at level 2 with random intervention effect and random intercepts was fitted

to evaluate how sensitive the impact estimate was to placement of the pretest score at level 1 rather than at level 2.[17]

Group differences on baseline covariates. Any baseline variables that were not statistically significant at $p < .05$ but were at $p < .10$ were included in the multilevel model as a sensitivity analysis. Specifically, each variable was included in the multilevel model (grand mean centered) as a teacher-level covariate in addition to the pretest classroom mean covariate (grand mean centered) to address the confirmatory research question. This analysis indicated whether the estimate and statistical significance were sensitive to excluding these variables from the model.[18]

Missing data

Two approaches were used to handle missing data: listwise deletion and dummy variable adjustment. The listwise deletion was used as the primary approach and a dummy variable adjustment as a sensitivity analysis.

Listwise deletion. Listwise deletion was used for missing data at the student level for four reasons. First, the study design planned for a 20 percent attrition rate. Any attrition rate greater than 20 percent would result in statistical power of less than .80 (for an assumed minimum detectable effect size of 0.20). Student-level attrition was only 13 percent and therefore did not result in a reduction in statistical power (see appendix B for power analysis assumptions). Second, the teacher-classroom was the level of random assignment, and there were no missing data at that level. Thus, there was no evidence that the impact estimate was biased at the level of random assignment due to attrition.

Third, and most important, based on conversations with school principals during pre- and posttesting, a reasonable assumption was that test data were missing completely at random in both the intervention and control groups. In other words, the probability that a student did not take the pre- or posttest was unrelated to treatment condition, teacher characteristics, or any other variables in the multilevel model but was due to such causes as illness or family trips. When data can be assumed to be missing completely at random, Allison (2001, p. 7) demonstrates empirically that listwise deletion produces statistically unbiased estimates of effect and is thus the best method for dealing with missing data.

Finally, there are several other advantages in using listwise deletion. It can be used for any type of statistical analysis. No special computational methods are needed. Bias is often minimal when pretest variables are included in the model as covariates (Graham 2009). And the most serious penalty for its use, loss of sample size, is transparent. Even if the weaker assumption of missing at random were invoked because the assumption of missing completely at random was considered too strong, the limited amount of missing data

---

[17] As is shown in the results section, this did not increase statistical precision nor did it alter the interpretation of the estimate of the effect of Odyssey Math on student achievement.
[18] As is shown in the results section, this did not increase statistical precision nor did it alter the interpretation of the estimate of the effect of Odyssey Math on student achievement.

combined with the low level of differential attrition across intervention and control conditions still suggests that listwise deletion is a reasonable choice.[19]

Thus, although there are other techniques that could have been used such as nonresponse weighting adjustments and multiple imputation, analyses based on listwise deletion were sufficient because statistical power was not reduced below .80 and the low (statistically nonsignificant) differential attrition across study conditions did not threaten the validity of the impact estimate.

Dummy variable adjustment. A sensitivity analysis was conducted to determine how sensitive the impact estimate was to missing pretest data. Students who completed the posttest but not the pretest were included in the model with grand mean or class mean pretest scores substituted for missing pretest data. A missing dummy indicator (with 1 = pretest score absent and 0 = pretest score present) was used to adjust for the effect of missing pretest scores. Both student pretest scores (grand mean centered) and the missing dummy indicator were entered as level 1 covariates. As in the model used to generate the impact estimate for the core analysis, class mean pretest score (grand mean centered) was entered as a level 2 covariate, the intervention group indicator was included in level 2 (classroom level), and a random intervention effect was estimated.[20] These two models were estimated with the dummy variable indicator for missing data but differed in the choice of mean substitution for the missing pretest score to test whether the impact estimate was invariant to the choice of the substitute mean (classroom or grand mean) for the unobserved (or missing) pretest score as part of the dummy variable adjustment.

Students missing posttest scores were deleted from the analysis, even if they had pretest scores.

---

[19] Among the missing data techniques explored by Allison (2001), listwise deletion is the most robust to violations of the missing at random assumption in regression models. However, it is not clear from his work whether this extends to random coefficient regression models such as multilevel models.

[20] As is shown in the results section, this did not increase statistical precision nor did it alter the interpretation of the estimate of the effect of Odyssey Math on student achievement.

3. IMPLEMENTATION OF
THE ODYSSEY MATH INTERVENTION

This chapter covers implementation of the Odyssey Math intervention. It describes the full CompassLearning Odyssey® software package and its Odyssey Math component, and the various professional development packages available from CompassLearning, including the professional development option selected for the study and the rationale for its selection. It also presents statistics on the actual use of Odyssey Math by students in the study and summarizes the observations of intervention and control classrooms.

ODYSSEY PRODUCT OPTIONS AND
THE ODYSSEY MATH COMPONENT SELECTED FOR THE STUDY

The CompassLearning Odyssey software package provides access to language arts, math, science, social studies, brain buzzers, thematic projects, and language arts extensions (see exhibit G1 in appendix G for a sample screen of the student launch pad from the CompassLearning Odyssey software package). The CompassLearning Odyssey software package also contains instruction, activities, and assessments to support K–12 students.

This study focused on the grade 4 Odyssey Math portion of the full CompassLearning Odyssey software package, for the reasons presented in chapter 1. Although the intervention teachers and students had access to the full CompassLearning Odyssey software package, teachers were instructed during professional development to use only the Odyssey Math link. Monthly reviews of the CompassLearning software computer logs showed that all users followed these instructions. In addition, Odyssey Math software for grades 3 and 5 were made available to intervention teachers to facilitate their tailoring of instruction. The grade 3 package could be used for remediation purposes and the grade 5 package for advanced instruction.

The use of the Odyssey Math software required a computer for each student and headphones for the multimedia presentations. Each teacher and student had a unique username and password to access the software.

Although a search of CompassLearning's materials do not suggest a specific theory of change, the developer indicates that teachers who use Odyssey Math will have access to instructional techniques such as using on-screen manipulatives, using formative assessment to monitor student progress toward learning objectives, providing related feedback, and generating individualized instructional plans to provide a form of instructional scaffolding. CompassLearning reports that its professional development for teachers focuses on developing skills such as applying individualized, scaffolded assignments that can be incorporated in overall lesson plans, as noted in appendix A.

The following paragraphs describe what a typical student might have seen during an Odyssey Math lesson. (For a sample learning activity screen on two-digit divisors, see exhibit G2 in appendix G.) They showcase the content, student interactions, assessment, and feedback associated with a lesson on number theory and systems, with four subactivities (shown in exhibit G3). The example includes descriptions of software presentations made to students for correct and incorrect item responses.

Selected lesson

The first screen of the selected lesson from a series on number theory and systems presents a lesson on standard and expanded form and offers a text description, three activities, and a quiz.

Text description: "Convert numbers containing two to nine digits from standard form to expanded form and vice versa."

Activity 1: standard exchange. The first activity, a pre-lesson activity, begins with a timed "matching game" (exhibit 1). The game area is a four-by-four group of blank squares. If the student clicks on the "How to play" button, the web page displays the following directions: "Click on boxes to match each number to its name." Two squares can be clicked on at a time to reveal their contents. If the two revealed squares match, they turn into parts of a picture. If the squares do not match, they turn back into blank squares. Play continues until the timer runs out or all squares are revealed. The lesson then proceeds automatically.

The first page of the lesson offers a graphic with the lesson outline and a button the student can click on to proceed.

The next display is the "Galactic Arcade," with a "ticket exchange booth" that allows students to exchange tickets for virtual prizes. Narration explains: "You are needed at the ticket exchange booth. Some kids want to cash in their tickets for prizes."

The next display shows and narrates an example of converting a number from standard to expanded form (exhibit 2) and explains a place-value chart (for example, the place values for the digits in the number 6,503,825, where 6 is depicted as a value in the millions, 5 as a value in the hundreds of thousands, and so on). Then the ticket booth displays a number in standard form, and the student is to re-create the number in expanded form by clicking on arrows. Students click on a button labeled "exchange" to submit their answer. If the answer is correct, a graphic pops up depicting the student receiving a prize. If the answer is incorrect on the first try, an example is displayed. Following a second and third incorrect response, a pop-up window shows a place-value chart. After a third incorrect attempt, the correct answer is filled onto the ticket booth and the student is prompted to move onto the next question. There are six questions in this lesson.

Exhibit 1. Pre-lesson activity "matching game"
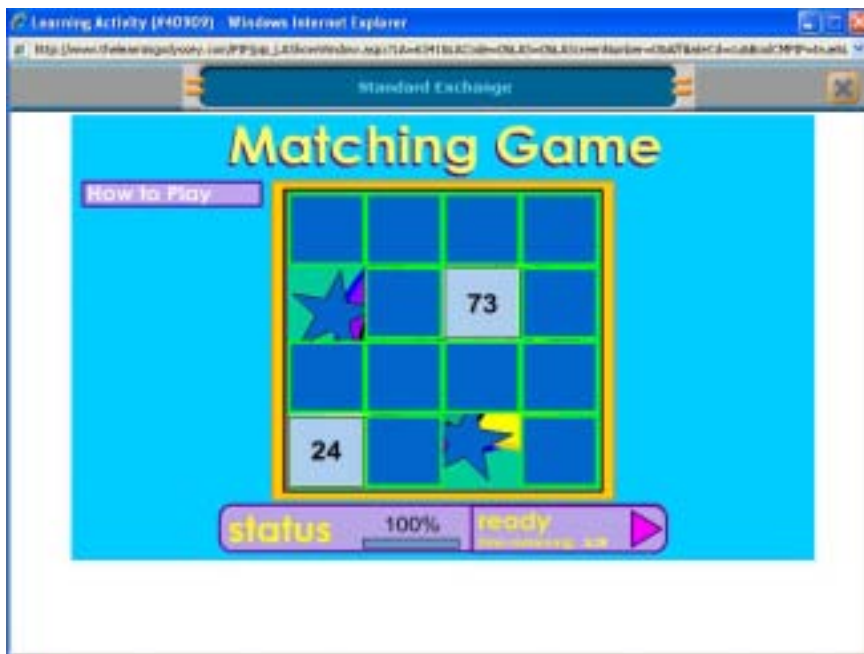
Source: CompassLearning Odyssey Math®.

Each lesson also has a navigation bar in the bottom right corner (see exhibit 2). This bar includes a graphic that charts the student's progress, a button that repeats the last narration, a button that repeats the lesson portion of the activity, a button that gives the student another look at the topic lesson, and a button that lets the student move forward in the lesson.

Exhibit 2. Standard and expanded form of numbers
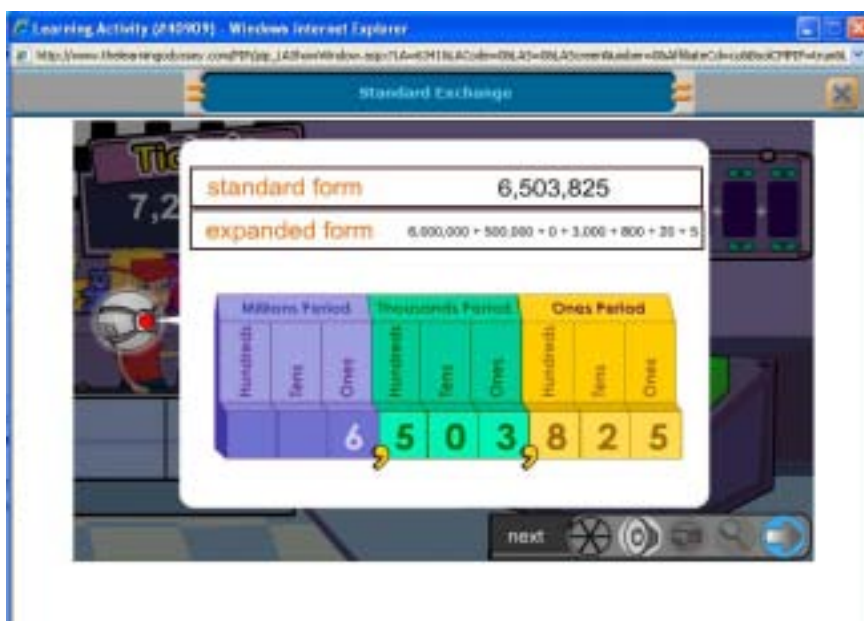
Source: CompassLearning Odyssey Math®.

Activity 2: expanded form exploratory. The next activity is an unstructured learning exercise with six activities (exhibits 3 and 4). Answers are not scored. Students can view the correct answer by clicking on the key icon at the bottom of the answer area. The help button gives generic directions for the activity. Students either type in a box or click on numbered boxes to answer the questions.

Exhibit 3. Expanded form exploratory
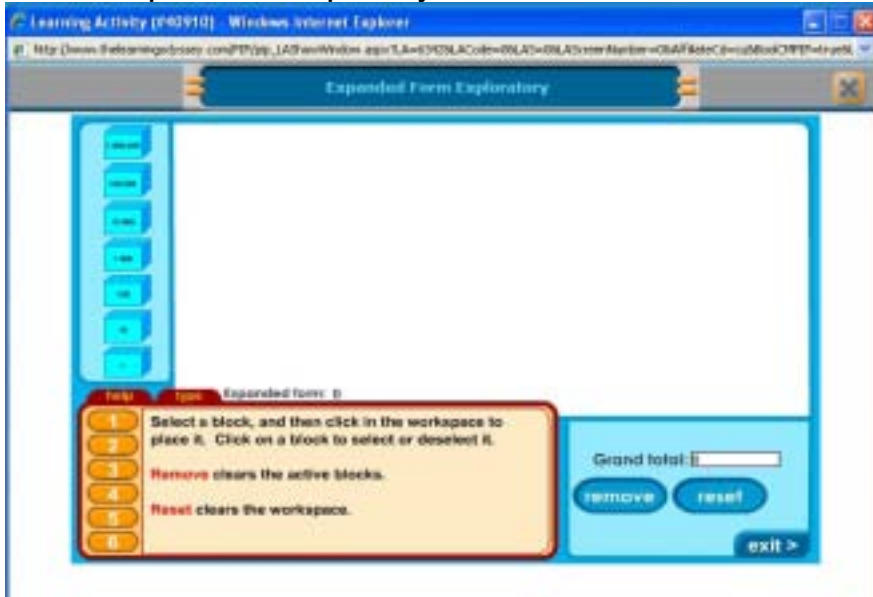
Source: CompassLearning Odyssey Math®.

Exhibit 4. Expanded form exploratory activity with student response
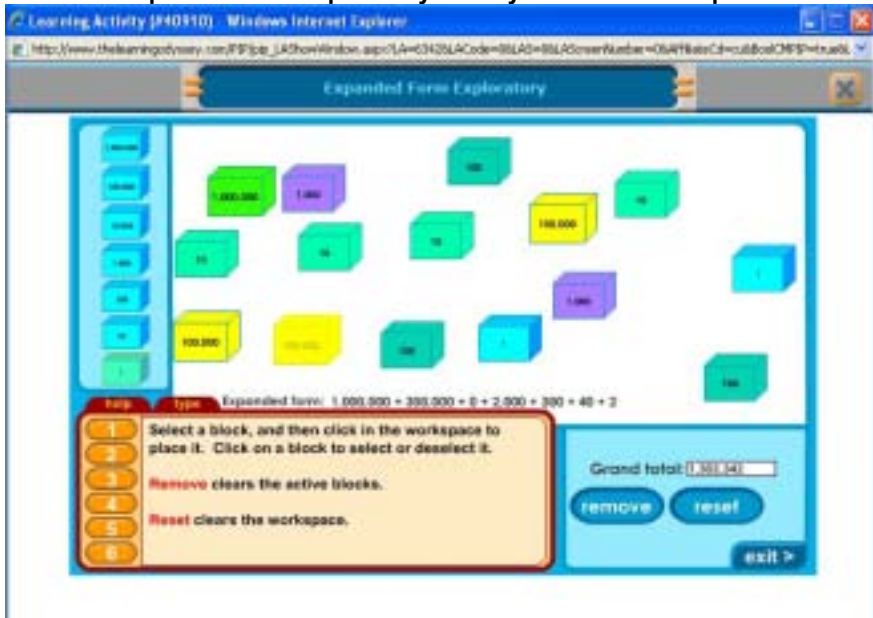
Source: CompassLearning Odyssey Math®.

Activity 3: expanded form handbook. This activity is an in-depth explanation of converting from standard to expanded form (exhibit 5). Explanations are given for the student to read (not narrated), then students are asked to answer questions by choosing from a dropdown list. Feedback is given through a pop-up window that tells students whether the answers are correct (exhibit 6).

Exhibit 5. Expanded form handbook

Source: CompassLearning Odyssey Math®.

Exhibit 6. Depiction of feedback for a correct answer to an assessment item

Source: CompassLearning Odyssey Math®.

At the end of the lesson students are given a multiple-choice quiz on standard and expanded form (exhibit 7).

Exhibit 7. Standard and expanded form quiz
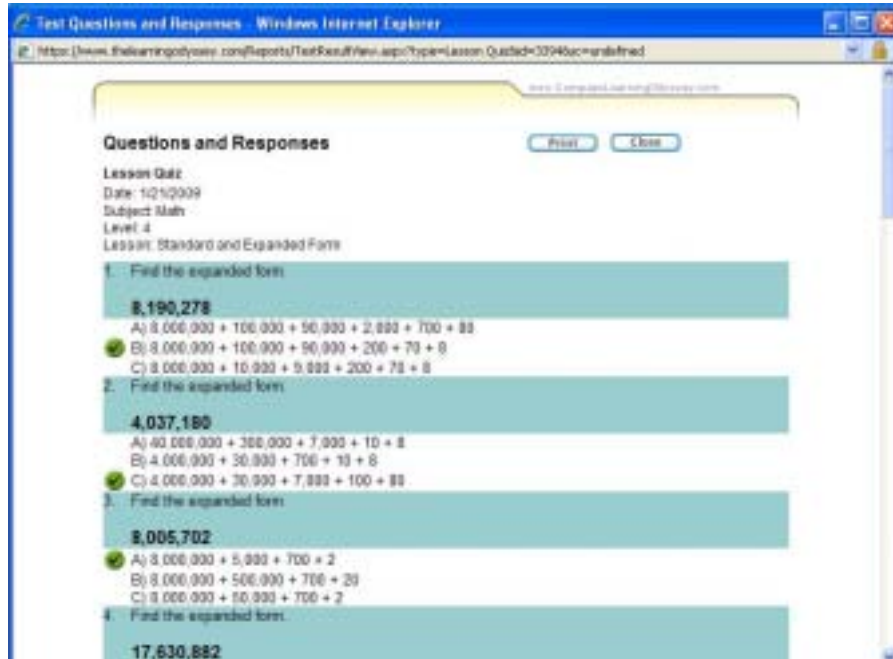
Source: CompassLearning Odyssey Math®.

Alignment of Odyssey Math with state and national standards

Odyssey Math software allows teachers to choose activities such as the ones presented above for students to practice. The software has built-in assessments and multimedia capabilities. The developer's web site states that "CompassLearning's research-based Odyssey curriculum is aligned with state and national standards and provides a stimulating learning environment. A variety of instructional approaches supports multiple learning styles and levels of achievement" (CompassLearning 2008b). On request, CompassLearning provided documentation showing the alignment of the Odyssey Math curriculum with state standards in Delaware, New Jersey, and Pennsylvania.

ODYSSEY MATH PROFESSIONAL DEVELOPMENT PACKAGE

CompassLearning offers several professional development packages to train teachers in Odyssey Math software. According to the developer, schools may purchase 6, 12, or 24 "days" of professional development based on the subjects and the number of grade levels using the Odyssey software. The five-day professional development package was selected because the study focused only on the Odyssey Math subset of the Odyssey suite and only on one grade level. The 12- and 24-day packages are used to support the full range of subjects in Odyssey and also a larger range of grades.

Two large group professional development sessions were offered to the intervention teachers and any school administrators who wanted to attend (table 9; appendix A presents the detailed agenda for the professional development sessions). The first large group session, over two calendar days in August 2007, was offered in four regional locations and attended by 37 teachers. Makeup sessions were offered to teachers who could not attend the initial scheduled sessions. The second large group professional development session was offered for one calendar day in January 2008. These large group sessions were followed by one-on-one coaching sessions with intervention teachers in their classrooms. All intervention teachers received the Odyssey Math professional development in addition to their regular professional development opportunities.

Table 9. Description of professional development offered to intervention teachers

Professional development "day" number[a]	Type of setting	Month and duration	Number of attendees	Content[b]
1 Makeup "day"	Large group instruction in computer labs at universities in Altoona and Scranton, Pennsylvania, and Rutgers, New Jersey In-school "day"	August 2007 • Day 1: 5 hours • Day 2: 3 hours • Compressed to 1 full day	• 37 intervention teachers and 4 administrators • 2–4 members of the study team • 23 intervention teachers • 1 member of the study team	• Student launch pad • Overview of curriculum, tests, and assessments • Student launch pad • Overview of curriculum, tests, and assessments
2	In-school, one-on-one coaching	October–November 2007 • 1–2 hours	• 60 intervention teachers	• Startup, management, logistics
3	Large group instruction in computer labs at universities in Altoona, Beaver, and Scranton, Pennsylvania; and New Brunswick, New Jersey	January 2008 • 6 hours	• 60 intervention teachers • 2–3 members of the study team	• Incorporating Odyssey Math in lesson plans
4	In-school, one-on-one coaching	February 2008 • 1–2 hours	• 60 intervention teachers	• Developing assessments and reports
5	In-school, one-on-one coaching	March 2008 • 1–2 hours	• 60 intervention teachers	• Scaffolding assignments and tailoring to individual students

a. The developer uses the term "day" for financial accounting purposes and not to describe actual instructional contact time between CompassLearning staff and teachers. A "day" is roughly the amount of time the developer needs to prepare and deliver the intended curriculum. Summer training "days" average 5–6 hours of training time. Coaching "days" average 1–2 hours of instruction for an individual teacher.
b. The complete agenda for the professional development sessions are shown in appendix A.
Source: Authors' compilation.

MATH INSTRUCTIONAL TIME

The study encouraged equivalent total instructional time in math across intervention and control classrooms, communicated in writing through the memorandum of

understanding and consistently throughout the study to CompassLearning and school personnel. However, the study team did not verify this expectation empirically.[21]

In the memorandum of understanding participating schools also agreed to use the software for approximately 60 minutes each week, and CompassLearning professional development trainers instructed the teachers about the 60-minute usage.

Implementation in intervention classrooms was measured as Odyssey Math usage time by students, which was tracked through software access logs. Since this was an effectiveness trial, the study team reported any low usage rates to CompassLearning personnel to enable them to address problems that might inhibit typical implementation (such as technology problems and miscommunication around expectations). The developer reported that having access to this data did not alter their standard practices during the study.

At the classroom level the mean usage time was 754 minutes and the standard deviation was 343 minutes with a maximum time of 1,450 minutes. Student-level time on Odyssey Math ranged from 0 to 1,918 minutes, with a standard deviation of about 370 minutes and a mean of 749 minutes (approximately 38 minutes each week on average based on 20 weeks of implementation, below the expected 60 minutes.)

Figure 2 shows monthly mean usage time for each intervention teacher's classroom.

Figure 2. Average total time on Odyssey Math per month by classroom, October 2007–April 2008
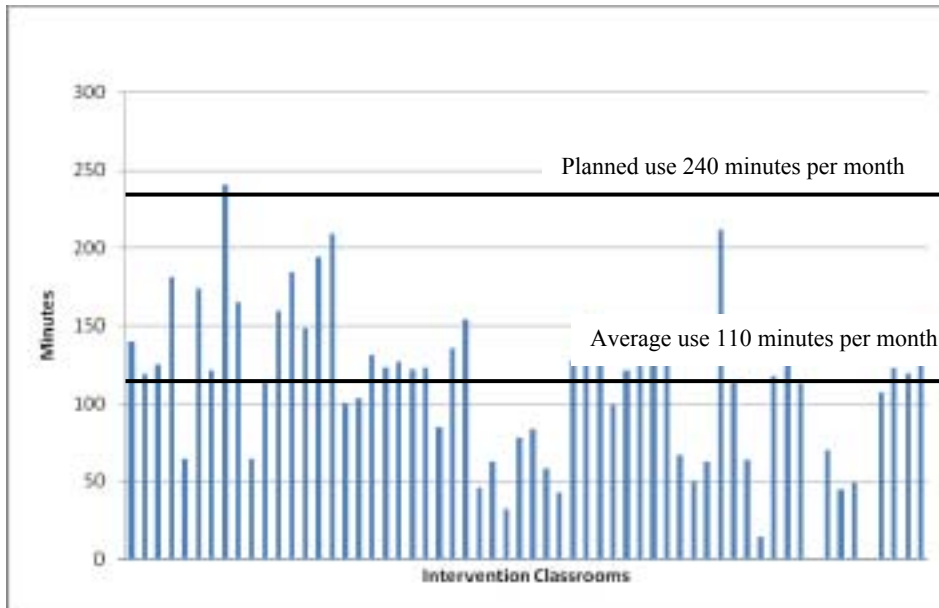
Source: Authors' analysis using data from end-of-year backup of the Odyssey Math log created by CompassLearning.

---

[21] Three fidelity observations were planned to document the math instructional time, but because of high costs only one observation was conducted in each classroom. During this observation the math instructional time was the same in intervention and control classrooms in the same school.

Figure 3 shows average monthly time on Odyssey Math over the October 2007–April 2008 intervention period.

Figure 3. Average total time on Odyssey Math by month during 2007/08 school year
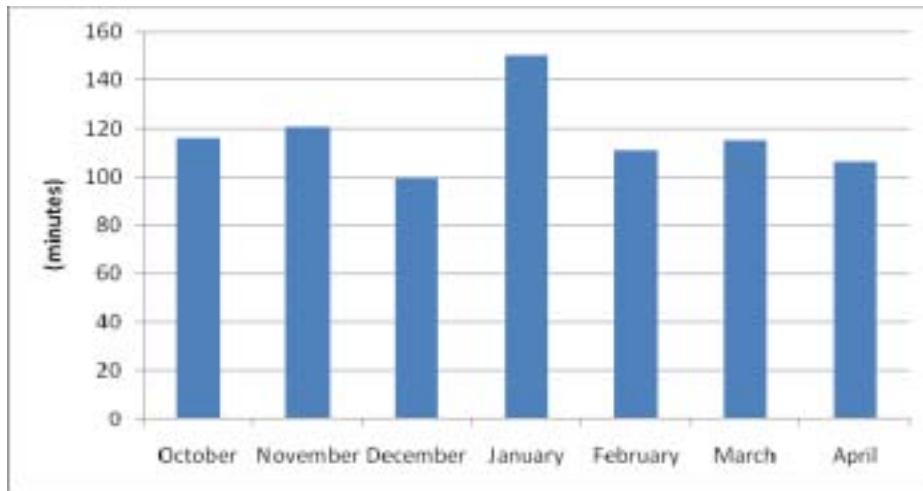
Source: Authors' analysis using data from end-of-year backup of the Odyssey Math log created by CompassLearning.

The mean usage time ranged from 0 to 240 minutes. One teacher maintained the prescribed level of usage at 240 minutes for the month (60 minutes each week). Two intervention teachers are shown with 0 minutes using Odyssey Math (fifth and ninth position from the right in figure 2). One teacher did not carry out the intervention after participating in the summer training but did allow pre and posttest student data to be collected. Students in this classroom were still considered intervention participants and were thus included in intent-to-treat analyses, which yielded the primary findings presented in chapter 4.

The other teacher showing no usage time in the intervention condition used paper versions of the Odyssey Math program instead of the web-based software. The CompassLearning team was consulted in conference calls and through email, and the study team was assured that this is typical of some implementations of the software (A. Manilla, CompassLearning educational consultant, personal communication February 5, 2008). This decision produced a slightly downward bias on usage times reported above, but otherwise did not affect the analyses. The teacher was treated as an intervention teacher because, again, the developer considers paper-based implementation to be a legitimate approach for Odyssey Math use.

During implementation the study team downloaded the monthly software usage report (shown in figure 3) and reviewed the logged times, monitoring progress and notifying the developer of the usage statistics. The CompassLearning team assured the study team that the professional development instructors assigned to each teacher would follow up during the four in-school coaching sessions and remind the teachers of the planned 60-minute usage time. CompassLearning also regularly noted that reported usage times were typical of routine

implementation (A. Manilla, CompassLearning educational consultant, personal communication, January 9, February 13, and March 12, 2008).

In summary, the Odyssey Math usage time varied by intervention classroom and by month across intervention classrooms and did not meet the average usage time prescribed by the study. As one aim of this study was to estimate the impact of Odyssey Math under typical implementation conditions, the study team took no additional steps beyond providing the monthly reports to persuade the CompassLearning implementation coaches to intervene with teachers to increase the time on task. Thus, the study team concluded that the study impact estimates (chapter 4) measure the impact of Odyssey Math with usage times that varied and were under the prescribed rate but that were considered typical of the implementation of the program.

CLASSROOM OBSERVATIONS AND FIDELITY OF INTERVENTION IMPLEMENTATION

The study team conducted 118 observations in intervention and control classrooms. Four additional planned observations of intervention classrooms did not occur because of scheduling inconsistencies. All observational data were used for descriptive purposes by providing context for the impact estimates described in chapter 4.

A total of 18 students were not using headphones, either by choice or because the headphones were missing or not operating properly. Headphone use is a required hardware component for some Odyssey Math applications, and failure to use them can contribute to a noisy classroom environment. Other problems noted during classroom observations were poor Internet connectivity and missing software components ("plugins").

The observations documented that nine curricula were being used by the 32 participating schools (control and intervention teachers in these schools used the same main curriculum). Table 10 documents the four curricula used in 27 of the 32 study schools.

Table 10. Regular curricula in use in participating schools

Regular curriculum	Number of schools
Everyday Math (Everyday Math 2009)	10
Scott Foresman (Pearson 2009)	7
Harcourt Brace (Harcourt School 2009)	5
Saxon Math (Saxon 2009)	5

Source: Authors' compilation based on study team classroom observations.

Since the within-school random assignment of classrooms ensured that both the intervention and control classrooms within each school followed the same math instructional curriculum, the difference between the intervention and control classrooms was the use of Odyssey Math.

Teachers were not instructed on what part of the regular math curriculum to replace with Odyssey Math. Teachers could substitute Odyssey Math for any combination of the

following: traditional practice tasks (for example, hands-on activities using a ruler), assessment, or whole instructional modules.

The Everyday Math curriculum (http://everydaymath.uchicago.edu/about/) used in the greatest number of participating schools reports similar instructional goals as Odyssey Math. The approach differs from that of Odyssey Math in that the teacher presents the instruction and the learning modules using materials in the classroom. Everyday Math uses real-life examples to present the instruction for learners and for student practice. A review of the other curricula used in the participating schools showed similar formats and strategies, with the teacher leading the instruction, practice tasks, and assessments.

Some classrooms used certain types of curriculum supplements that are not part of the regular curriculum and therefore are not included in table 10. Twelve participating schools (37.5 percent) used Study Island software (www.studyisland.com) as a supplement to the regular curriculum in control classrooms. During the observed class periods there was no use of the software to extend math instructional time beyond the typical math period in which the regular curriculum was used. No additional data are available on the frequency of Study Island use. Another three schools used other existing curriculum supplements, though use was not seen during classroom observations. Thus, 47 percent of participating schools reported use of other software in their control classrooms.

From the classroom observations the authors concluded that Odyssey Math was implemented with fidelity and that there were no noteworthy differences between conditions (see appendix H for a summary of information gathered during these observations). Classroom observers could see the software in use and confirm that teachers used intervention guidelines (each student had access to a computer, and students appeared to be comfortable using the software). They could also confirm that the software was not used in control classrooms. The study team also reviewed the Odyssey Math usage logs to confirm that no students or teachers from control classrooms had usernames and passwords to access the system.

4. RESULTS: DID ODYSSEY MATH IMPROVE MATH ACHIEVEMENT?

This chapter presents evidence on whether grade 4 classrooms using Odyssey Math as a partial substitute for the standard math curriculum outperformed control classrooms on the math subtest of the TerraNova Basic Battery, the confirmatory question. After comparing intervention and control classrooms (across schools) on baseline characteristics, the chapter presents findings, generated by the multilevel models, to address the confirmatory research question. The chapter also reports analyses of tests of how sensitive the empirical findings are to estimating a random effects rather than a fixed effects model, to including the pretest covariate at different levels of the multilevel model, to including baseline characteristics in the model that were statistically significantly different between intervention and control classrooms (at $p < .10$), and to using a dummy variable adjustment rather than listwise deletion for missing data on the pretest. The impact estimate with the pretest as a covariate is the empirical result that addresses the primary confirmatory question.

BASELINE CHARACTERISTICS OF ANALYTIC SAMPLE

The intervention and control classrooms were shown to be statistically equivalent at pretest (see table 8 in chapter 2). This continues to be the case when comparing the groups at pretest using the sample of students who completed both the pre- and posttests (the analytic sample). Table 11 presents the baseline characteristics for the analytic sample of 122 teachers (and 124 classrooms) with 2,456 students. There was no statistical difference at the $p < .05$ level between the intervention and control groups on any of the characteristics compared. In other words, sample loss between the pretesting and analysis phases of the study did not alter the statistical equivalence of the intervention and control groups on measured baseline characteristics.

Table 11. Mean baseline characteristics for intervention and control group classrooms at pretest for the analytic sample

Baseline characteristics	Intervention classrooms	Control classrooms	Difference	Test statistic[a]	p-value
Student characteristics					
Proportion of girls (percent)	51.00 (sd = 9.81 n = 60)	48.40 (sd = 7.95 n = 62)	2.60	$t = 1.61$ (1.61)	.11
Proportion of racial/ethnic minority students (percent)[b]	24.99 (sd = 32.81 n = 44)	24.71 (sd = 32.21 n = 43)	0.28	$t = 0.04$ (6.97)	.97
Proportion of English language learner students (percent)	6.28 (sd = 18.88 n = 60)	6.72 (sd = 21.66 n = 62)	–0.44	$t = 0.12$ (3.68)	.90
Proportion of students eligible for free or reduced-price lunch (percent)	19.06 (sd = 21.48 n = 60)	16.75 (sd = 19.48 n = 62)	2.31	$t = 0.62$ (3.71)	.54

	115.61	116.01			
	(sd = 2.13	(sd = 2.94		t = 0.85	
Student age (months)	n = 60)	n = 62)	−0.40	(0.47)	.40
Classroom average pretest score					
	621.81	622.32			
	(sd = 14.40	(sd = 14.30		t = 0.20	
TerraNova Basic Battery math subtest	n = 60)	n = 62)	−0.51	(2.60)	.84

a. Numbers in parentheses are standard errors.

b. Students in some participating schools did not complete their racial/ethnic code during the pretest. Both the control and intervention classrooms within the school did not complete the information, so the report includes statistics for only 86 classrooms.

Source: Authors' analysis based on data described in text.

PRELIMINARY ANALYSES: ESTIMATED INTRACLASS CORRECTION AND UNADJUSTED MEAN DIFFERENCES

Before the conditional multilevel models (hierarchical linear models) with at least one covariate were estimated, an unconditional model without covariates was estimated (also known as a random effects analysis of variance model) using HLM6 to assess clustering at the student and teacher levels. The estimated intraclass correlation (ICC) between any two students sharing the same teacher in the same school (or teacher-level ICC) was 0.12 (see appendix I). There was less clustering in the observed data than had been assumed during the design phase (ICC = 0.20), one of several indicators that the study was adequately powered to detect the target minimum effect size of 0.20 standard deviation.[22]

As discussed, the presence of clustering justified the use of the multilevel model to assess the impact of Odyssey Math on math achievement. The analytic sample for estimating the model included 2,456 students with both pre- and posttest scores, 122 teachers, and 32 schools. The number of students per teacher ranged from 6 to 34, with an average of 20. The number of teachers per school ranged from two to six, with an average of four.

Table 12 compares the intervention and control classrooms on their unadjusted pre- and posttest means for the TerraNova Basic Battery math subtest, taking into account the clustering data structure (a random intercepts model with fixed intervention effect and no covariates). The TerraNova scaled scores on level 14 (grade 4) were used for both pre- and posttest. The minimum observed score was 403 and the maximum was 770 on both the pretest and posttest in the study sample. The average pretest difference between intervention and control classrooms was estimated at 0.11 scale score points (SE = 2.51), and the average posttest difference was 0.81 scale score points (SE = 2.36). Both intervention and control classrooms showed essentially the same gains from pre- to posttest (see table 12). The difference between the intervention and control classrooms at both pre- and posttest was less than 1 scale score point on the TerraNova Basic Battery. Neither difference was statistically significant at the $p < .05$ level with the statistical test based on the proper standard error taking clustering into account.

---

[22] The pretest teacher level ICC was also 0.12, indicating that any two students with the same teacher in the same school did not become any more homogeneous on math achievement from the start of the school year to the end.

Table 12. Intervention and control classroom means and estimated differences on math achievement at pre- and posttest and estimated impact of Odyssey Math on math achievement

Outcome measure	Intervention classrooms	Control classrooms	Estimated difference[a]	p-value	95 percent confidence interval	Effect size[b]
Pretest score	621.46	621.35	0.11 (2.51)	.964	−4.81, 5.03	na
Posttest score unadjusted for class pretest mean	647.41	646.60	0.81 (2.36)	.734	−3.82, 5.44	0.02
Posttest score adjusted for class pretest mean	648.29	647.50	0.78 (1.27)	.543	−1.71, 3.27	0.02

na is not applicable.
a. Numbers in parentheses are standard errors.
b. Standardized difference by student-level pooled standard deviation of posttest scores.
Source: Authors' analysis based on data described in text.

Another way to interpret the average posttest difference between intervention and control classrooms is to standardize the difference as an effect size. The pooled standard deviation for student-level posttest scores was 38.69 and the control group student level standard deviation was 38.18. The effect size on posttest was 0.02 standard deviation regardless of whether pooled or control group standard deviation was used to standardize the difference. This effect size represents a very small difference in posttest achievement between the two groups (see Rosnow and Rosenthal 2003) and is likely due to random fluctuations from zero standard deviation units. The results from this unconditional model (without covariates) indicate that the intervention did not have a statistically significant effect on the posttest mean or its variability.

RESULTS OF MULTILEVEL MODEL WITH PRETEST COVARIATE

The results from the multilevel model with pretest covariate also indicate that Odyssey Math did not yield a statistically significant impact on end-of-year student achievement (see table 12, last row). The impact is quantified by the multilevel model posttest mean difference between intervention and control classrooms adjusted for class mean pretest scores (γ_{010} = 0.78, SE = 1.27). The adjusted posttest mean difference (for class mean pretest scores) was slightly smaller than the unadjusted posttest mean difference in table 12 (unadjusted posttest mean difference = 0.81, SE = 2.36). Both differences are less than one scale point on the math achievement test (see appendix J for a complete table of parameter estimates for the model).[23]

SENSITIVITY ANALYSIS: ALTERNATIVE MODELS

Several sensitivity tests were run to assess whether the results were affected by the decision to estimate a random effects (rather than fixed effects) model, potential group differences on two professional development variables (whether teachers received "short

---

[23] So the reader can evaluate the statistical power of the design to detect a less than one scale point difference between groups on math achievement, a comparison of assumed statistical power population parameters with corresponding actual sample statistics is presented in appendix K.

training" of one-half day or less of professional development and whether teachers received "long training," defined as more than one-half day of professional development), different ways of treating missing data on the pretest, and inclusion of the pretest covariate at different levels of the multilevel model.

Pretest covariate at different levels of the model

Student achievement pretest scores were aggregated to the teacher-classroom level (level 2), grand mean centered at level 2, and entered as a covariate in the model at level 2 for the core analysis to address the confirmatory question. As an alternative, the first model was replicated but with student achievement pretest scores entered at level 1 as grand mean centered to evaluate how sensitive the impact estimate was to placement of the pretest score at level 1 rather than at level 2.

Based on the results of these models it can be concluded that the impact estimate (γ_{010} = 0.73) and standard error (SE = 1.28, t_{31} = .571, p = .572) were invariant to the decision to include student achievement pretest scores at level 1 or level 2 in the multilevel model.

Random or fixed effects model

To evaluate how sensitive the impact estimate (or treatment effect) and standard error are to the decision to model school effects as random in the core analysis, a series of fixed effect models were estimated as a sensitivity analysis:

- A two-level model with students at level 1 and classrooms at level 2 as specified previously but with the impact estimate (or treatment effect), β_{01k}, modeled as fixed across schools (a two-level model estimated without the school level). The results showed that the impact estimate β_{01} = 0.58 (SE = 1.51, t_{119} =.386, p = .700).
- A two-level model with students at level 1 and classrooms at level 2, as specified previously, but with the impact estimate (or treatment effect), β_{01k}, modeled as fixed and school effects modeled as fixed by including Z – 1 dummy variables (where Z is the total number of schools in the sample) at the classroom level. The results show that the impact estimate β_{01} = 0.91 (SE = 1.48, t_{88} =.617, p = .538).

Based on the results of these models, it can be concluded that the impact estimate and the standard errors are insensitive to the choice of a random effects or fixed effects models.

Group difference on math professional development variables

A sensitivity analysis was conducted by including the two professional development variables for which there was a statistically significant mean difference between intervention and control classrooms at p < .10: p = .053 (favoring the control group) for long training (more than a half day) and p = .07 (also favoring the control group) for short training. Each variable was included in the impact multilevel model as a teacher-level covariate (grand mean centered) to address the first research question. The fixed effect parameter estimates did not change substantially, nor did the statistical tests when teacher long training and pretest class

means were controlled for (impact estimate = 1.00, SE = 1.56, p = .53) or when teacher short training and pretest class means were controlled for (impact estimate = 0.59, SE = 1.55, p = .71), indicating that the impact estimate and statistical significance were insensitive to excluding these variables from the model.

Missing data on the pretest

The impact model was reanalyzed with two additional level 1 covariates: grand mean–centered student pretest scores with grand mean substitution for missing data and missing dummy variables to adjust for the effect of missing student-level pretest data. The impact estimate (0.65), its standard error estimate (1.24), and its p-value (p = .60) were similar to the corresponding estimates obtained from the complete data analysis that used listwise deletion to address missing data.

To test whether the impact estimate was invariant to the choice of the substitute mean (classroom or grand mean) for the unobserved (or missing) pretest score as part of the dummy variable adjustment, a model was estimated with the dummy variable indicator as defined previously but substituting the class mean for the missing pretest score. For class mean substitution for missing pretest score at the student level (level 1), class mean pretest score as covariate at the classroom level (level 2), and random treatment effect across school level (level 3), the impact estimate γ_{010} = 0.59 (SE = 1.23, t_{31} = .482, p = .633).

Based on the results of these two models, it can be concluded that the impact estimate and standard errors were invariant to the choice of the substitute mean for missing pretest scores with the dummy variable indicator adjustment.

Potential group differences on professional development

The models with each of the additional level 2 professional development covariates were also reanalyzed with the missing dummy variable adjustment for missing data on the pretest. The impact estimates for long training (estimate = 0.94, SE = 1.34, p = .492) and for short training (estimate = 0.58, SE = 1.35, p = .672) were also similar to the corresponding estimates with complete data. These results demonstrate that the impact estimate was insensitive to the two different approaches for handling missing data on the pretest.

The models that generated the results in table 12 and the model that generated the sensitivity results for long training professional development are in appendix K.

5. SUMMARY OF FINDINGS AND STUDY LIMITATIONS

This section summarizes the findings on the effect of Odyssey Math on grade 4 math achievement and describes the study limitations.

EFFECT OF ODYSSEY MATH ON MATH ACHIEVEMENT

The main finding from this study is that Odyssey Math did not cause a statistically significant overall effect on grade 4 math achievement. The magnitude of the effect was less than one scale score point and did not show statistically significant variability across schools. Stated differently, grade 4 classrooms using Odyssey Math as a partial substitute for their regular curriculum performed no differently than did the control classrooms on the mathematics subtest of the TerraNova Basic Battery administered at the end of the 2007/08 school year. Sensitivity analysis showed that this conclusion did not change when teacher professional development variables were added to the analysis or when missing data on the pretest were addressed using an alternative approach to listwise deletion.

CHARACTERISTICS OF AN EFFECTIVENESS TRIAL

When designing the Odyssey Math study, REL Mid-Atlantic applied Flay's (1986) definitions of an effectiveness trial. As such, the effectiveness trial was designed to test the effects of an intervention under typical conditions. The purpose was to test CompassLearning's claim that Odyssey Math has a positive effect on student learning in the instructional environment that would naturally occur had school districts purchased and implemented Odyssey Math as they normally do. Therefore, implementation features required for an efficacy trial are not applicable to this effectiveness trial.

FIRST EFFECTIVENESS TRIAL ON ODYSSEY MATH

This study was the first randomized controlled trial to assess the impact of Odyssey Math on student achievement. The study was rigorous in that it was sufficiently powered, designed as a cluster randomized effectiveness trial, and documented fidelity of intervention implementation. As a result, the study generated statistically unbiased estimates of the effects of Odyssey Math, implemented in naturalistic conditions, on student achievement. In contrast, previous research studies on Odyssey Math lacked the control groups formed by random assignment that are needed to conclude that the software caused the achievement gains observed in those studies.

LIMITATIONS

No one study can address all questions about the effectiveness of an intervention. Regardless of rigor, all studies have limitations, especially in terms of generalizability to other settings and contexts. This study is no different. The findings apply to typical

implementations of Odyssey Math software as a partial substitute for the existing curriculum at the grade 4 level:

- Because teachers were instructed to use the software for 60 minutes a week but were allowed to vary from that recommendation, it should not be inferred that this study indicates that the same results would be produced under other conditions.

- The effect demonstrated in this study applies to the Odyssey Math portion of the software and should not be generalized to the other components of the Odyssey Software Suite.

- The results apply only to the Odyssey Math curriculum at the grade 4 level and not to Odyssey Math software developed for other grade levels.

- As noted in the report, Odyssey Math may be implemented as a partial substitute within the curriculum, a supplement to the curriculum, or as a replacement for the curriculum. Findings of this study are applicable only to the partial substitute implementation option.

- The use of a volunteer sample limits the findings of this study to the schools, teachers, and students in the Mid-Atlantic Region that voluntarily participated in the study. Results should not be generalized beyond this sample.

APPENDIX A. DETAILED PROFESSIONAL DEVELOPMENT AGENDA SESSIONS

This appendix describes the professional development package CompassLearning developed for treatment teachers at the outset of the study. This description was vetted with the developer to ensure its accuracy. To convey the sense that this appendix describes planned activities, it is presented in the future tense.

GOALS OF THE COMPASSLEARNING TRAINING PACKAGE

CompassLearning has identified three broad goals of the training package:

- *Goal 1*. Intervention classroom teachers will integrate software into their weekly teaching.
   o All teachers will attend training on the Odyssey Math management system and curriculum.
   o All teachers will attend training for Odyssey Math diagnostic/prescriptive assessments aligned to TerraNova objectives and state standards.
   o Math teachers will incorporate Odyssey Math activities into their weekly lesson plans.
- *Goal 2*. Intervention classroom students will use Odyssey Math to increase their math achievement (as measured by the grade 4 TerraNova Basic Battery math test) and demonstrate growth on state assessment tests.
   o Intervention students will attend the Odyssey Math lab for at least 60 minutes a week and use the Odyssey Math assessment and learning paths customized by their coach, along with learning activities that correlate to classroom instruction.
   o Teachers will plan for student access to the computer lab and or classroom computers.
- *Goal 3*. Intervention classroom teachers will monitor and evaluate student progress in order to design student intervention plans that reflect differentiated instruction and integration of available materials.
   o Teachers will attend at least four consultant-led coaching sessions (one to two hours long) between September 2007 and April 2008.
   o Teachers will attend a full-day session on integration that uses technology, Odyssey Math resources, instructional strategies, and differentiated instruction.

ADDITIONAL TRAINING DETAILS

The two "days" of summer training will focus on showing teachers how to operate and navigate the Odyssey Math system. Teachers will receive a full review of how the software works and will learn how to use the assessment system, assign curricula

components to students, and get a sense of how the software can be used to meet state standards. The overall goal of the introductory training will be to ensure that teachers are able to implement the Odyssey Math package at the beginning of the school year. CompassLearning's stated session objectives for the summer training session are as follows:

- Understand the relationship of CompassLearning resources and materials to state standards.
- Operate the management system.
- Assign appropriate standards-based math curriculum components to students.
- Orient participants to student launch pad.
- Review the basic operation of the management system.
- Use Test Builder and preview TerraNova assessments.
- Access/generate/analyze reports.
- Create purposeful assignments.

COACHING SESSION 1

In October teachers will receive job-embedded coaching that focuses on system management training to reinforce concepts learned during the summer. The timing of the training allows for revisiting Odyssey Math features after class has been in session for a few weeks. This will give teachers a chance to use the system with students while working with a coach. In addition to reviewing properties of the software package, teachers will have a chance to troubleshoot problems they have been experiencing, begin to learn about differentiated instruction (more on this below), and use high-stakes assessment data to determine skill gaps.

Stated session objectives for the first coaching sessions are as follows:

- Teachers will create the class list and assign the TerraNova-aligned pretest as well as an initial curriculum assignment.
- Teachers will review and discuss the orientation process for students accessing the software.
- Teachers will plan for student access to complete the TerraNova-aligned Odyssey Math assessment.
  Specific training tasks include:

- Access the Set-Up Module and populate the class list with intervention students.
- Access the Assignment Archive and assign a math assignment to support instruction.
- Access the Assignment Archive and assign the TerraNova-aligned assessment.
- Distribute student orientation brochure and discuss test administration strategies.

- Encourage teachers to orient students with math curriculum assignment first.
- Review CompassLearning Odyssey Skills Checklist with teachers and provide coaching in areas that indicate nonmastery.

After the session the coach will edit each student's profile in class list to access Math 4 only.

COACHING SESSION 2

A second coaching session will occur in November–December, focusing on the individual learning needs of teachers and development of student progress data. CompassLearning's objectives for the second coaching session are as follows:

- Teachers will generate and review student progress reports.
- Teachers will generate and review student assessment reports.
- Teachers will use Odyssey data to assist with classroom instructional interventions.
  Specific training tasks include:

- Guide teachers as they access the following reports: Student Progress, Progress Summary, Class Progress, Test Results, Test Objective Summary, and Learning Path Status.
- Revisit the "Which report do I use?" handout, and discuss most relevant reports for classroom planning.
- Access the Assignment Status tool and modify student assignments if needed.
- Revisit the CompassLearning Odyssey Skills Checklist with teachers and provide coaching in areas that indicate nonmastery.

Specific training tasks entail the following:

- Introduce teachers to the principles of differentiated instruction.
  - o Build an assignment that helps teachers address a specific instructional objective for their students.
  - o Ask teachers to consider the underlying process of each Odyssey Math activity; identify the best match between students and given activities.
  - o Identify resources to help teachers target assignments for students in a way that supports content learning.
- Develop ways to evaluate student learning in the context of differentiated instruction.
  - o Adjust evaluation to help students understand whether they have achieved mastery of a concept.

COACHING SESSION 3

Session 3 will occur sometime in January or February. The focus will be on fully infusing Odyssey Math tools (including offline resources) into daily lesson planning and instructional delivery. CompassLearning's stated session objectives for the third coaching session are as follows:

- Teachers will incorporate Odyssey Math into their weekly lesson plans.
- Coach will provide an overview of the Offline Resources CD and discuss strategies for use of the materials.
- Teachers will experience an Odyssey Math Handbook activity using a student study guide.

Specific training tasks include:

- Distribute and view the contents of the Offline Resources CD.
- Discuss strategies to integrate CD materials.
- Coach teachers on incorporating online and offline activities into their math instructional day.
- Distribute Student Handbook Study Guides, and plan for instructional use with students.
- Access and review available Odyssey Reports.

COACHING SESSION 4

This final coaching session should occur in March (April at the latest). Training objectives assume that teachers have strong working knowledge of the Odyssey Math software and use it regularly. With this base, they should be ready to tailor lesson plans to individual student learning needs. CompassLearning's stated session objectives for the fourth coaching session are as follows:

- Teachers will create scaffolded assignments to address varying student abilities within the same skill set.
- Teachers will make assignments for specific students.
- Teachers will plan for student interventions using Learning Path Status student data.
  Specific training tasks include:

- Revisit the Assignment Module and use Assignment Builder to create scaffolded (tiered) assignments.

- Demonstrate the use of folders and subfolders within assignments as well as folder settings for activity functionality.

- Revisit Decision Points and Passing Scores that can be attached to activities within assignments.

- Access and interpret student reports.

APPENDIX B. STATISTICAL POWER ANALYSIS

This appendix describes the statistical power analysis laid out in the proposal for the design of this randomized controlled trial (Wijekumar and Hitchcock 2006). The analysis was conducted using the multisite cluster randomized trial option in the Optimal Design software package (Spybrook et al. 2006).

The lack of internal validity of previous empirical studies of Odyssey Math made it difficult to form an empirical basis for a hypothesized effect size to be used in power calculations. As Bloom (2005) notes, Cohen (1977) suggested that a small effect size is approximately .20 standard deviations, a medium is .50, and a large is .80. Lipsey and Wilson (2001) have generated empirical support for this suggestion. More recently, Agodino et al. (2003) presented empirical evidence for setting the minimally detectable effect size for technology-based interventions in which the outcome measure is standardized achievement in the range of d = .25–.35. Previous studies of Odyssey Math suggest medium effect sizes, but these results are based on designs with questionable causal validity. Furthermore, because Odyssey Math is used in this study as a partial substitute for the standard curriculum, a conservative approach was taken, setting the minimally detectable effect size at 0.20. Based on this choice, the study was sufficiently powered to detect smaller yet educationally meaningful effects of the curriculum, if they existed. The following additional assumptions were made:

- Statistical power of .8.
- Statistical significance level at α = .05 for a two-tailed test.
- 25 students per classroom, but with an 80 percent posttest response rate so that both pre- and posttest data are available for 20 students per classroom.[24]
- Balanced allocation with four teachers (or classrooms) per school.
- A minimum detectable effect size of 0.20, but with power analyses also presented for 0.25, for comparison.
- Explanatory power (R^2) classroom-level covariates (math pretest of the math outcome measure) of .56 and .62.
- Intraclass correlation (ICC) ρ–values of .10, .15 and .20. Limited information is available in the research literature to guide assumptions about ICC values for education outcomes. Schochet (2005) presents ICC values that suggest that .10 marks the low range, .15 the mid-range, and .20 the upper range.
- Power analyses were performed for fixed effects analyses as well as random effects. Random effects models consider additional sources of variance and thus tend to

---

[24] Cluster-level attrition was assumed to be minimal for a one-year intervention. Research suggests that most teacher attrition occurs during the summer, so it could be assumed that schools and classrooms would generally stay with a study. For a more conservative estimate, we multiplied the required sample size by 1.1 to provide a margin for error.

require larger sample sizes, although the differences were not dramatic in this design and results for random effects models are presented in table B1.

Table B1. A priori power analysis for multisite randomized controlled trial with schools as random effects

Proportion of the explained variance in the level 2 covariate	$\rho = .10$		$\rho = .15$		$\rho = .20$	
	Classrooms	Schools	Classrooms	Schools	Classrooms	Schools
Minimum detectable effect size = 0.20						
$R^2 = .56$	84	20	100	25	112	28
$R^2 = .62$	84	18	92	23	104	26
Minimum detectable effect size = 0.25						
$R^2 = .56$	56	14	68	17	76	19
$R^2 = .62$	52	13	60	15	68	17

Note: This model assumes a .01 variance of effect size across schools, and each school produces its own effect size, which can vary. The degree to which effect sizes vary affects power. The .01 value is a default for the Optimal Design software and is recommended when trying to detect a 0.20 effect size. No blocking effect is assumed ($B = 0$).
Source: Authors' analysis based on data described in text.

The power analyses suggest that under the most conservative assumptions ($R^2 = .56$, ICC = .20, MDE = 0.20, with random effects), the study would need to recruit 28 schools (112 classrooms) to achieve power. To allow an additional margin of error, the study attempted to recruit 33 schools with at least four classrooms each. This allowed for scenarios where classroom-level attrition occurs or where schools had fewer than four grade 4 classrooms that could be assigned to conditions.

APPENDIX C. PROBABILITY OF ASSIGNMENT TO STUDY CONDITIONS

The probability of assignment was 50 percent for each teacher in the sample using the school as a blocking factor. The random assignment was conducted for schools with 2, 3, 4, 5, and 6 teachers. Because the main text describes the random assignment process for schools with three teachers, the examples that follow describe the process for a school with two teachers, four and six teachers (to show how the process applied to larger groups), and three and five teachers (to demonstrate how the process worked with an odd number of teachers). Second, the explanation is modified to demonstrate why the probability of selection was 50 percent.

Random assignment of conditions to teachers was conducted independently in each school. In general, within each school all teachers enrolled in the study were listed in the spreadsheet, assigned a random number, and sorted in ascending order by these numbers. Each teacher was assigned to either the intervention or the control condition, and each assigned condition was assigned a random number. The conditions (listed beside each teacher) were sorted by that number. Table C1 provides an example for a school with two teachers.

Table C1. Random assignment for a school with two teachers

District	School	Number of teachers	Number of students	Teacher identification	Teacher random number (sorted ascending)	Condition	Condition random number (sorted ascending)
1	A	2	18	B	0.005059943	Control	0.317672024
			19	C	0.442152720	Intervention	0.451865140

Source: Authors' analysis.

In this two-teacher scenario the probability of random assignment to either the intervention or the control condition is clearly 50 percent. This probability applies to all schools with an even number of teachers. When there are four teachers, each teacher has a two in four chance of being assigned to either the intervention or the control group, and when there are six teachers, the chance is three in six (table C2).

Table C2. Random assignment for schools with four or six teachers

District	School	Number of teachers	Number of students	Teacher identification	Teacher random number (sorted ascending)	Condition	Condition random number (sorted ascending)
2	B	4	29	A	0.022143812	Intervention	0.151401646
			28	B	0.375630698	Control	0.346167298
			28	C	0.758037054	Intervention	0.357526685
			27	D	0.777492445	Control	0.881163748
	C	6	24	A	0.0277311635	Intervention	0.282777251
			23	B	0.3552814269	Control	0.306743025
			24	C	0.7099579051	Control	0.423735487
			24	D	0.7869448344	Intervention	0.659483027
			24	E	0.8620487790	Control	0.660952959
			24	F	0.9570748475	Intervention	0.778937978

Source: Authors' analysis.

For schools with an odd number of teachers the probability of assignment is also 50 percent because there are *n* + 1 occurrences (where *n* is the number of teachers) of intervention or control conditions (table C3).

Table C3. Random assignment for schools with three or five teachers

District	School	Number of teachers	Number of students	Teacher identification	Teacher random number (sorted ascending)	Condition	Condition random number (sorted ascending)
1	D	3	21	A	0.193462905	Control	0.514158344
			21	B	0.399362138	Intervention	0.567417901
			19	C	0.879538643	Control	0.646899288
						Intervention	0.809666408
	E	5	24	A	0.3525713234	Control	0.3331299163
			24	B	0.4479692658	Intervention	0.3919477578
			24	C	0.5251795640	Control	0.4951489155
			24	D	0.8091025645	Control	0.6330112624
			24	E	0.8693979724	Intervention	0.7128600351
						Intervention	0.8083222680

Source: Authors' analysis.

Because of the *n* + 1 occurrences of alternative study conditions, in schools with three teachers there was a two in four chance of each teacher being randomly assigned to either the intervention or the control condition. In schools with five teachers there was a three in six chance.

APPENDIX D. SAMPLE SIZE FROM RANDOM ASSIGNMENT TO DATA ANALYSIS

Table D1 shows the sample size from random assignment through posttest.

Table D1. Sample sizes at different levels from random assignment to posttest phases

Level	Schools	Classrooms Intervention	Classrooms Control	Teachers Intervention	Teachers Control	Enrollment Intervention	Enrollment Control	Total
Random assignment	33	62	65	61	64	na	na	na
At professional development	33	62	65	61	64	na	na	na
Estimated enrollment	na	na	na	na	na	1,399	1,477	2,876
Enrollment from rosters	na	na	na	na	na	1,448	1,492	2,940
Not eligible to participate (special education student, English language learner student, Title I math, not enrolled)	na	na	na	na	na	−45	−41	−86
Eligible to participate	na	na	na	na	na	1,403	1,451	2,854
Parents did not consent	na	na	na	na	na	−15	−16	−31
Other						−27	−84	−111
Absent at pretest	na	na	na	na	na	39	33	72
Pretested	32	61	63	60	62	1,322	1,318	2,640
Posttested	32	61	63	60	62	1,300	1,284	2,584
Total analytic sample[a]	32	61	63	60	62	1,223	1,233	2,456

na is not applicable.

a. The students and classrooms in the analytic sample were those that had completed both the pre- and posttests. Students who moved out of the district during the academic year would have a pretest but no posttest and as a result were excluded from the analytic sample. Students who moved into the district and students crossing over from their randomly assigned condition were included in the analytic sample.

Note: Two of the participating teachers were each assigned to two classrooms in one participating school district. Both classrooms for the same teacher were assigned to the same research condition. Therefore, this table shows more classrooms than teachers (124 classrooms and 122 teachers). Student assent was 100 percent. There were 32 schools at pretest because one school in the random assignment pool was deemed ineligible to participate after random assignment.

APPENDIX E. TEACHER SURVEY, FALL 2007

Dear Teacher:

The Odyssey Math® study is a groundbreaking national study designed to test an innovative method for teaching math in grade 4. Your participation is important and appreciated, but you do have the right to skip any question that you do not wish to answer. Below are answers to some general questions concerning this survey.

What is the purpose of this survey?

The purpose of this survey is to collect background information, such as years of teaching experience, about the teachers participating in the study.

Who is conducting this survey?

The Odyssey Math study was commissioned by the Department of Education's Institute of Education Sciences and is administered by its Mid-Atlantic Regional Educational Laboratory, a consortium of the Pennsylvania State University, Rutgers University, ICF-Caliber, The Metiri Group, and Analytica.

Why should you participate in this survey?

Policymakers and education leaders rely on findings from studies like the Odyssey Math study to make decisions about curricula or, in this case, supplements to curricula. The current study will help determine if Odyssey Math software can help students with mathematics achievement. Your participation in the study is critical when it comes to answering this question.

Will your responses be kept confidential?

All responses that relate to or describe identifiable characteristics of individuals may be used only for statistical purposes and may not be disclosed, or used, in identifiable form for any other purposes, unless otherwise compelled by law. Your responses are protected from disclosure by federal statute (PL 107-279, Title I, Part E, Sec.183).

How will your information be reported?

The information you provide will be combined with the information provided by other teachers in statistical reports. No information that links your name, address, or telephone number with your responses will be included in any reports related to the study.

Where should you return your completed survey?

Please return the completed survey to the person who gave you the survey.

Who can you contact about the survey?

If you have any questions about the survey, you can ask the person who gave you the survey, or you can contact the coordinator of data collection, <insert name>.

Thank you for your cooperation in this very important effort!

BACKGROUND INFORMATION

Education

1. Have you earned any of the following degrees, certificates, or credentials? (Check no or yes in each row, and write in the major code from table 1 and the year if applicable.)

	Degree	Earned	Major code (from table 1)	Year
a.	Bachelor's degree	1☐ No 2☐ Yes➔	. .	. .
b.	Master's degree	1☐ No 2☐ Yes➔	. .	. .
c.	Educational specialist or professional diploma (at least one year beyond master's level)	1☐ No 2☐ Yes➔	. .	. .
d.	Certificate of advanced graduate studies	1☐ No 2☐ Yes➔	. .	. .
e.	Doctorate or professional degree (Ph.D., Ed.D., M.D., L.L.B., J.D., D.D.S.)	1☐ No 2☐ Yes➔	. .	. .

Table 1. Major field of study codes

Major code	Major field
01	Elementary education
02	Secondary education
03	Special education
04	Arts/music
05	English/language arts
06	English as a second language
07	Foreign languages
08	Mathematics

09	Computer science
10	Natural sciences
11	Social sciences
12	Other

Experience

2. How do you classify your position at THIS school, that is, the activity at which you spend most of your time during this school year? *Mark (X) only one box.*

☐   Regular full-time teacher

☐   Regular part-time teacher

☐   Itinerant teacher (i.e., your assignment requires you to provide instruction at more than one school)

☐   Long-term substitute (i.e., your assignment requires that you fill the role of a regular teacher on a long-term basis, but you are still considered a substitute)

3. How many years of teaching experience do you have (write in number of years, and count the current year as one full year):

	Number of years
a. Teaching in total	. . Years .
b. Teaching grade 4	. Years .
c. Teaching at this school	. . Years

PROFESSIONAL DEVELOPMENT EXPERIENCES

Types of professional development

In answering the following items, consider all the professional development activities related to math instruction or use of computers to teach (second section) in which you have participated during the summer of 2007 or the 2006/07 school year.

Professional development refers to a variety of activities intended to enhance your professional knowledge and skills, including teacher networks, coursework, institutes, workshops, committee work, coaching, and mentoring. Workshops are short-term learning opportunities that can be located in your school or elsewhere. Institutes are longer term professional learning opportunities, for example, of a week or longer in duration.

4. Since completing your degree, what is the total number of hours you have spent in the following professional development activities for math instruction?

Write the total number of *hours* you spent in these activities. Mark "0" if you participated in none.

	Number of hours
a. Attended short, stand-alone training or workshop in math (half-day or less)	
b. Attended longer institute or workshop in math (more than half-day)	
c. Attended a college course in math (include any courses you are currently attending)	
d. Received coaching or mentoring related to math instruction	
e. Acted as a coach or mentor related to math instruction	
f. Other informal professional development (e.g., participated in teacher study group, network, or collaboration supporting professional development in math, participated in committee or task force related to math, visited or observed math instruction in other schools)	

5. What is the total number of hours you spent in the following professional development involving *the use of computer technology* (i.e., any software, hardware, Internet, or peripheral components) in a teaching context?
Write the total number of *hours* you spent in these activities. Mark "0" if you participated in none.

	Number of hours
a. Attended short, stand-alone training or workshop in using computers (half-day or less)	
b. Attended longer institute or workshop in using computers (more than half-day)	
c. Attended a college course focusing on computer technology (include any courses you are currently attending)	
d. Received coaching or mentoring related to computers	
e. Acted as a coach or mentor related to using computers in a teaching context	
f. Other informal professional development (e.g., participated in teacher study group, network, or collaboration supporting professional development in computer use, participated in committee or task force related to computer-technology, visited or observed the use of computers in other schools)	

You are done with the survey. Thank you.

APPENDIX F. OBSERVATION PROTOCOLS

This appendix contains fidelity checklists for control classroom and intervention classroom observations.

FIDELITY CHECKLIST FOR CONTROL CLASSROOM OBSERVATIONS

Basic data

School name	Teacher name	Date of visit	Timeframe of observation

Classroom environment and technical observations—control group

Question	Answer	Further comments
Number of students		
Number of absent students		
Including teacher aides, how many teachers are in the classroom?		
Have students with disabilities been accommodated?		
Are all students working on math learning or is this time being used to supplement class time? (Making up missed exams or regular class work would be an example)[1]	Y/N (Circle one and add notes as needed)	
Is the classroom environment quiet?	Y/N	
Do all students have access to their own computer workstation and/or are they working at their desk?	Y/N	
Do all students have their books?	Y/N	
Do students stay in the classroom for the whole period? (An example would be leaving for another class or extracurricular activity; an exception would be leaving to use the restroom)	Y/N	
Do students work on their own, or do they tend to ask for or take help from their neighboring classmates?[2]	Y/N	
Further comment about classroom environment		

1. If all students are working on Odyssey Math, the reviewer will mark "Yes." Otherwise, the reviewer will note how many students are doing other work and document what type of work they are doing.
2. If students ask other classmates for help, the reviewer would mark "Yes."

Teacher-student interactions—control group

Criteria	Scale of 1–5, with 1 being least favorable, 5 being exceptional	Comments
Teacher listened to student questions carefully	1 2 3 4 5	
Teacher intervened with students appropriately	1 2 3 4 5	
Students were treated with respect	1 2 3 4 5	
Teacher answered student questions correctly and reasonably	1 2 3 4 5	
Teacher used computer applications (List what was used)	1 2 3 4 5	
Teacher was comfortable answering any computer-related student questions	1 2 3 4 5	
Teacher had control of the classroom	1 2 3 4 5	
Students asked questions when necessary	1 2 3 4 5	
Students used examples and tools as needed to learn the content	1 2 3 4 5	
Additional comments or concerns		

Math content—control group

Criteria	Scale of 1–5 (1 is least favorable and 5 is exceptional)	Comments/notes
Learning objectives for the class period		
Teacher clearly articulated the objectives for the class period	1 2 3 4 5	
Motivational component to the learning objectives included	1 2 3 4 5	
Teacher used such techniques as asking questions to assess the different students' skills in the content	1 2 3 4 5	
Students used learning strategies appropriate for the learning objective	1 2 3 4 5	
Teacher presented different types of learning strategies for students with different interest and/or skills in the classrooms	1 2 3 4 5	
Teacher was able to break larger learning objectives into smaller units	1 2 3 4 5	
Teacher explained the real-life applications of the learned content	1 2 3 4 5	
Teacher used examples to explain how the content is applied	1 2 3 4 5	
Other domain related observations	1 2 3 4 5	
	1 2 3 4 5	
	1 2 3 4 5	
	1 2 3 4 5	
Additional comments or concerns		

FIDELITY CHECKLIST FOR ODYSSEY MATH INTERVENTION CLASSROOM OBSERVATION

Basic data

School name	Teacher name	Date of visit	Timeframe of observation

Classroom environment and technical observations—Odyssey intervention group

Question	Answer	Further comments
Number of students		
Number of absent students		
Including teacher's aides, how many teachers are in the classroom?		
Have students with disabilities been accommodated?	Y / N (add notes here if necessary)	
Are all students working on Odyssey Math, or is this time being used to supplement class time? (Making up missed exams or regular class work would be an example)[a]	Y/N	
Is the classroom environment quiet?	Y/N	
Do all students have access to their own computer workstation?	Y/N	
Are all computers in proper working order (are they usable throughout the class period, batteries stay charged on mobile workstations, etc.)	Y/N	
Do all students have working headphones?	Y/N	
Do students stay in the classroom for the whole period? (An example would be leaving for another class or extracurricular activity; an exception would be leaving to use the restroom)	Y/N	
Do students work on their own, or do they tend to ask for or take help from their neighboring classmates?	Y/N	
Further comment about classroom environment		

a. If all students are working on Odyssey Math, the reviewer will mark "Yes." Otherwise, the reviewer will note how many students are doing other work and document what type of work they are doing.

Teacher-student interactions—Odyssey intervention group

Criteria	Scale of 1–5, with 1 being least favorable, 5 being exceptional	Comments
Teacher listened to student questions carefully	1 2 3 4 5	
Teacher intervened with students appropriately	1 2 3 4 5	
Students were treated with respect	1 2 3 4 5	
Teacher answered student questions regarding Odyssey Math correctly and reasonably	1 2 3 4 5	
Teacher was comfortable using the computer	1 2 3 4 5	
Teacher was comfortable answering any computer-related student questions	1 2 3 4 5	
Teacher had control of the classroom	1 2 3 4 5	
Teacher followed all Odyssey Math guidelines as presented during training	1 2 3 4 5	
Students were comfortable using the Odyssey Math program	1 2 3 4 5	
Students asked questions when necessary	1 2 3 4 5	
Students were excited to be doing Odyssey Math	1 2 3 4 5	
Students only worked on Odyssey Math while using the computer workstations	1 2 3 4 5	
Students were encouraged to use all of the tools incorporated into Odyssey Math to enhance the learning experience	1 2 3 4 5	

Math content—Odyssey intervention group

Criteria	Scale of 1–5 (1 is least favorable and 5 is exceptional)	Comments/notes
Learning objectives for the class period		
Teacher clearly articulated the objectives for the class period	1 2 3 4 5	
Motivational component to the learning objectives included	1 2 3 4 5	
Teacher used such techniques as asking questions to assess the different students' skills in the content	1 2 3 4 5	
Students used learning strategies appropriate for the learning objective	1 2 3 4 5	
Teacher presented different types of learning strategies for students with different interests and/or skills in the classrooms	1 2 3 4 5	
Teacher was able to break larger learning objectives into smaller units	1 2 3 4 5	
Teacher explained the real-life applications of the learned content	1 2 3 4 5	
Teacher used examples to explain how the content is applied	1 2 3 4 5	
Other domain-related observations	1 2 3 4 5	
	1 2 3 4 5	
	1 2 3 4 5	
	1 2 3 4 5	
Additional comments or concerns		

APPENDIX G. ODYSSEY MATH SAMPLE SCREENS
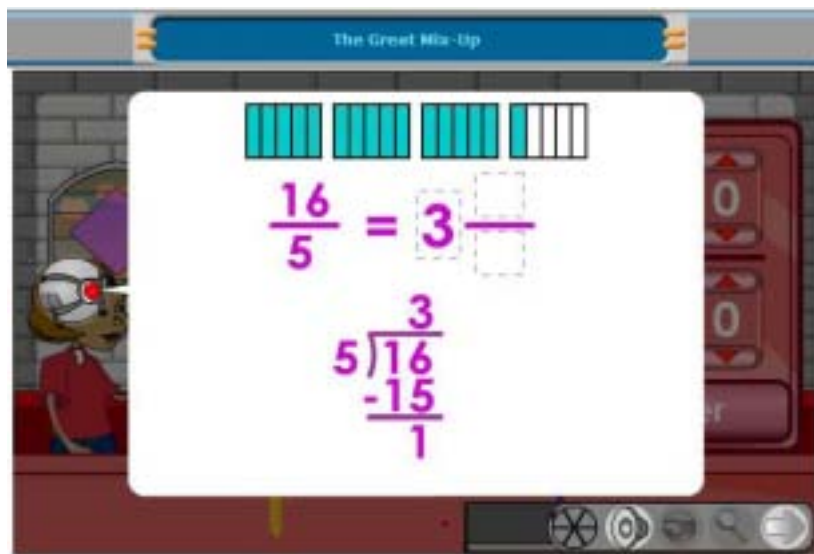
This appendix contains screenshots of sample Odyssey Math screens.

Exhibit G1. Odyssey Math launch pad

Source: CompassLearning Odyssey Math®.

Exhibit G2. Sample Odyssey Math learning activity

Source: CompassLearning Odyssey Math®.

Exhibit G3. Sample assessment from Odyssey Math

Chapter Test (HA4A02) - Microsoft Internet Explorer

Question 1 of 15

Next ▶   Quit ✗

Find the sum.

1,189
+ 3,442

○ A) 4,521
○ B) 4,531
○ C) 4,631
○ D) 4,731

Scored Quiz

D) 736

11. The cows in the dairy has produced 78,932 gallons of milk. The dairy delivered 22,040 gallons of milk on Friday and 15,893 gallons of milk on Saturday. Estimate the amount of milk left at the dairy.
✓ A) 40,000 gallons of milk
   B) 50,000 gallons of milk
   C) 60,000 gallons of milk

12. Estimate the sum by rounding to the nearest hundred thousand.

674,032 + 397,192 =
   A) 1,000,000
✓ B) 1,100,000
   C) 1,200,000
   D) 2,000,000

13. Estimate the difference by rounding to the nearest ten thousand.

86,302 - 37,204 =
   A) 40,000
   B) 50,000
   C) 60,000
✗ D) 130,000

14. Estimate the difference by rounding to the nearest hundred thousand.

389,302 - 205,859 =
   A) 100,000
   B) 200,000
   C) 300,000
✗ D) 600,000

15. Choose the expression that best fits this description.
   The estimated difference is 2,000,000.
   A) 8,666,953 - 4,562,642
✓ B) 7,387,486 - 4,599,425
   C) 8,061,642 - 7,903,742

You have finished the quiz. Please click Continue to move to the next screen.

Continue ▶

Source: Retrieved August 21, 2008, from www.compasslearningodyssey.com.

APPENDIX H. FIDELITY OBSERVATION COMPARISONS

Table H1. Comparisons of class observations between control teachers' classrooms and intervention teachers' classrooms

Observation item	Aggregate response for Odyssey® Math classrooms	Aggregate response for control classrooms
Average number of students during the observation	20.77 (3.314)	20.24 (3.607)
Average number of students absent during the observation	1.27 (1.127)	2.11 (5.463)
Including teacher aides, average number of teachers in the classroom	1.39 (.788)	1.21 (.585)
Percentage of classrooms with apparent accommodations for students with a disability	60.3 (49.3)	67.8 (47.1)
Percentage of classrooms that had a "quiet" environment	84.7 (36.3)	96.6 (18.4)
Percentage of classrooms where students stayed in the room for the entire instructional period	90.7 (28.6)	91.5 (28.1)
Percentage of classrooms that used group-based work (students working together) as opposed to individualized work	84.7 (36.3)	83.1 (37.8)
Percentage of classrooms using an individual work/textbook	na	84.5 (36.5)
Percentage of classrooms specifically working on math activities	93.2 (23.6)	100 (0.00)
Percentage of classrooms where students had individualized access to a computer	96.6 (18.3)	66.1 (47.7)
Percentage of classrooms that appeared to have computers in working order	81.4 (39.3)	N/A
Percentage of classrooms with available headphones	76.3 (42.9)	N/A
Did teachers listen carefully to students?	4.23 (.745)	4.32 (.730)
Did teachers intervene with student appropriately?	4.25 (.703)	4.36 (.693)
Were students treated respectfully?	4.36 (.712)	4.48 (.655)
Were teachers comfortable using a computer?	4.18 (.948)	N/A
Were teachers in control of the classroom?	4.48 (.732)	4.49 (.679)
Did students ask questions when necessary?	4.12 (.888)	4.12 (.839)
Were teachers comfortable answering computer related student questions?	4.05 (.840)	N/A
Did students use examples and tools as needed to learn content?	3.11 (1.413)	4.19 (.789)
Did teachers use computer applications?	Not in Odyssey Math	Only 12 responses
Did Odyssey Math teachers use guidelines presented during training?	3.98 (.995)	N/A
Were Odyssey Math students comfortable using the program?	4.13 (.685)	N/A
Did Odyssey Math students appear to be excited when using the	3.95	N/A

program?	(.705)	
Did Odyssey Math students use Odyssey Math only when working with a computer?	4.41 (.814)	N/A
Did teacher clearly articulate learning objectives for the period?	3.40 (1.272)	4.03 (.837)
Did teachers ask students questions to assess their skill level?	3.66 (1.121)	4.29 (.756)
Did students use strategies appropriate for the objective?	3.85 (.911)	4.19 (.687)
Did teachers use different types of learning strategies for students with different interests and skills?	3.50 (1.109)	3.98 (1.068)
Was teacher able to break larger learning objectives into smaller units?	3.64 (1.056)	4.17 (.841)
Did teacher explain real life applications of learning content?	2.81 (1.312)	3.45 (1.245)
Did teachers use examples of how content was applied?	2.93 (1.330)	3.72 (1.136)

Source: Authors' analysis based on data described in text.

APPENDIX I. MODEL VARIANCE AND INTRACLASS CORRELATIONS

The variance components from the unconditional (or null) three-level multilevel model estimates can be partitioned as follows:

$$\sigma^2 = 1{,}312.56$$

$$\tau_\pi = 102.63$$

$$\pi_\beta = 76.42$$

$$V(y_{ijk}) = \pi_\beta + \tau_\pi + \sigma^2 = 1{,}491.61.$$

Table I1 presents the variance component ratios and intraclass correlations (ICCs). For example, the proportion of variance within teachers' classrooms is σ^2 divided by total variance $(\pi_\beta + \tau_\pi + \sigma^2)$, or $1{,}312.56/1{,}491.61 = .88$ (88 percent). The proportion of variance among teachers' classrooms within schools is τ_π divided by the total variance $(\pi_\beta + \tau_\pi + \sigma^2)$, or $102.63/1{,}491.61 = .07$ (7 percent). Finally, the proportion of variance among schools is π_β divided by the total variance, which is $.05$ (5 percent). Each ratio quantifies how much student-, classroom-, and school-level characteristics contribute to the total variance in the model.

Table I1. Estimated proportion of variance by level and intraclass correlations based on a three-level unconditional model

Partitioned variance/intraclass correlation	Estimate	Description
Proportion of variance within teachers' classrooms	0.88	About 88 percent of the variance in achievement is due to student characteristics
Proportion of variance among teachers within schools	0.07	About 6.9 percent of the variance is due to differences among teachers within schools
Proportion of variance among schools	0.05	About 5.1 percent of the variance is due to differences among schools
$ICC1 = \dfrac{\tau_\beta}{\pi_\beta + \tau_\pi + \sigma^2}$	0.05	Correlation between any two students who go to the same school but have different teachers
$ICC2 = \dfrac{\tau_\beta + \tau_\pi}{\pi_\beta + \tau_\pi + \sigma^2}$	0.12	Correlation between any two students who share the same teacher at the same school
$ICC_{\beta_k} = \dfrac{\tau_\beta}{\pi_\beta + \tau_\pi}$	0.43	Correlation of average student achievement among teachers within schools

Source: Authors' analysis based on data described in text.

APPENDIX J. COMPLETE MULTILEVEL MODEL RESULTS FOR RESEARCH QUESTION 1

Tables J1 and J2 present the fixed effects and random effects multilevel model results for research question 1: Do grade 4 classrooms using Odyssey Math as a partial substitute for the standard math curriculum outperform control classrooms on the math subtest of the TerraNova Basic Battery in a typical school setting?

Table J1. Multilevel fixed effects model estimates for the impact assessment of Odyssey Math on student math achievement

Fixed effects model	Coefficient	Standard error	*t*-ratio	Degrees of freedom	*p*-value
γ_{000}, adjusted grand school mean in control condition	647.15	1.22	531.45	31	0.000
γ_{010}, adjusted average Odyssey Math effect across all schools	0.80	1.47	0.55	31	0.588
γ_{020}, average effect of class mean pretest on student outcome across all schools	0.94	0.06	16.33	119	0.000

Source: Authors' analysis based on data described in text.

Table J2. Multilevel random effects model estimates for the impact assessment of Odyssey Math on student math achievement

Random effects	Standard deviation	Variance component	Degrees of freedom	Chi-square	*p*-value
e_{ijk}, random error associated with student i in teacher j's class in school k	36.01	1,296.45			
r_{0jk}, random error associated with teacher j in school k on class average student outcome	0.60	0.36	57	49.10	>.500
u_{00k}, random error associated with school k on adjusted school average student outcome	3.49	12.20	31	33.08	0.365
u_{01k}, random error associated with school k on intervention effect	.66	.44	31	13.86	>.50

Source: Authors' analysis based on data described in text.

APPENDIX K. COMPARISON OF ASSUMED POPULATION PARAMETERS FOR STATISTICAL POWER (DURING PLANNING PHASE) WITH CORRESPONDING SAMPLE STATISTICS (DURING ANALYSIS PHASE)

Table K1. Comparison of assumed parameter values and observed sample statistics for statistical power analysis

Statistical power parameter	Assumed parameter value (design phase)	Observed sample statistic (analysis phase)
Effect size variability, σ_δ^2	.01	.01
School-level intraclass correlation	.15	.12
Classroom-level R_{L2}^2	.56	.74
Proportion of variance explained by blocking variable B	0	.50
Average number of classrooms per school	4	3.81
Average number of students per class	20	20

Note: The reader should interpret the sample statistics with caution as the standard errors are not reported.

APPENDIX L. EQUATIONS FOR MULTILEVEL MODEL ANALYSES

The model that generated results in table 12:

Level 1 (student level):

$$Y_{ijk} = \pi_{0jk} + e_{ijk}.$$

Level 2 (teacher level):

$$\pi_{0jk} = \beta_{00k} + \beta_{01k} \, (Odyssey)_{jk} + r_{0jk}.$$

Level 3 (school level):

$$\beta_{00k} = \gamma_{000} + u_{00k}$$

$$\beta_{01k} = \gamma_{010} + u_{01k}.$$

Model that generated results in table 12, bottom row, and tables J1 and J2:

Level 1 (student level):

$$Y_{ijk} = \pi_{0jk} + e_{ijk}.$$

Level 2 (teacher level):

$$\pi_{0jk} = \beta_{00k} + \beta_{01k} \, (Odyssey)_{jk} + \beta_{02k} \, (Pretest)_{jk} + r_{0jk}.$$

Level 3 (school level):

$$\beta_{00k} = \gamma_{000} + u_{00k}$$

$$\beta_{01k} = \gamma_{010} + u_{01k}$$

$$\beta_{02k} = \gamma_{020}.$$

Model that generated sensitivity results for long training math professional development reported in chapter 4:

Level 1 (student level):

$$Y_{ijk} = \pi_{0jk} + e_{ijk}.$$

Level 2 (teacher level):

$$\pi_{0jk} = \beta_{00k} + \beta_{01k} (Odyssey)_{jk} + \beta_{02k} (Pretest)_{jk} + \beta_{03k} (Long\ training)_{jk} + r_{0jk}.$$

Level 3 (school level):

$$\beta_{00k} = \gamma_{000} + u_{00k}$$

$$\beta_{01k} = \gamma_{010} + u_{01k}$$

$$\beta_{02k} = \gamma_{020}$$

$$\beta_{03k} = \gamma_{030}.$$

REFERENCES

Agodino, R., Dynarski, M., Honey, M., and Levin, D. (2003, May). *The effectiveness of educational technology: issues and recommendations for the national study.* Princeton, NJ: Mathematica Policy Research, Inc.

Allison, P.D. (2001). *Missing data* (Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-136). Thousand Oaks, CA: Sage.

Bailey, S., and Majors, D. (2007). *Odyssey® School Effectiveness Report: Maple Leaf Intermediate Unit.* Retrieved August 30, 2008, from www.compasslearning.com/files/GarfieldHeights_OH.pdf.

Baldi, S., Jin, Y., Skemer, M., Green, P.J., and Herget, D. (2007). *Highlights from PISA 2006: performance of U.S. 15-year-old students in science and mathematics literacy in an international context* (NCES 2008-016). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Bloom, H.S. (Ed.) (2005). *Learning more from social experiments: evolving analytic approaches.* New York: Russell Sage.

Bloom, H.S., Richburg-Hayes, L., and Black, A.R. (2007). Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis, 29*(1), 30–59.

Boruch, R.F. (1997). *Randomized experiments for planning and evaluation: a practical guide.* Thousand Oaks, CA: Sage.

Bracy, G.W. (2004). Research: international comparisons—less than meets the eye. *Phi Delta Kappan, 85*(6), 477–80.

Brandt, W.C., and Hutchinson, C. (2006). *Romulus Community Schools comprehensive school reform evaluation—spring/summer 2006.* Naperville, IL: Learning Point Associates. Retrieved September 25, 2007, from www.compasslearning.com/files/Romulus_Report_2.pdf.

Business Coalition for Education Reform. (1998, May). *The formula for success: a business leader's guide to supporting math and science achievement.* Washington, DC: U.S. Department of Education.

Campbell, P.B., and Clewell, B.C. (1999). Science, math, and girls. *Education Week 19*(2), 50–52.

Caraisco-Alloggiamento, J. (2008). *A comparison of the mathematics achievement, attributes, and attitudes of fourth-, sixth-, and eighth-grade students.* Unpublished doctoral dissertation, St. John's University, School of Education and Human Services, New York.

Clariana, R. (2007). *Odyssey school effectiveness report: Pemberton Township School District.* Retrieved August 30, 2008, from www.compasslearning.com/files/Pemberton_NJ.pdf.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

CompassLearning, Inc. (2005). *CompassLearning Odyssey® school effectiveness report: Boone County School District*. Retrieved August 30, 2008, from www.compasslearning.com/files/DanielBooneAreaSchoolDistrict_PA.pdf.

CompassLearning, Inc. (2006). *CompassLearning Odyssey® school effectiveness report: Lillie Burney Elementary School*. Retrieved August 30, 2008, from www.compasslearning.com/files/Hattiesburg_MS.pdf.

CompassLearning, Inc. (2007). *Impact of CompassLearning Odyssey® reading/language arts & mathematics on NWEA RIT scores and lexile range*. Retrieved August 30, 2008 from www.compasslearning.com/files/Akron.pdf.

CompassLearning, Inc. (2008a). *Elementary school uses technology to improve math scores. (Scotch Elementary School)*. Retrieved August 30, 2008, from www.compasslearning.com/files/SER_Scotch.pdf.

CompassLearning, Inc. (2008b). *Odyssey® helps Milwaukee students improve performance on NWEA MAP Test*. Retrieved August 30, 2008, from www.compasslearning.com/files/SER_Milwaukee.pdf.

CTB/McGraw-Hill. (2000). *TerraNova: frequently asked questions second edition*. Retrieved August 30, 2008, from www.ctb.com/terranova_faq.pdf.

Deno, S.L. (2003). Developments in curriculum-based measurement. *Journal of Special Education, 37*(3), 184–92.

Elledge, A, Le Floch, K.C., Taylor, J., and Anderson, L. (2009). *State and local implementation of the No Child Left Behind Act*. Volume V, *Implementation of the 1 percent rule and 2 percent interim policy options*. Washington, DC: U.S. Department of Education.

Everyday Math (2009). *The University of Chicago School Mathematics Project*. Retrieved September 18, 2009, from http://everydaymath.uchicago.edu.

Faulkner, L.R., Benbow, C.P., Ball, D.L., Boykin, A.W., Clements, D.H., Embretson, S., Fennell, F., Fristedt, B., et al. (2008). *Final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education. Retrieved January 20, 2009, from www.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf.

Fuchs, L. S., Deno, S. L., and Mirkin, P. K. (1984). Effects of frequent curriculum-based measurement of evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal, 21*(2), 449–60.

Fuchs, L. S., and Fuchs, D. (2002). Curriculum-based measurement: describing competence, enhancing outcomes, evaluating treatment effects, and identifying treatment nonresponders. *Peabody Journal of Education, 77*(2), 64–84.

Fuchs, L. S., Fuchs, D., Prentice, K., Burch, M., Hamlett, C. L., Owen, R., Hosp, M., and Jancek, D. (2003). Explicitly teaching for transfer: effects on third-grade students' mathematical problem solving. *Journal of Educational Psychology, 95*(2), 293–305.

Gin, S.B. (2001). *Mathematics: the path to math success*. Allen, TX: Benziger.

Gonzalez, P., Guzman, J.C., Partelow, L., Pahlke, E., Jocelyn, L., Kastberg, D., and Williams, T.

(2004). *Highlights from the Trends in International Mathematics and Science Study (TIMSS)* (NCES 2005-005). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Gonzalez, P., Williams, T., Jocelyn, L., Roey, S., Kastberg, D., and Brenwald, S. (2009). *Highlights from TIMSS 2007: mathematics and science achievement of U.S. fourth and eighth-grade students in an international context (*NCES 2009-001). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Graham, J.W. (2009). Missing data analysis: making it work in the real world. *Annual Review of Psychology, 60*, 549–76.

Harcourt-School. (2009). *Harcourt School Math.* Retrieved September 18, 2009, from www.harcourtschool.com.

Houghton-Mifflin. (2009a). *Houghton-Mifflin Math.* Retrieved September 18, 2009, from www.eduplace.com/math/mw.

Houghton-Mifflin. (2009b). *Houghton-Mifflin Math Central.* Retrieved September 18, 2009, from www.eduplace.com/math/mathcentral/index.html.

Investigations. (2009). *Investigations in number, data, and space.* Retrieved September 18, 2009, from http://investigations.terc.edu.

Jitendra, A. K. (2007). *Solving math word problems: teaching students with learning disabilities using schema-based instruction.* Austin, TX: PRO-ED.

Lipsey, M.W., and Wilson, D.B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

Liu, O.L., and Wilson, M. (2009). Gender differences and similarities in PISA 2003 mathematics: a comparison between the United States and Hong Kong. *International Journal of Testing, 9*(1), 20–40.

Luke, D.A. (2004). *Multi-level modeling.* Thousand Oaks, CA: Sage.

Macmillan McGraw-Hill. (2009). *Math Connects.* Retrieved September 18, 2009, from www.macmillanmh.com/math/2003/student/index.html.

Martin, R.L. (2005). *Effects of cooperative and individual integrated learning system on attitudes and achievement in mathematics.* Unpublished doctoral dissertation, Florida International University, Miami.

McCaffrey, D.F., Hamilton, L.S., Stecher, B.M., Klein, S.P., Bugliari, D., and Robyn, A. (2001). Interactions among instructional practices, curriculum, and student achievement: the case of standards-based high school mathematics. *Journal for Research in Mathematics Education, 32*(5), 493–517

Moore, D.S., McCabe, G.P., and Craig, B.A. (2009). *Introduction to the practice of statistics.* New York: W.H. Freeman and Company.

National Assessment of Educational Progress. (2007). *The Nation's Report Card.* Retrieved August 30,

2008, from http://nces.ed.gov/nationsreportcard.

National Commission on Excellence in Education. (1983). *A nation at risk: the imperative for educational reform: an open letter to the American people.* A report to the nation and the secretary of education. Washington, DC: National Commission on Excellence in Education.

National Council of Teachers of Mathematics. (2008). *Principles and standards for school mathematics.* Retrieved August 30, 2008, from www.nctm.org/standards/content.aspx?id=268.

National Mathematics Advisory Panel. (2008). *Reports of the task groups and subcommittees.* Washington, DC: National Mathematics Advisory Panel.

National Research Council. (2002). *Scientific research in education: committee on scientific principles for education research.* Washington, DC: National Academy Press.

Neuschmidt, O., Barth, J., and Hastedt, D. (2008). Trends in gender differences in mathematics and science (TIMSS 1995-2003). *Studies in Educational Evaluation*, *34*(2), 56–72.

No Child Left Behind Act of 2001. (2009). Pub. L. No. 107–110, 115 Stat. 1425. Retrieved August 30, 2009, from www.ed.gov/policy/elsec/leg/esea02/index.html.

Pearson. (2009). *Pearson Scott Foresman.* Retrieved September 18, 2009, from www.pearsonschool.com/index.cfm?locator=PSZ1B7.

Raudenbush, S.W., and Bryk, A.S. (2002*). Hierarchical linear models: applications and data analysis methods.* Thousand Oaks, CA: Sage.

Raudenbush, S.W., Bryk, A.S., and Congdon, R. (2008*). HLM: hierarchical linear and nonlinear modeling* [Computer program]. Lincolnwood, IL: Scientific Software International.

Raudenbush, S.W., Martinez, A., and Spybrook, J. (2005). Strategies for improving precision in group-randomized experiments. *Educational Evaluation and Policy Analysis, 29*(1), 5–29.

Raudenbush, S.W., Spybrook, J., Liu, X., and Congdon, R. (2005, October). *Optimal design for longitudinal and multi-level research* (version 1.555). Retrieved August 30, 2008, from www.wtgrantfoundation.org/info-url_nocat5241/info-url_nocat.htm.

Rosnow, R.L., and Rosenthal, R. (2003). Effect sizes for experimenting psychologists. *Canadian Journal of Experimenting Psychology, 57*(3), 221–37.

Saxon. (2009). *Saxon Math.* Retrieved September 18, 2009, from http://saxonpublishers.hmhco.com/en/sxnm_home.htm.

Schochet, P.Z. (2005). *Statistical power for random assignment evaluations of education programs.* Princeton, NJ: Mathematica Policy Research.

Shadish, W.R., Cook, T.D., and Campbell, D.T. (2001). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin.

Sowell, E. (1989). Effects of manipulative materials in mathematics instruction. *Journal for Research in Mathematics Education, 20*(5), 498–505.

Spybrook, J., Raudenbush, S.W., Liu, X., and Congdon, R. (2006). *Optimal Design for longitudinal and multilevel research: documentation for the "Optimal Design" software.* National Institute of Mental Health and William T. Grant Foundation.

Stonewater, J.K. (1996). The standards observation form: feedback to teachers on classroom implementation of the standards. *School Science and Mathematics, 96*(6), 290–97.

Tournaki, N. (2003). The differential effects of teaching addition through strategy instruction versus drill and practice to students with and without learning disabilities. *Journal of Learning Disabilities, 36*(5), 449–58.

Trends in International Mathematics and Science Study. (2003). Retrieved August 30, 2008, from www.nces.ed.gov/timss/results03.asp.

U.S. Department of Education, National Center for Education Statistics, Common Core of Data Public School Universe. (2008). Retrieved September 1, 2005, from www.nces.ed.gov/ccd.

Wiersma, W., and Jurs, S.G. (2005). *Research methods in education.* Boston, MA: Pearson.

Wijekumar, K., and Hitchcock, J. (2006). *The Effects of CompassLearning Odyssey® Math Software on the mathematics achievement of selected fourth grade students in the Mid-Atlantic Region: a multi-site cluster randomized trial.* Available on request from the U.S. Department of Education, Institute of Education Sciences, Washington, DC.