NCEE 2011-4002

U. S. DEPARTMENT OF EDUCATION

Achievement Effects of Four Early Elementary School Math Curricula

Findings for First and Second Graders

Executive Summary

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE

Institute of Education Sciences

Achievement Effects of Four Early Elementary School Math Curricula

Findings for First and Second Graders

Executive Summary

October 2010

Roberto Agodini
Barbara Harris
Melissa Thomas
Mathematica Policy Research, Inc.

Robert Murphy
Lawrence Gallagher
SRI International

Audrey Pendleton
Project Officer
Institute of Education Sciences

NCEE 2011-4002
U.S. DEPARTMENT OF EDUCATION

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

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
John Q. Easton
Director

National Center for Education Evaluation and Regional Assistance
Rebecca Maynard
Commissioner

October 2010

This report was prepared for the Institute of Education Sciences under Contract No. ED-04-CO-0112/0003. The project officer was Audrey Pendleton in the National Center for Education Evaluation and Regional Assistance.

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

To order copies of this report,

- Write to ED Pubs, Education Publications Center, U.S. Department of Education, P.O. Box 22207, Alexandria, VA 22304.
- Call in your request toll free to 1-877-4ED-Pubs. If 877 service is not yet available in your area, call 800-872-5327 (800-USA-LEARN). Those who use a telecommunications device for the deaf (TDD) or a teletypewriter (TTY) should call 800-437-0833.
- Fax your request to 703-605-6794.
- Order online at www.edpubs.org.

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

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

ACHIEVEMENT EFFECTS OF FOUR EARLY ELEMENTARY SCHOOL MATH CURRICULA: FINDINGS FOR FIRST AND SECOND GRADERS

EXECUTIVE SUMMARY

National achievement data show that elementary school students in the United States, particularly those from low socioeconomic backgrounds, have weak math skills (National Center for Education Statistics 2009). In fact, data show that, even before they enter elementary school, children from disadvantaged backgrounds are behind their more advantaged peers in basic competencies such as number-line ordering and magnitude comparison (Rathburn and West 2004). Furthermore, after a year of kindergarten, disadvantaged students still have less extensive knowledge of mathematics than their more affluent peers (Denton and West 2002).

This study examines whether some early elementary school math curricula are more effective than others at improving student math achievement in disadvantaged schools.[1] A small number of curricula, which are based on different approaches for developing student math skills, dominate elementary math instruction—7 curricula make up 91 percent of those used by K–2 educators, according to a 2008 survey (Resnick et al. 2010). Little rigorous evidence exists to support one approach over another, however, which means that research does not provide educators with much useful information when choosing a math curriculum to use.

This study helps to fill that knowledge gap by examining the relative student achievement effects of four elementary school math curricula during the first year of implementation in the first and second grades:

- ***Investigations in Number, Data, and Space (Investigations)*** is published by Pearson Scott Foresman (Wittenburg et al. 2008a) and uses a student-centered approach encouraging metacognitive reasoning and drawing on constructivist learning theory. The lessons focus on understanding, rather than on students answering problems correctly, and build on students' knowledge and understanding. Students are engaged in thematic units of three to eight weeks in which they first investigate and then discuss and reason about problems and strategies.

- ***Math Expressions*** is published by Houghton Mifflin Harcourt (Fuson 2009a; Fuson 2009b) and blends student-centered and teacher-directed approaches to mathematics. Students question and discuss mathematics but are also explicitly taught effective procedures. There is an emphasis on using multiple specified objects, drawings, and language to represent concepts and also on learning through the use of real-world situations. Students are expected to explain and justify their solutions.

- ***Saxon Math (Saxon)*** is published by Harcourt Achieve (Larson 2008) and is a scripted curriculum that blends teacher-directed instruction of new material with daily distributed practice of previously learned concepts and procedures. The teacher introduces concepts or efficient strategies for solving problems. Students observe and then receive guided practice, followed by distributed practice. Students hear the correct answers and are explicitly taught procedures and strategies. Frequent

monitoring of student achievement is built into the program. Daily routines are extensive and emphasize practice of number concepts and procedures and use of representations.

- ***Scott Foresman-Addison Wesley Mathematics (SFAW)*** is published by Pearson Scott Foresman (Charles et al. 2005a; Charles et al. 2005b) and is a basal curriculum that combines teacher-directed instruction with a variety of differentiated materials and instructional strategies. Teachers select the materials that seem most appropriate for their students, often with the help of the publisher. The curriculum is based on a consistent daily lesson structure, which includes direct instruction, hands-on exploration, the use of questioning, and practice of new skills.

Generally speaking, the curricula vary in the extent to which they emphasize student-centered or teacher-directed approaches.

A randomized controlled trial involving 110 elementary schools was implemented to determine the relative effects of the curricula—about a quarter of the schools were randomly assigned to each of the study's four curricula. Random assignment of curricula to schools was conducted separately for each participating district, which established an experiment in each study district.

Among the 110 schools, 39 (cohort one) began study participation during the 2006–2007 school year and during that first year, curriculum implementation occurred only in the first grade. The remaining 71 schools (cohort two) began study participation during the 2007–2008 school year and during that first year, curriculum implementation occurred in both the first and second grades—except in one school, where curriculum implementation occurred only in the second grade.

The study's first report examined first-grade effects during the first year of curriculum implementation among the 39 cohort-one schools (Agodini et al. 2009). Implementation analyses indicated that all teachers received training on their assigned curriculum and, according to teacher surveys, nearly all (99 percent in the fall, and 98 percent in the spring) reported using their assigned curriculum as their core curriculum. In terms of progress with the curricula, as of the spring survey, 88 percent of teachers reported completing at least 80 percent of their assigned curriculum's lessons. This progress with the lessons is consistent with the timing of the spring survey, which was administered about 80 percent through the school year. There was one notable difference in math instruction between the curriculum groups—on average, Saxon teachers reported spending one more hour on math instruction per week than did teachers in the other curriculum groups. Analyses of first-grade math achievement indicated that there were significant differences in achievement across the curriculum groups. In particular, after one year of study participation, average spring first-grade math achievement of Math Expressions and Saxon students was similar and higher than both Investigations and SFAW students. Achievement of the latter two groups (Investigations and SFAW) was similar.

The current report updates the first report in two ways. First, it examines first-grade effects during the first year of curriculum implementation among all study schools (cohort-one and cohort-two schools combined). Given the school-level curriculum implementations described

above, this first-grade analysis is based on 109 schools—39 from cohort one and 70 from cohort two (as mentioned above, one of the 71 cohort-two schools did not implement its assigned curriculum in the first grade). The other way in which the current report updates the previous one is by examining second-grade effects during the first year of curriculum implementation among the 71 cohort-two schools (as mentioned above, the cohort-one schools did not implement the curricula in the second grade during their first year of study participation).[2]

The key findings in this report include the following:

- **Teachers used their assigned curriculum, and the instructional approaches of the four curriculum groups differed as expected.** At least 98 percent of teachers reported using their assigned curriculum, according to fall and spring surveys. Classroom observations conducted by the study team revealed that the instructional approaches of the four curriculum groups differed as expected—student-centered instruction and peer collaboration were highest in Investigations classrooms, and teacher-directed instruction was highest in Saxon classrooms. These curriculum-group differences, as well as all others that are noted, are statistically significant at the 5 percent level of confidence, which means that there is no more than a 5 percent chance that the differences mentioned occurred by chance.

  Math instruction varied in other notable ways across the curriculum groups. Saxon teachers reported spending an average of about one more hour on math instruction per week than did teachers in the other curriculum groups. The number of lessons taught in many math content areas also differed across the curriculum groups. In first-grade classrooms, the number of lessons taught in 15 of the 20 content areas examined was significantly different across the curriculum groups. In second-grade classrooms, the number of lessons taught in 19 of 20 content areas examined was significantly different across the curriculum groups. When looking at the six pair-wise comparisons that can be made between the curricula for each significantly different content area,[3] some curriculum pair differences are significant whereas others are not; there is no clear pattern to which curriculum pair differences are consistently significant across the content areas.

- **In terms of student math achievement, the curriculum used by the study schools mattered.** In first grade classrooms, average math achievement of Math Expressions students was 0.11 standard deviations higher than that of both Investigations and SFAW students; in second grade classrooms average math achievement of Math Expressions and Saxon students was 0.12 and 0.17 standard deviations higher than that of SFAW students, respectively. None of the other curriculum differentials are statistically significant. (As mentioned above, the study's first report based on cohort-one schools showed that average spring first-grade math achievement of Math Expressions and Saxon students was similar and higher than both Investigations and SFAW students.)

- **The curriculum used in different contexts also mattered, and some of these findings are consistent with findings based on all students whereas others are not.** The study examined the relative effects of the curricula for subgroups of schools and teachers with different characteristics, and for the schools and teachers in each

study district.[4] Among the first-grade subgroups, 22 curriculum differentials are statistically significant, of which 14 are consistent with the findings based on all first graders—that is, average math achievement of Math Expressions students was higher than that of Investigations and SFAW students. Among the 8 statistically significant differentials that are not consistent, 4 of them indicate that average math achievement of Saxon students was higher than that of Investigations students, 3 indicate that average achievement of Saxon students was higher than SFAW students, and the last one indicates that achievement of Investigations students was higher than Saxon students. Among the second-grade subgroups, 23 curriculum differentials are statistically significant, of which 16 are consistent with the findings based on all second graders—that is, average math achievement of Math Expressions and Saxon students was higher than that of SFAW students. Among the 7 statistically significant differentials that are not consistent, 4 indicate that average math achievement of Saxon students was higher than Investigations students, 2 show that average achievement of Investigations students was higher than SFAW students, and the last one shows that achievement of Saxon students was higher than Math Expressions students.

Below we discuss features of the study that help establish the context for the findings. We also provide more details about the overall first- and second-grade student achievement results summarized above, including the size of the relative curriculum effects.

Study Participants

The 110 elementary schools included in the evaluation were recruited by the study team and are not a representative sample of all elementary schools in the United States, but they are geographically dispersed and they are in areas with different levels of urbanicity. The participating schools also serve a higher percentage of students eligible for free or reduced-price meals than the average U.S. elementary school. As the national achievement data mentioned earlier show, identifying ways to improve math achievement of students from low socioeconomic backgrounds is critical. Focusing on disadvantaged schools is also consistent with the policy interest that underlies Title I of the No Child Left Behind Act for studying effective approaches to help low-income children meet state standards for academic achievement.

Outcome Measure

To measure the achievement effects of the curricula, the study team tested students at the beginning and end of the school year using the math assessment developed for the Early Childhood Longitudinal Study-Kindergarten Class of 1998–99 (ECLS-K) (West et al. 2000). The ECLS-K assessment is a nationally normed test designed to measure achievement gains both within and across elementary grades. The first- and second-grade results are based on students who were tested in both the fall and spring in those respective grades.

The assessment includes questions in five math content areas: (1) number sense, properties, and operations; (2) measurement; (3) geometry and spatial sense; (4) data analysis, statistics, and probability; and (5) patterns, algebra, and functions. On the first-grade test, about three-quarters

of the items can be classified as number sense, properties, and operations; the remaining items are predominantly related to data analysis, statistics, and probability and patterns, algebra, and functions. On the second-grade test, about half of the test is comprised of items pertaining to number sense, properties, and operations; the other half is predominantly related to measurement; geometry and spatial sense; and patterns, algebra, and functions.

Other Data Collection

To help interpret the measured achievement effects, teachers completed surveys about curriculum implementation, and the study team observed each first- and second-grade classroom once during the school year. Together, the survey and observation data are useful for assessing teacher participation in curriculum training, use of the assigned curriculum, and supplementation of the assigned curriculum with other materials. The data were also useful for assessing adherence to each curriculum's specific features and for examining curriculum-group differences in teaching approaches and practices that could be measured consistently across the curricula.

Relative Effects of the Curricula

The graphs in Figure 1 summarize the achievement results for first- and second-grade students. Each graph includes a symbol for each of the four curricula, where the dot in the middle of each symbol indicates the average spring math score of students in the respective curriculum groups, adjusted for the baseline characteristics of students, teachers/classrooms, and schools;[5] the bars that extend from each dot represent the 95 percent confidence interval around each average score. As described in Chapter III, hierarchical linear modeling (HLM) techniques, which account for the extent to which students are clustered in classrooms and schools, were used to adjust the average spring scores for baseline characteristics and to calculate the 95 percent confidence interval around each score. Curricula with non-overlapping confidence intervals have average scores that are significantly different at the 5 percent level—the statistical significance criterion we used in this study.

The results discussed below are presented in effect size units, which were calculated by dividing each pair-wise curriculum comparison by the pooled standard deviation of the spring score for the two curricula being compared—Hedges' *g* formula (with the correction for small-sample bias) was used to calculate the effect sizes. Chapter III, Table III.2 presents the magnitude and statistical significance for the six unique pair-wise curriculum comparisons at each grade level. Appendix D, Table D.5 presents the simple average (that is, non-HLM-adjusted) and standard deviation of the fall and spring math scores, and the average gain (spring minus fall score), separately by grade and curriculum group.

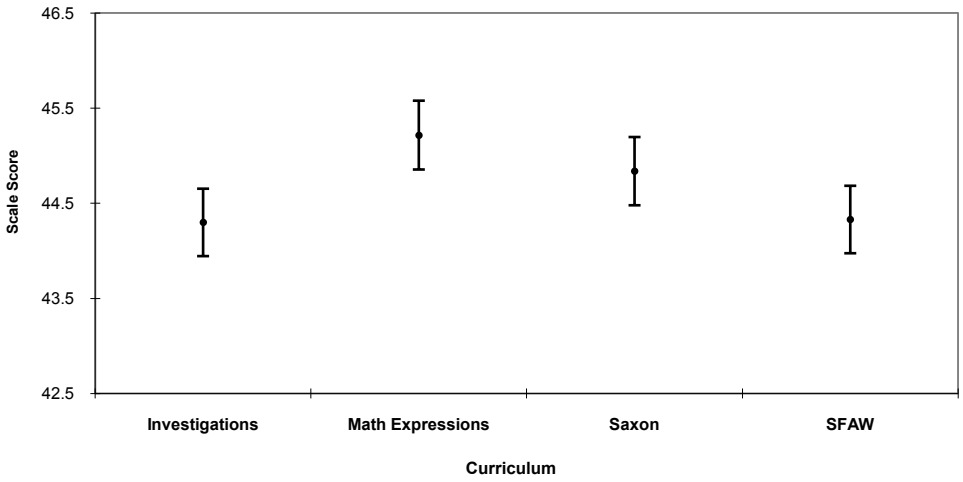
As Figure 1 shows, two of the curriculum differentials are statistically significant at the 5 percent level in both the first and second grades.

- At the first-grade level, average math achievement of Math Expressions students was 0.11 standard deviations higher than that of both Investigations and SFAW students, which is equivalent to moving a student from the 50th to the 54th percentile. None of the other curriculum-pair differentials are statistically significant.[6]

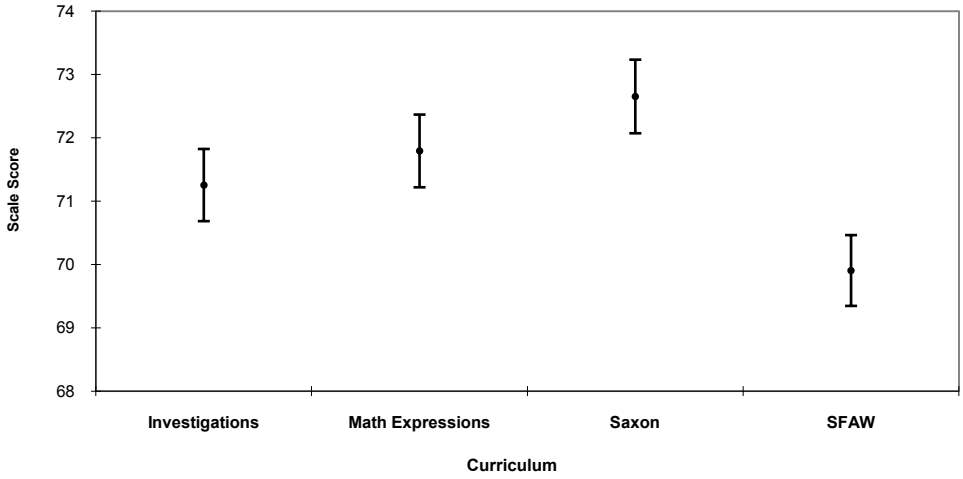
FIGURE 1

AVERAGE HLM-ADJUSTED SPRING STUDENT MATH SCORE WITH CONFIDENCE INTERVAL, BY GRADE AND CURRICULUM

First-Grade Students



Second-Grade Students



Note: The dots in each symbol represent the average HLM-adjusted spring student math score for each curriculum, and the bars that extend from each dot represent the 95 percent confidence interval around each average. Curricula with non-overlapping confidence intervals have significantly different average scores at the 5 percent level. Each curriculum was randomly assigned to about 27 schools, 116 classrooms, and 1,180 students for the first-grade analysis, and to about 18 schools, 82 classrooms, and 835 students for the second-grade analysis. Chapter I, Table I.3 provides the exact school, classroom, and student sample sizes that are the basis for these results.

- At the second-grade level, average math achievement of Math Expressions and Saxon students was 0.12 and 0.17 standard deviations higher than that of SFAW students, respectively, which is equivalent to moving a student from the 50th to the 55th or 57th percentile. None of the other curriculum-pair differentials are statistically significant.

These findings are based on statistical tests that have not been adjusted for the six unique pair-wise curriculum comparisons that can be made. Results based on statistical tests that have been adjusted for the multiple comparisons made indicate that only the Saxon-SFAW differential of 0.17 standard deviations for second graders is statistically significant. There is a large literature that considers the issue of multiple comparison adjustments, but, to our knowledge, there is no consensus about whether statistical tests should or should not be adjusted (see, for example, Saville 1990 and Westfall et al. 1999). For this reason, we present both sets of results.

What the Relative Curriculum Effects Include

The relative effects of the curricula reflect all differences between the curricula, including differences in teacher training, instructional strategies, content coverage, and curriculum materials. Of course, the relative effects ultimately depend on how teachers implemented their curriculum, and actual implementation reflects what publishers and teachers achieved, not some level of implementation specified by the study.

What Accounts for the Relative Curriculum Effects Observed?

The four curriculum groups differ along several implementation measures, including the amount of teacher curriculum training, amount of time teachers spent on math instruction, number of lessons taught in various math content areas, and scales about instructional approaches. We conducted correlational analyses focusing on one curriculum pair at a time, for the curriculum pairs that had significantly different achievement. For those significant curriculum-pair differentials, we examined whether the teaching approaches and practices that are significantly different across the four curriculum groups are related to student achievement of the curriculum pairs with significantly different achievement.

For three of the four curriculum-pair differentials that are statistically significant across the two grade levels, the results show that the student achievement differences are related to differences in the teaching approaches and practices of these curriculum pairs. The curriculum differentials that are related to the implementation measures examined include both of the first-grade differentials (Math Expressions-Investigations and Math Expressions-SFAW) that are statistically significant, and one of the two second-grade differentials (Saxon-SFAW) that is statistically significant. The teaching approaches and practices that were related to the curriculum differentials include curriculum training, math instructional time, coverage in many math content areas, and at least one of the scales about instructional approaches. None of the teaching approaches and practices examined was related to the other second-grade differential that is statistically significant (Math Expressions-SFAW). It is important to note, however, that this part of the analysis was confined to identifying correlational patterns, which may not be causal.

Next Steps for the Study

Some of the schools participated in the study for a second year, and a smaller number participated for a third (the last year of the study). In those subsequent years, curriculum implementation was repeated in grades where it began, and expanded to higher grades. For example, during the second year of participation for cohort-one schools, curriculum implementation was repeated in the first grade and expanded to the second. Data from these follow-up years can be used to examine the relative effects of the curricula among teachers and students that have two-to-three years of experience with them, and a future report is planned that will present results based on those data.

REFERENCES

Agodini, Roberto, Barbara Harris, Sally Atkins-Burnett, Sheila Heaviside, Timothy Novak, and Robert Murphy. "Achievement Effects of Four Early Elementary School Math Curricula: Findings from First Graders in 39 Schools." Report submitted to the U.S. Department of Education, Institute of Education Sciences. Princeton, NJ: Mathematica Policy Research, February 2009.

Charles, Randall, Warren Crown, Francis (Skip) Fennell, Janet H. Caldwell, Mary Cavanagh, Dinah Chancellor, Alma B. Ramirez, Jeanne F. Ramos, Kay Sammons, Jane F. Schielack, William Tate, Mary Thompson, and John A. Van de Walle. *Scott Foresman-Addison Wesley Mathematics.* Grade 1. Glenview, IL: Pearson Scott Foresman, 2005a.

Charles, Randall, Warren Crown, Francis (Skip) Fennell, Janet H. Caldwell, Mary Cavanagh, Dinah Chancellor, Alma B. Ramirez, Jeanne F. Ramos, Kay Sammons, Jane F. Schielack, William Tate, Mary Thompson, and John A. Van de Walle. *Scott Foresman-Addison Wesley Mathematics.* Grade 2. Glenview, IL: Pearson Scott Foresman, 2005b.

Denton, K., and Jerry West. *Children's Reading and Mathematics Achievement in Kindergarten and First Grade*. Publication No. NCES 2002-125. Washington, DC: U.S. Department of Education, National Center for Education Statistics, 2002.

Fuson, Karen C. *Math Expressions.* Grade 1. Orlando, FL: Houghton Mifflin Harcourt Publishing Company, 2009a.

Fuson, Karen C. *Math* Expressions. Grade 2. Orlando, FL: Houghton Mifflin Harcourt Publishing Company, 2009b.

Larson, Nancy. *Saxon Math*. Orlando, FL: Harcourt Achieve Inc., 2008.

National Center for Education Statistics (NCES). *The Nation's Report Card: Mathematics 2009.* Publication No. NCES-2010-451. Washington, DC: Institute of Education Sciences, U.S. Department of Education, 2009.

Rathburn, A., and J. West. *From Kindergarten Through Third Grade: Children's Beginning School Experiences.* Publication no. NCES-2004-007. U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office, 2004.

Resnick, Robert M., Glenn Sanislo, and Stephanie Oda. *The Complete K–12 Report: Market Facts and Segment Analyses.* Rockaway Park, NY: Education Market Research, 2010.

Saville, D.J. "Multiple Comparison Procedures: The Practical Solution." *The American Statistician*, vol. 44, 1990, pp. 174–180.

West, Jerry, Kristin Denton, and Elvira Germino-Hausken. *America's Kindergartners*. Publication No. NCES 2000-070. Washington, DC: U.S. Department of Education, National Center for Education Statistics, 2000.

Westfall, P.H., R. Tobias, D. Rom, R. Wolfinger, and Y. Hochberg. *Multiple Comparisons and Multiple Tests Using SAS.* Cary, NC: SAS Institute, Inc., 1999.

Wittenberg, Lucy, K. Economopoulos, V. Bastable, K.H. Bloomfield, K. Cochran, D. Earnest, A. Hollister, N. Horowitz, E. Leidl, M. Murray, Y. Oh, B.W. Perry, S.J. Russell, D. Schifter, and K. Sillman. *Investigations in Number, Data, and Space.* 2nd ed. Glenview, IL: Pearson Scott Foresman, 2008a.

ENDNOTES

1. The context for the study is "disadvantaged" schools, which is defined as those that have a relatively high schoolwide Title I eligibility rate—57 percent of the study's elementary schools are schoolwide Title I eligible, compared to 44 percent of U.S. elementary schools. The Title I program provides financial assistance to schools with high numbers or percentages of poor children to help all students meet state academic standards. Schools in which children from low-income families make up at least 40 percent of enrollment are eligible to use Title I funds for schoolwide programs that serve all children in the school.

2. Some of the cohort-one schools participated in the study during the 2007–2008 school year (the year when the cohort-two schools began study participation). In this second year of participation, curriculum implementation was repeated in the first grade and expanded to the second. As mentioned below, these data, together with data collected in a subset of cohort-one and cohort-two schools during the 2008–2009 school year (the last year of the study), will be examined in a third planned report.

3. With the four curricula included in the study, six unique pair-wise comparisons of student achievement can be made: (1) Investigations relative to Math Expressions, (2) Investigations relative to Saxon, (3) Investigations relative to SFAW, (4) Math Expressions relative to Saxon, (5) Math Expressions relative to SFAW, and (6) Saxon relative to SFAW.

4. Subgroups were constructed separately for each grade. Baseline measures of school characteristics were used to create five subgroups that include students in schools with different math achievement (three subgroups), and different poverty status (two subgroups). Baseline measures of teacher characteristics were used to create eight subgroups that include students in classrooms led by teachers with different levels of education (two subgroups), experience (two subgroups), and math content and pedagogical knowledge (two subgroups), and teachers who did and did not have prior experience with their assigned curriculum (two subgroups). Examining results for each study district is supported by the study's design that created an experiment in each district, as mentioned above.

5. Student characteristics included fall ECLS-K math test score, age at fall test, number of days between the start of the school year and the fall test, number of days between the fall and spring tests, gender, race/ethnicity, whether the student is limited English proficient or is an English language learner, and whether the student has an individualized education plan or receives special services. Teacher/classroom characteristics included teacher race, education, experience, prior use of the assigned curriculum at the K–3 level, and score on the math content and pedagogical test administered before curriculum training; and three classroom characteristics that may affect student achievement—class size, variance of the fall student math score, and skewness of the score. School characteristics included curriculum assigned to the school, Title I eligibility, the percentage of students eligible for free or reduced-price meals, and the random assignment block.

6. As mentioned above, the study's first report, which examined first-grade effects during the first year of study participation among the 39 cohort-one schools, found that average spring first-grade math achievement of Math Expressions and Saxon students was similar and higher than both Investigations and SFAW students. Achievement of the latter two groups (Investigations and SFAW) was similar. In particular, average spring first-grade math achievement of Math Expressions and Saxon students was 0.30 standard deviations higher than Investigations students, and 0.24 standard deviations higher than SFAW students.

This page intentionally left blank for double-sided copying.


DEPARTMENT OF EDUCATION
UNITED STATES OF AMERICA