

# Achievement Effects of Four Early Elementary School Math Curricula

Findings from First Graders in 39 Schools



# Achievement Effects of Four Early Elementary School Math Curricula

Findings from First Graders in 39 Schools

February 2009

**Roberto Agodini**  
**Barbara Harris**  
**Sally Atkins-Burnett**  
**Sheila Heaviside**  
**Timothy Novak**  
Mathematica Policy Research, Inc.

**Robert Murphy**  
SRI International

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

**U.S. Department of Education**

Arne Duncan

*Secretary*

**Institute of Education Sciences**

Sue Betka

*Acting Director*

**National Center for Education Evaluation and Regional Assistance**

Phoebe Cottingham

*Commissioner*

**February 2009**

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.

This publication is in the public domain. Authorization to reproduce it in whole or in part is granted. While permission to reprint this publication is not necessary, the citation should be: Roberto Agodini, Barbara Harris, Sally Atkins-Burnett, Sheila Heaviside, Timothy Novak, and Robert Murphy (2009). *Achievement Effects of Four Early Elementary School Math Curricula: Findings from First Graders in 39 Schools* (NCEE 2009-4052). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

**To order copies of this report,**

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

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

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

## ACKNOWLEDGMENTS

This study was made possible by the collaboration and hard work of many individuals beyond the study authors. We appreciate the willingness of the participating districts, schools, and teachers to use the study's curricula, and to respond to the data requests that are the basis for this report. We also appreciate the willingness of the curriculum publishers to take part in this evaluation. We benefited from useful comments of the study's technical working group: Richard Askey, Douglas Clements, Thomas Cook, Lynn Fuchs, Tom Loveless, Kevin Miller, Donald Rock, and Hung-Hsi Wu.

We thank Melissa Thomas for helping to direct all phases of the student testing and other data collection efforts, and for her contributions to the development of the survey instruments. Alejandra Lopez helped design the teacher surveys. Tom Barton, Tim Bruursema, and Kristina Rall helped design the data collection training programs and manage the data collection. Season Bedell-Boyle and Loring Funaki coordinated the large team of student testers. Mark Brinkley, Andrew Frost, and Joel Zief provided systems support, Erin Slyne programmed the computer-assisted test instruments, and Donsig Jang provided sampling support. We thank Carol Razafindrakato for his programming expertise, and Emily Sama Martin for processing the teacher-reported adherence data. Brian Gill provided useful comments on an earlier version of the report. Marjorie Mitchell and William Garrett produced the report.

Last, but far from least, we thank the team of site recruiters who diligently worked to secure the study's first cohort of participating districts and schools. The recruiters included several report authors, Alex Bogin, Larissa Campuzano, John Deke, Patricia Del Grosso, Benita Kim, Jeffrey Max, and Melissa Thomas.

The Authors



## **DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST**

The research team for this evaluation consists of a prime contractor, Mathematica Policy Research, and a main subcontractor, SRI International. Neither organization nor its key staff have financial interests that could be affected by the findings from the evaluation. None of the study's Technical Working Group members, which were convened by the research team to provide advice on key features of the study, have financial interests that could be affected by the evaluation's findings.



## CONTENTS

Chapter	Page
EXECUTIVE SUMMARY .....	xvii
I INTRODUCTION AND STUDY FEATURES.....	1
A. THE NEED FOR A LARGE-SCALE STUDY OF MATH CURRICULA.....	1
B. DESIGNING THE EVALUATION AND SELECTING THE CURRICULA.....	3
1. Selecting the Curricula.....	4
2. Evaluation Design.....	5
C. RECRUITING PARTICIPANTS .....	6
1. Suitable Districts.....	6
2. Recruiting Districts and Schools.....	7
3. Characteristics of Participants.....	9
D. RANDOM ASSIGNMENT AND STATISTICAL POWER.....	9
E. DATA COLLECTION .....	14
1. Outcome Measure .....	14
2. Other Data Collection .....	16
F. FUTURE PUBLICATION PLANS.....	18
II CURRICULUM IMPLEMENTATION.....	19
A. CONTEXT FOR CURRICULUM IMPLEMENTATION.....	20
B. TEACHER CURRICULUM TRAINING .....	24
1. All Teachers Attended Initial Training on Their Assigned Curriculum.....	24
2. Ninety-Six Percent of Teachers Attended Follow-Up Training on Their Assigned Curriculum .....	27
3. Other Sources of Professional Development .....	28

**CONTENTS** *(continued)*

**II**  
*(continued)*

C. SCHOOL-BASED INSTRUCTIONAL SUPPORT.....	28
1. Seventy-Three Percent of Teachers Had Access to a Math Coach.....	29
2. Teachers Reported Having a Supportive Instructional Environment .....	31
D. SOME BASICS ABOUT TEACHER USE OF THE ASSIGNED CURRICULUM .....	31
1. Nearly All Teachers (99 percent in the fall and 98 percent in the spring) Reported Using Their Assigned Curriculum .....	31
2. Eighty-Eight Percent of Teachers Completed at Least 80 Percent of Their Curriculum .....	33
3. One-Third of Teachers Supplemented with Other Materials.....	33
4. Saxon Teachers Spent One More Hour on Math Instruction per Week .....	37
E. TEACHER ADHERENCE TO THE ESSENTIAL FEATURES OF THE CURRICULA.....	37
1. Descriptions of the Curricula and Teacher Adherence .....	38
a. Investigations .....	39
b. Math Expressions.....	42
c. Saxon.....	45
d. SFAW .....	48
2. Content Coverage.....	50
III CURRICULUM EFFECTS ON FIRST GRADE ACHIEVEMENT .....	53
A. METHODS USED TO CALCULATE CURRICULUM EFFECTS .....	53
B. RELATIVE EFFECTS OF THE CURRICULA .....	58
1. Student Math Achievement was Significantly Higher in Math Expressions and Saxon Schools than in Investigations and SFAW Schools.....	60
2. Some Curriculum Differentials Also Exist in Several Subgroups.....	61
C. NEXT STEPS FOR THE STUDY.....	68
REFERENCES.....	69

**CONTENTS** *(continued)*

TABLE OF ACRONYMS .....	73
APPENDIX A: DATA COLLECTION AND RESPONSE RATES.....	A.1
APPENDIX B: TEACHER-REPORTED FREQUENCY OF IMPLEMENTING OTHER CURRICULUM- SPECIFIC ACTIVITIES .....	B.1
APPENDIX C: GLOSSARY OF CURRICULUM-SPECIFIC TERMS .....	C.1
APPENDIX D: CONSTRUCTING THE ANALYSIS SAMPLES AND ESTIMATING CURRICULUM EFFECTS .....	D.1



## TABLES

<b>Table</b>	<b>Page</b>
I.1	CHARACTERISTICS OF U.S. DISTRICTS AND COHORT-ONE PARTICIPATING DISTRICTS.....10
I.2	CHARACTERISTICS OF U.S. ELEMENTARY SCHOOLS AND COHORT-ONE PARTICIPATING SCHOOLS .....11
I.3	NUMBER OF COHORT-ONE SCHOOLS, CLASSROOMS, AND STUDENTS, BY CURRICULUM.....13
I.4	RESEARCH QUESTIONS AND SUPPORTING DATA COLLECTION EFFORTS: 2006-2007 STUDY PARTICIPANTS .....15
II.1	TEACHER BASELINE CHARACTERISTICS, BY CURRICULUM .....22
II.2	CURRICULA PREVIOUSLY USED BY TEACHERS .....25
II.3	TEACHER TRAINING ON THE ASSIGNED CURRICULUM .....26
II.4	NON-STUDY MATH PROFESSIONAL DEVELOPMENT DURING THE 2006-2007 SCHOOL YEAR.....29
II.5	INSTRUCTIONAL SUPPORT AT STUDY SCHOOLS.....30
II.6	INSTRUCTIONAL CLIMATE AT STUDY SCHOOLS .....32
II.7	TEACHER INSTRUCTION AS REPORTED IN THE FALL .....34
II.8	TEACHER INSTRUCTION AS REPORTED IN THE SPRING .....35
II.9	INVESTIGATIONS: TEACHER-REPORTED FREQUENCY OF IMPLEMENTING ESSENTIAL CURRICULUM ACTIVITIES (N = 31).....41
II.10	INVESTIGATIONS: TEACHER-REPORTED SUCCESS AT FACILITATING DISCUSSIONS FOCUSED ON PROCESS (N = 31).....42
II.11	MATH EXPRESSIONS: TEACHER-REPORTED FREQUENCY OF IMPLEMENTING ESSENTIAL CURRICULUM ACTIVITIES (N = 27).....44
II.12	MATH EXPRESSIONS: TEACHER-REPORTED SUCCESS AT FACILITATING DISCUSSIONS FOCUSED ON PROCESS (N = 27) .....45

**TABLES** *(continued)*

II.13	SAXON: TEACHER-REPORTED FREQUENCY OF IMPLEMENTING ESSENTIAL CURRICULUM ACTIVITIES (N = 30) .....	47
II.14	SFAW: TEACHER-REPORTED FREQUENCY OF IMPLEMENTING ESSENTIAL CURRICULUM ACTIVITIES (N = 32) .....	49
II.15	AVERAGE EMPHASIS IN VARIOUS MATH CONTENT AREAS.....	51
III.1	BASELINE CHARACTERISTICS OF COHORT-ONE SCHOOLS BY CURRICULUM .....	55
III.2	BASELINE CHARACTERISTICS OF COHORT-ONE LONGITUDINAL STUDENTS, BY CURRICULUM .....	56
III.3	AVERAGE DIFFERENCE BETWEEN PAIRS OF CURRICULA IN HLM-ADJUSTED SPRING STUDENT MATH ACHIEVEMENT, IN EFFECT SIZES .....	59
III.4	AVERAGE DIFFERENCE BETWEEN PAIRS OF CURRICULA IN HLM-ADJUSTED SPRING STUDENT MATH ACHIEVEMENT, BY SUBGROUPS AND IN EFFECT SIZES .....	66
A.1	NUMBER OF SCHOOLS AND FIRST GRADE CLASSROOMS PARTICIPATING IN THE STUDY DURING THE 2006-2007 SCHOOL YEAR, BY CURRICULA.....	A.6
A.2	NUMBER AND PERCENTAGE OF CLASSROOMS IN WHICH THE PRIMARY MATHEMATICS TEACHER COMPLETED THE TEACHER KNOWLEDGE ASSESSMENT, AND THE FALL AND SPRING SURVEYS, BY CURRICULA: 2006-2007 SCHOOL YEAR PARTICIPANTS.....	A.8
A.3	PARENT CONSENT RATES BY CURRICULA AND SAMPLED STUDENTS' ENTRY INTO THE STUDY: 2006-2007 SCHOOL YEAR.....	A.9
A.4	NUMBER AND PERCENTAGE OF SAMPLED STUDENTS TESTED AND TYPES OF NONRESPONSE: 2006-2007 SCHOOL YEAR.....	A.11
A.5	NUMBER AND PERCENTAGE OF BASELINE STUDENTS AND NEW ARRIVERS SAMPLED FOR TESTING BY ROUND OF TESTING AND CURRICULA: 2006-2007 SCHOOL YEAR.....	A.11

**TABLES** *(continued)*

A.6	NUMBER OF SAMPLED STUDENTS ENROLLED AND ELIGIBLE FOR TESTING AT BOTH BASELINE AND SPRING FOLLOWUP, AND NUMBER AND PERCENTAGE TESTED IN BOTH THE FALL AND SPRING: 2006-2007 SCHOOL YEAR.....	A.14
A.7	NUMBER OF SAMPLED STUDENTS ENROLLED AND ELIGIBLE FOR TESTING IN THE SPRING AND NUMBER AND PERCENTAGE TESTED, BY CURRICULUM AND TYPE OF SAMPLE—CROSS-SECTION SAMPLE: SPRING 2007.....	A.14
A.8	TESTING DATES AND NUMBER OF DAYS BETWEEN FALL AND SPRING TESTING START DATES AND END DATES, BY CURRICULA: 2006-2007 SCHOOL YEAR.....	A.15
A.9	NUMBER AND PERCENTAGE OF STUDENTS FOR WHOM STUDENT DEMOGRAPHIC RECORDS AND INDIVIDUAL DEMOGRAPHIC ITEMS WERE COLLECTED, BY TYPE OF SAMPLE AND ITEM: 2006-2007.....	A.15
B.1	INVESTIGATIONS: TEACHER-REPORTED FREQUENCY OF IMPLEMENTING OTHER CURRICULUM-SPECIFIC ACTIVITIES (N = 31) .....	B.3
B.2	MATH EXPRESSIONS: TEACHER-REPORTED FREQUENCY OF IMPLEMENTING OTHER CURRICULUM-SPECIFIC ACTIVITIES (N = 27) .....	B.4
B.3	SAXON MATH: TEACHER-REPORTED FREQUENCY OF IMPLEMENTING OTHER CURRICULUM-SPECIFIC ACTIVITIES (N = 31) .....	B.5
B.4	SFAW MATH: TEACHER-REPORTED FREQUENCY OF IMPLEMENTING OTHER CURRICULUM-SPECIFIC ACTIVITIES (N = 30) .....	B.6
D.1	MODEL-BASED IMPUTATION OF MISSING DATA, LONGITUDINAL SAMPLE .....	D.5
D.2	MODEL-BASED IMPUTATION OF MISSING DATA, CROSS-SECTIONAL SAMPLE .....	D.6
D.3	AVERAGE UNADJUSTED STUDENT MATH SCORES, BY CURRICULUM.....	D.8
D.4	HIERARCHICAL LINEAR MODEL ESTIMATES FOR THE LONGITUDINAL SAMPLE .....	D.11

**TABLES** *(continued)*

D.5	HIERARCHICAL LINEAR MODEL ESTIMATES FOR THE CROSS-SECTIONAL SAMPLE .....	D.15
D.6	AVERAGE DIFFERENCE BETWEEN PAIRS OF CURRICULA IN HLM- ADJUSTED SPRING STUDENT MATH ACHIEVEMENT FOR THE CROSS SECTIONAL SAMPLE, IN EFFECT SIZES .....	D.16
D.7	SAMPLE SIZES USED IN SUBGROUPS ANALYSES.....	D.17

## FIGURES

Figure	Page
I.1 DATA COLLECTION TIMELINE DURING THE 2006-2007 SCHOOL YEAR .....	15
III.1 AVERAGE HLM-ADJUSTED SPRING MATH SCORE WITH CONFIDENCE INTERVAL, BY CURRICULUM.....	58
III.2 AVERAGE HLM-ADJUSTED SPRING MATH SCORE IN STANDARD DEVIATIONS, BY DISTRICT AND CURRICULUM.....	63
III.3 AVERAGE HLM-ADJUSTED SPRING MATH SCORE IN STANDARD DEVIATIONS, BY SCHOOL FALL ACHIEVEMENT AND CURRICULUM .....	63
III.4 AVERAGE HLM-ADJUSTED SPRING MATH SCORE IN STANDARD DEVIATIONS, BY SCHOOL FREE/REDUCED PRICE MEALS ELIGIBILITY AND CURRICULUM.....	64
III.5 AVERAGE HLM-ADJUSTED SPRING MATH SCORE IN STANDARD DEVIATIONS, BY TEACHER EDUCATION AND CURRICULUM .....	64
III.6 AVERAGE HLM-ADJUSTED SPRING MATH SCORE IN STANDARD DEVIATIONS, BY TEACHER EXPERIENCE AND CURRICULUM .....	65
III.7 AVERAGE HLM-ADJUSTED SPRING MATH SCORE IN STANDARD DEVIATIONS, BY TEACHER MATH CONTENT/PEDAGOGICAL TEST SCORE AND CURRICULUM.....	65
A.1 FLOW OF SCHOOLS THROUGH THE STUDY .....	A.7
A.2 FLOW OF STUDENTS THROUGH THE STUDY.....	A.12



# ACHIEVEMENT EFFECTS OF FOUR EARLY ELEMENTARY SCHOOL MATH CURRICULA: FINDINGS FROM FIRST GRADERS IN 39 SCHOOLS

## EXECUTIVE SUMMARY

Many U.S. children start school with weak math skills and there are differences between students from different socioeconomic backgrounds—those from poor families lag behind those from affluent ones (Rathburn and West 2004). These differences also grow over time, resulting in substantial differences in math achievement by the time students reach the fourth grade (Lee, Gregg, and Dion 2007).

The federal Title I program provides financial assistance to schools with a high number or percentage of poor children to help all students meet state academic standards. Under the No Child Left Behind Act (NCLB), Title I schools must make adequate yearly progress (AYP) in bringing their students to state-specific targets for proficiency in math and reading. The goal of this provision is to ensure that all students are proficient in math and reading by 2014.

The purpose of this large-scale, national study is to determine whether some early elementary school math curricula are more effective than others at improving student math achievement, thereby providing educators with information that may be useful for making AYP. A small number of curricula dominate elementary math instruction (seven math curricula make up 91 percent of the curricula used by K-2 educators), and the curricula are based on different theories for developing student math skills (Education Market Research 2008). NCLB emphasizes the importance of adopting scientifically-based educational practices; however, there is little rigorous research evidence to support one theory or curriculum over another. This study will help to fill that knowledge gap. The study is sponsored by the Institute of Education Sciences (IES) in the U.S. Department of Education and is being conducted by Mathematica Policy Research, Inc. (MPR) and its subcontractor SRI International (SRI).

## BASIS FOR THE CURRENT FINDINGS

This report presents results from the first cohort of 39 schools participating in the evaluation, with the goal of answering the following research question: What are the relative effects of different early elementary math curricula on student math achievement in disadvantaged schools? The report also examines whether curriculum effects differ for student subgroups in different instructional settings.

***Curricula Included in the Study.*** A competitive process was used to select four curricula for the evaluation that represent many of the diverse approaches used to teach elementary school math in the United States:

- Investigations in Number, Data, and Space (Investigations) published by Pearson Scott Foresman (Russell, Economopoulos, Mokros, Kliman, Wright, Clements, Goodrow, Murray, and Sarama 2006)

- Math Expressions published by Houghton Mifflin Company (Fuson 2006a)
- Saxon Math (Saxon) published by Harcourt Achieve (Larson 2004)
- Scott Foresman-Addison Wesley Mathematics (SFAW) published by Pearson Scott Foresman (Charles, Crown, Fennel, Caldwell, Cavanagh, Chancellor, Ramirez, Ramos, Sammons, Schielack, Tate, Thompson, and Van de Walle 2005)

The process for selecting the curricula began with the study team inviting developers and publishers of early elementary school math curricula to submit a proposal to include their curricula in the evaluation. A panel of outside experts in math and math instruction then reviewed the submissions and recommended to IES curricula suitable for the study. The goal of the review process was to identify widely used curricula that draw on different instructional approaches and that hold promise for improving student math achievement.

***Study Design.*** An experimental design was used to evaluate the relative effects of the study’s four curricula. The design randomly assigned schools in each participating district to the four curricula, thereby setting up an experiment in each district. The relative effects of the curricula were calculated by comparing math achievement of students in the four curriculum groups.

The study does not include a control group of schools (or a “business as usual” group) that continue to use whatever math curriculum they were using before joining the study. The study team decided not to include such a control group because it would contain a variety of curricula used by the participating districts, thereby making it difficult to compare effects of the study’s curricula to effects for this group.

***Participating Districts and Schools.*** The study compares the effects of the selected curricula on math achievement of students in disadvantaged schools. The study team identified and recruited districts that (1) have Title I schools, (2) are geographically dispersed, and (3) contain at least four elementary schools interested in study participation, so all four of the study’s curricula could be implemented in each district.

Participating sites are not a representative sample of districts and schools, because interested sites are likely to be unique in ways that make it difficult to select a representative sample. Interested districts were willing to use all four of the study’s curricula, allowed the curricula to be randomly assigned to their participating schools, and were willing to have the study team test students and collect other data required by the evaluation (as described below). It would have been extremely costly to recruit a representative sample of districts and schools that met these criteria.

The 39 schools examined in this report are contained in four districts that are geographically dispersed in four states and in three regions of the country. The districts also fall in areas with different levels of urbanicity—two districts are in urban areas, one is in a suburban area, and the other is in a rural area.

In this first cohort, curriculum implementation occurred in the first grade during the 2006-2007 school year. Data were collected from the 131 first-grade teachers in the study schools, and from 1,309 students—a random sample of about 10 students in each classroom was sufficient to support the analyses. Each of the four curricula was assigned about 10 schools with 33 classrooms and 325 students. The table below presents the exact number of schools, classrooms, and students included in the analysis, in total and by curriculum group.

NUMBER OF COHORT-ONE SCHOOLS, CLASSROOMS, AND STUDENTS,  
IN TOTAL AND BY CURRICULUM

	All	Curriculum			
		Investigations	Math Expressions	Saxon	SFAW
Schools	39	10	9	9	11
Classrooms	131	33	31	31	36
Average # of classrooms/school	3.4	3.3	3.4	3.4	3.3
Students Both Fall and Spring Tested	1,309	332	314	304	359
Average # of students/classroom	10	10	10	10	10

An inspection of baseline school, teacher, and student characteristics shows that random assignment achieved its objective of creating four groups with similar characteristics before curriculum implementation began. The baseline characteristics include 7 school characteristics (see Table III.1 in the body of the report) 21 teacher characteristics (see Table II.1 in the body of the report), and 7 student characteristics (see Table III.2 in the body of the report), including student fall math achievement. Statistical tests indicate that none of the school and student characteristics are significantly different at the 5 percent level of confidence across the curriculum groups.<sup>1</sup> One of the 21 teacher characteristics (race) is significantly different across the curriculum groups;<sup>2</sup> however, as described in Chapter III, the approach for calculating curriculum effects adjusted for teacher race.

**Statistical Power.** The effect size that can be detected with the first cohort is as small as 0.22, where effect size is defined as a fraction of the standard deviation of the test score. Specifically, the minimum detectable effect (MDE) equals the difference in average student math scores of any two curriculum groups, divided by the pooled standard deviation of the score for the two curricula being compared.<sup>3</sup>

<sup>1</sup> The 5 percent level of confidence means there is no more than a 5 percent chance that the finding (that none of the school and student characteristics are different across the curriculum groups) could have occurred by chance.

<sup>2</sup> At least 93 percent of Investigations, Math Expressions, and Saxon teachers classified themselves as white, whereas 78 percent of SFAW teachers did so.

<sup>3</sup> The MDE calculation accounts for the extent to which students in the first cohort are clustered in classrooms and schools according to their baseline achievement, after adjusting for other baseline student, teacher, and school

The MDE of 0.22 means that, when comparing student achievement of any two curriculum groups, it must differ by at least 15 percent of the gain made by the average first grader from a low income family to be detectable in this report. Chapter I provides more details about the computation of the MDE and what it represents.

***Outcome Measure and Other Data Collection.*** 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, Denton, and Germino-Hausken 2000). The ECLS-K assessment is a nationally normed test that meets the study’s requirements of: assessing knowledge and skills mathematicians and math educators feel are important for early elementary school students to develop; having accepted standards of validity and reliability; being administered to students individually; being able to measure achievement gains over the study’s grade range (which ultimately will include the first, second, and third grades); and being able to accurately capture achievement of students from a wide range of backgrounds and ability levels.

Another important feature of the ECLS-K assessment is that it is an adaptive test, which is an approach used to measure achievement that is tailored to a student’s achievement level. In particular, the test begins by administering to each student a short, first-stage routing test used to broadly measure each examinee’s achievement level. Depending on the score on the routing test, the student is then administered one of three longer, second-stage tests: (1) an easy test, (2) a middle-difficulty test, or (3) a difficult test. Some of the items on the second-stage tests overlap, and this overlap is used to place the scores on the different tests on the same scale. Item response theory (IRT) techniques (Lord 1980) were used to develop the scale score, which, according to the test developers, are the appropriate scores to analyze for our purposes (Rock and Pollack 2002).<sup>4</sup> Adaptive tests are useful for measuring achievement because they limit the amount of time children are away from their classrooms and reduce the risk of ceiling or floor effects in the test score distribution—something that can have adverse effects on measuring achievement gains.

The assessment includes questions in the 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. The items in each of the second-stage tests administered to the study’s first graders can primarily be classified as

---

*(continued)*

characteristics. The calculation also uses the Tukey-Kramer method (Tukey 1952, 1953; Kramer 1956) to account for the six unique pair-wise comparisons that can be made with the study’s four curricula: (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.

<sup>4</sup> Student answers on the assessment were sent to the Educational Testing Service (ETS) for scoring—ETS was a developer of the ECLS-K Mathematics Assessment. A three-parameter IRT model was used to place scores from the different tests students took on the same scale. Reliabilities for the study’s sample (0.93 for the fall score and 0.94 for the spring score) were consistent with the national ECLS-K sample (Rock and Pollack 2002, pp. 5-7 through 5-9)—reliabilities are based on the internal consistency (alpha) coefficients. Also, there were no floor or ceiling effects observed in either the fall or spring scores.

Number Sense, Properties, and Operations, with the remainder from the other areas. The easy test contained only a few items from each of the remaining areas, whereas the middle-difficulty and difficult tests contained more such items. On the middle-difficulty test, the remaining items were mainly about Patterns, Algebra, and Functions, whereas those on the difficult test were mainly about Data Analysis, Statistics, and Probability.

To help interpret the measured effects of the curricula, teachers were surveyed about curriculum implementation. The survey data are useful for assessing teacher participation in curriculum training, usage of the assigned curriculum, and any supplementation with other materials. Teachers also reported their usage of the essential and secondary features of their assigned curriculum, which was useful for assessing adherence to each curriculum. Demographic information about teachers also was collected through the surveys, and student demographics were obtained from school records.

## MAIN FINDINGS

The study's main findings include information about curriculum implementation and the relative effects of the curricula on student math achievement. Statistical tests were used to assess the significance of all the results. Hierarchical linear modeling (HLM) techniques—which account for the extent to which students are clustered in classrooms and schools according to achievement—were used to conduct the statistical tests. When comparing results for pairs of curricula, the Tukey-Kramer method (Tukey 1952, 1953; Kramer 1956) was used to adjust the statistical tests for the six unique pair-wise comparisons that can be made with four curricula, as described above. Only results that are statistically significant at the 5 percent level of confidence are discussed.<sup>5</sup>

Before presenting the main findings, it is worth mentioning the information that is and is not provided by the study. The relative effects of the curricula presented below reflect 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 the curricula, and implementation reflects what publishers and teachers achieved, not some level of implementation specified by the study. Information about curriculum implementation presented in this report is based only on teacher reports—the study team is observing classrooms and plans to present that information in a future report.<sup>6</sup> Also, the relative effects of the curricula are based only on the ECLS-K math assessment administered by the study team—in the third grade and perhaps even the second grade, districts administer their own math assessments to students and the study team is investigating the possibility of obtaining those scores for our future analyses of second and third graders. Lastly, because the participating

---

<sup>5</sup> As mentioned above, the 5 percent level of confidence means there is no more than a 5 percent chance that any finding discussed could have occurred by chance.

<sup>6</sup> Each classroom in the current sample was observed once during the 2006-2007 school year. Those observations are not presented in this report because the reliability of those data cannot be assessed until observations have been completed in all the study schools.

sites are not a representative sample of districts and schools, the design does not support making statements about effects for districts and schools outside of the study.

Curriculum Implementation. The main findings from the implementation analysis are:

- All teachers received initial training from the publishers and 96 percent received follow-up training. Taken together, training varied by curriculum, ranging from 1.4 to 3.9 days.
- Nearly all teachers (99 percent in the fall, 98 percent in the spring) reported using their assigned curriculum as their core math curriculum according to the fall and spring surveys, and about a third (34 percent in fall and 36 percent in spring) reported supplementing their curriculum with other materials.
- Eighty-eight percent of teachers reported completing at least 80 percent of their assigned curriculum.<sup>7</sup>
- On average, Saxon teachers reported spending one more hour on math instruction per week than did teachers of the other curricula.

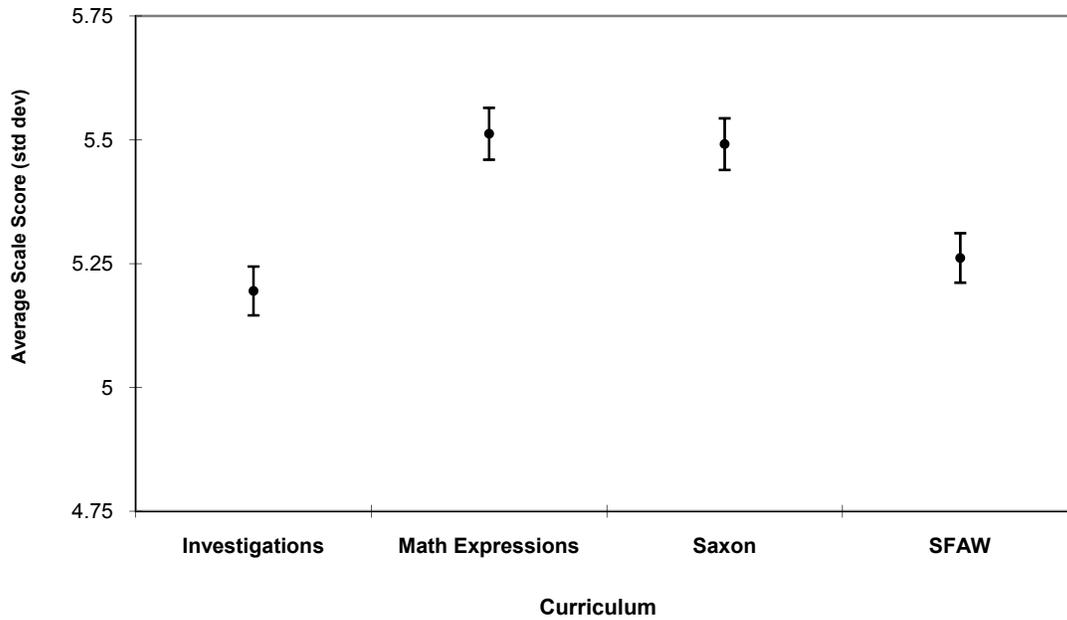
Achievement Effects. The figure below illustrates the relative effects of the study's curricula on student math achievement. The figure includes a symbol for each of the four curricula, where the dot in the middle of each symbol indicates the average spring math score of students in the respective curriculum groups. The average scores are adjusted for baseline measures of several student, teacher, and school characteristics related to student spring achievement (such as student fall math scores) to improve the precision of the results. The bars that extend from each dot represent the 95 percent confidence interval around each average score. HLM techniques were used to calculate the average scores and confidence intervals.

Curricula with non-overlapping confidence intervals have average scores that are significantly different at the 5 percent level of confidence. The results are presented in standard deviations, which means that subtracting the average values (the dots) for any two curricula indicates the effect size of using the first curriculum instead of the second. The effect sizes discussed below were calculated by dividing each pair-wise curriculum comparison by the pooled standard deviation of the spring score for the two curricula being compared, and Hedges' *g* formula (with the correction for small-sample bias) was used to calculate the pooled standard

---

<sup>7</sup> Adherence to the essential features of each curriculum also was examined and is presented in Chapter II. Several analytical approaches can be used to examine adherence, but only one approach could be supported by the relatively small teacher sample sizes that are currently available for each curriculum. We do not make any general statements about adherence in the executive summary because it would be useful to examine whether the results are sensitive to the other analytical approaches, and instead encourage readers interested in the adherence analysis we were able to conduct at this point to see Chapter II. A future planned report (described at the end of the executive summary) will have larger teacher sample sizes that can support the other analyses.

**Average HLM-Adjusted Spring Math Score with Confidence Interval, by Curriculum  
(in standard deviations)**



Note: The dots in each symbol represent the average HLM-adjusted spring math score (in standard deviations) 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 of confidence.

deviations. Appendix D presents averages of the unadjusted math scores (see Table D.3). The relative effects of the curricula described below are similar when based on the simple averages, although the confidence intervals are wider than those based on the HLM-adjusted averages, as expected.

The figure shows that:

- **Student math achievement was significantly higher in schools assigned to Math Expressions and Saxon, than in schools assigned to Investigations and SFAW.** Average HLM-adjusted spring 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. For a student at the 50th percentile in math achievement, these effects mean that the student's percentile rank would be 9 to 12 points higher if the school used Math Expressions or Saxon, instead of Investigations or SFAW.

- **Math achievement in schools assigned to the two more effective curricula (Math Expressions and Saxon) was not significantly different, nor was math achievement in schools assigned to the two less effective curricula (Investigations and SFAW).** The Math Expressions-Saxon and Investigations-SFAW differentials equal 0.02 and -0.07 standard deviations, respectively, and neither is statistically significant.

We also examined whether the relative effects of the curricula differ along six characteristics that differentiate instructional settings: (1) participating districts, (2) school fall achievement, (3) school free/reduced-price meals eligibility, (4) teacher education, (5) teacher experience, and (6) teacher math content/pedagogical knowledge that was measured before curriculum training began using an assessment administered by the study team. These characteristics were used to create 15 subgroups—one for each of the four districts, three based on school fall achievement, and two subgroups for each of the other four characteristics.

Eight of the fifteen subgroup analyses found statistically significant differences in student math achievement between curricula. The significant curriculum differences ranged from 0.28 to 0.71 standard deviations, and all of the significant differences favored Math Expressions or Saxon over Investigations or SFAW. There were no subgroups for which Investigations or SFAW showed a statistically significant advantage.

## **NEXT STEPS FOR THE STUDY**

Another 71 schools joined the study during the 2007-2008 school year (the year after the 39 schools examined in this report joined), and curriculum implementation occurred in both the first and second grades in all participating schools. A follow-up report is planned that will present results based on all 110 schools participating in the evaluation, and for both the first and second grades. The study also is supporting curriculum implementation and data collection during the 2008-2009 school year in a subset of schools, in which implementation will be expanded to the third grade. A third report is planned that will present those results.

## **I. INTRODUCTION AND STUDY FEATURES**

This report presents results for the first cohort of 39 schools participating in a large-scale, national study of four early elementary school math curricula: (1) Investigations in Number, Data, and Space, (2) Math Expressions, (3) Saxon Math, and (4) Scott Foresman-Addison Wesley Mathematics. These curricula represent many of the diverse approaches used to teach elementary school math in the United States, and the study is comparing the relative effects of the curricula on math achievement of students in disadvantaged schools. Experimental methods are being used to determine the relative effects of the curricula.

The results are based on first grade curriculum implementation during the 2006-2007 school year in the 39 cohort-one schools. A future report will be based on all 110 schools participating in the evaluation—an additional 71 schools joined the study during the 2007-2008 school year and curriculum implementation occurred in both the first and second grades in all study schools. The future report will both update the first grade results presented in this report by including the additional schools in the analysis and present results for curriculum implementation in the second grade. The study is sponsored by the Institute of Education Sciences (IES) in the U.S. Department of Education and is being conducted by Mathematica Policy Research, Inc. (MPR) and its subcontractor SRI International (SRI).

The rest of this chapter presents the rationale for the study, describes its key features, and presents more details about its future publication plans. Chapter II presents information about curriculum implementation in the first cohort of schools. Chapter III presents the relative effects of the curricula in those schools, both overall effects and effects for several subgroups.

### **A. THE NEED FOR A LARGE-SCALE STUDY OF MATH CURRICULA**

Math skills are critical for success in the workplace, more so today than was the case years ago. Scientific jobs have always required a strong math foundation, and growth rates in science- and technology-related jobs are exceeding job growth in the general labor force (National Science Board 2008). However, service jobs and jobs that once relied on strength and endurance now also require math skills for workers to perform successfully. For example, yesterday's assembly-line workers had to be physically fit and skillful with their hands. Today's assembly-line workers need math skills to effectively operate computerized equipment that automates tasks performed manually in the past.

Federal legislation recognizes the importance of developing math skills starting at an early age. Under Title I of the No Child Left Behind Act, schools must make adequate yearly progress (AYP) in student math performance as well as in reading performance beginning with third grade. AYP is a federally approved, state-specific standard that requires public schools to continuously and substantially improve student achievement in math and reading. The goal is to ensure that all students meet or exceed their state's standard for proficiency in math and reading by 2014.

Many schools face a major challenge meeting AYP in math. The 2007 National Assessment of Educational Progress showed that many U.S. students show mastery of only rudimentary mathematics, and only a small proportion achieve at high levels (Lee et al. 2007). Specifically, only 39 percent of fourth graders were judged “proficient” in mathematics, and 18 percent scored below “basic.” Differences in math performance also exist among fourth graders from different socioeconomic backgrounds (as measured by eligibility for free/reduced-price meals), with the math achievement of those from low income backgrounds lagging behind achievement of those from more affluent backgrounds.

What is taught to students and how it is taught (that is, curriculum and its pedagogical approach) may be important factors in a school’s ability to improve student math achievement, and elementary schools tend to use one of only a few curricula. A national survey conducted in 2008 found that seven math curricula make up 91 percent of the curricula used by kindergarten through second grade educators (Education Market Research 2008). The curricula are based on different theories for developing math skills, and debate exists over which theory is best. The debate about the different approaches is sometimes so intense that it is referred to as the “math war” (Whitehurst 2003; Schoenfeld 2004; Klein 2007).

The curricula and their corresponding instructional approaches are often categorized by terms such as “teacher-directed,” “student-centered,” “explicit,” “inquiry-based,” “traditional,” or “reform.” While all of these terms are used widely and some are used interchangeably, they are not often well defined (National Mathematics Advisory Panel 2008b; Klein 2007; National Research Council 2001). Also, a particular term may be used to categorize a curriculum, but it is possible that the curriculum includes some features or activities that could be categorized by another term. Because of the lack of clear definitions, each term can encompass an array of meanings. For example, teacher-directed approaches range “from highly scripted direct instruction approaches to interactive lecture styles” and student-centered approaches range “from students having primary responsibility for their own mathematics learning to highly structured cooperative groups” (National Mathematics Advisory Panel 2008b).

Despite the widespread use of these different instructional approaches, little research evidence exists about their effectiveness. Slavin and Lake (2007) reviewed studies on the achievement effects of different math curricula. They identified only 13 studies that met their inclusion criteria for review, and only 2 of those used an experimental evaluation design.<sup>8</sup> Other reports also point to the lack of rigorous evidence on the various curricular approaches (National Research Council 2004; What Works Clearinghouse 2006; National Mathematics Advisory Panel 2008b).

The lack of research evidence and the controversy about the different approaches were recognized in discussions held by the Title I Independent Review Panel, the Office of Elementary and Secondary Education, and a panel of curriculum experts. The discussions considered whether impact studies in mathematics should be conducted to provide information on the effectiveness of curricula to teach mathematics. The group ultimately concluded that,

---

<sup>8</sup> A study was included in their review if (1) it used a randomized or matched control group design, (2) treatment duration lasted at least 12 weeks, and (3) the achievement measure was not biased toward the treatment.

although there is little evidence on the effectiveness of specific instructional practices in mathematics, the Title I evaluation plan should include an evaluation of mathematics curricula (IES 2007).

Early in 2005, a panel of experts in mathematics, mathematics instruction, and evaluation design was convened to provide advice on an impact evaluation of mathematics curricula. The panel identified the early elementary grades as the most important level for the evaluation because, even before they enter elementary school, disadvantaged children fall behind their more advantaged peers in basic competencies such as number line ordering and magnitude comparison (Rathburn and West 2004). The panel also recommended that the evaluation compare different approaches to teaching early elementary math through an evaluation of commercially available curricula. It noted that many math curricula have been developed in recent years and are being widely implemented without evidence of effectiveness.

The 2008 report of the National Mathematics Advisory Panel also concluded that few rigorous studies of math curricula have been conducted and more are needed, so future decisions about the approaches used to teach math will be better informed. The panel indicated that a major goal for K-8 mathematics education is to develop student proficiency in content areas (such as whole numbers, fractions, and elements of geometry and measurement) that will help students succeed in Algebra. The panel focused on preparing students for Algebra because successful completion of Algebra is a prerequisite for other higher-level math such as Algebra II, which research shows is correlated with success in college and the labor market (Adelman 1999; Carnevale and Desrochers 2003).

## **B. DESIGNING THE EVALUATION AND SELECTING THE CURRICULA**

The study's goal is to select, implement, and evaluate the relative effects of commercially available early elementary school math curricula that use different instructional approaches. As described below, four curricula were selected for the study, and curriculum implementation and data collection that have been conducted thus far are presented in this report. The analysis in the report helps to answer the following main research questions:

- What are the relative effects of different early elementary math curricula on student math achievement in disadvantaged schools?
- What is the relationship between the effectiveness of the curricula and a school's instructional setting, including teacher knowledge of math content/pedagogy?

The first question examines effects for students overall, and the second examines whether curriculum effects differ for student subgroups in different instructional settings.<sup>9</sup>

---

<sup>9</sup> Additional curriculum implementation and data collection being supported by the study will be presented in a subsequent report, and will help to answer a third main research question: Which math curricula result in a

## 1. Selecting the Curricula

A competitive process was used to select the study's curricula, in which developers and publishers of early elementary school math curricula were invited to submit a proposal to include their curricula in the evaluation. Early in December 2005, the study team issued a request for proposals in an education publication with wide circulation and also sent the announcement to all the major publishers of early elementary school math curricula that could be identified. A total of eight submissions were received.

A panel of outside experts in math and math instruction reviewed the submissions and recommended to IES curricula suitable for the study. Six criteria were used to review the submissions: research support for the curriculum's conceptual framework; empirical evidence of effectiveness; objectives of the curriculum; quality of training and materials; institutional capability to train the number of teachers in the study; and appropriateness for grades one, two, and three in Title I schools. The goal was to identify widely used curricula that draw on different instructional approaches and that hold promise for improving student math achievement.

Late in February 2006, in-person meetings were held with publishers of curricula that were considered strong candidates for the study. The meetings began with publishers providing an overview of their curriculum, including a discussion of the curriculum's key principles, a first-grade lesson on estimation, and how a lesson on estimation in the second grade differs from one in first grade. Publishers were told in advance of the meeting that they should address two questions: (1) what pieces of math knowledge do you think need to be provided to teachers of first, second, and third grade students? and (2) what do you think are the best strategies for teaching students addition facts? The rest of the meeting was spent discussing those questions, as well as any other questions raised by IES, the study team, the panel of reviewers, and the publishers.

Early in March 2006, IES selected the following four curricula for the study:

- ***Investigations in Number, Data, and Space*** (Investigations) is published by Pearson Scott Foresman (Russell et al. 2006) and uses a student-centered approach encouraging metacognitive reasoning and drawing on constructivist learning theory. The lessons focus on understanding, rather than on "correct answers," and build on students' knowledge and understanding. Students are engaged in thematic units of three to eight weeks in which they first investigate, then discuss and reason about problems and strategies. Students frequently create their own representations.
- ***Math Expressions*** is published by the Houghton Mifflin Company (Fuson 2006a) and blends student-centered and teacher-directed approaches to mathematics. Students question and discuss mathematics, but are explicitly taught effective procedures. There is an emphasis on using multiple specified objects, drawings, and

---

*(continued)*

sustained impact on student achievement? This third question examines relative effects when students and teachers experience the study's curricula for more than one year.

language, to represent concepts, and an emphasis 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 2004) and is a scripted curriculum<sup>10</sup> 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. 2005) and is a basal curriculum<sup>11</sup> 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.

Investigations, Saxon, and SFAW are among the seven most widely used curricula in the United States, making up 32 percent of the curricula used by K-2 educators (Education Market Research 2008). Estimating usage of Math Expressions is difficult because it is a newer curriculum, for which market share data are not yet available. Chapter II provides more details about the study's curricula.

## 2. Evaluation Design

Experimental methods are being used to answer the research questions listed above. In particular, the evaluation is based on a school-level random assignment design, in which participating elementary schools in each participating district are randomly assigned to the curricula included in the study. Consider, for example, a district that has eight elementary schools interested in study participation. The study team randomly selects two schools to implement curriculum A, two schools to implement curriculum B, and so on. In each school, first grade teachers receive training and both teacher and student materials free of charge for the curriculum assigned to their school. Relative effects of the curricula are estimated using hierarchical linear modeling (HLM) techniques that compare average math achievement of students in the various curriculum groups.<sup>12</sup> For example, the relative effect of curriculum A

---

<sup>10</sup> Saxon provides teachers with a script to follow throughout each math lesson. The script is intended to help teachers deliver consistent and clear instruction to students (Larson and Saxon Publishers 2006).

<sup>11</sup> Basal curricula use a “hierarchical sequence of academic skills and corresponding instructional materials that are organized by learning objectives” (Erchul and Martens 2002).

<sup>12</sup> See Raudenbush (2002) for a detailed description of the theory and use of HLM.

versus curriculum B is estimated as the difference in average achievement between students in the schools assigned to curriculum A and those in the schools assigned to curriculum B. With the four curricula included in the study, six unique pair-wise comparisons of effects 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.

The study does not include a control group of schools that continue to use the math curriculum they were using before the study began. The study team decided not to include such a control group because it would be difficult to compare effects of the study's curricula to effects for the control group, since the control group would be using a variety of curricula found in the participating districts. Such a control group design could be difficult to interpret even at the district level, because schools in some districts have discretion in choosing their math curriculum. Therefore, the control schools in some districts also could be using a wide variety of curricula. Because students must take math in each of the elementary grades, it also was not possible to include a control group that does not use a math curriculum. The study team instead chose to compare the effects of curricula that represent many of the diverse approaches to teaching mathematics, as described above.

### C. RECRUITING PARTICIPANTS

The findings in this report are based on the four districts and 39 schools that participated in the study during the 2006-2007 school year. In each school, curriculum implementation and data collection were conducted in all first grade classrooms. Below, we summarize how the 2006-2007 school year participants were recruited.

#### 1. Suitable Districts

The study team's goal was to recruit districts that met the following criteria:

- ***Have Title I Schools.*** Including districts that have Title I schools is consistent with the policy interest that underlies Title I for studying effective approaches to help low-income children meet state standards for academic achievement. Participation was not limited to Title I schools, but an emphasis was placed on including Title I Schools in the study.
- ***Are Geographically Dispersed.*** Although (as described below) districts and schools were purposively selected, geographic diversity can help ensure that any variation in effects that could result from regional differences in instructional contexts is included.
- ***Contain at Least Four Schools Interested in Study Participation.*** Requiring that each district contain at least four elementary schools supports implementation of all

four curricula in each district, and makes it possible to examine whether curriculum effects vary across sites.<sup>13</sup>

Among districts that met the criteria above, those that were actually interested in study participation may be unique in other ways. For example, interested districts had to be willing to implement four very different curricula and each participating school had to be willing to use the curriculum randomly assigned by the study team. Sites that were comfortable with these participation requirements may value research evidence and be interested in obtaining direct evidence for their district to inform a future curriculum adoption decision. These participation requirements also may be acceptable to districts with tight budgets, because the free curriculum training and materials provided by the study could free up funds that districts could use in other ways. Of course, districts may have participated for other reasons. For example, an influential district leader who believed the study would be a valuable experience may have promoted the study to all the “right” people in the district—people who could be difficult for outsiders (such as members of the study team) to identify and contact.

The study team also sought schools with teachers who had not previously used the study’s curricula, so that no one curriculum had a potential advantage over another. However, the study team needed to commit to districts and schools that were interested in participation before an assessment of teacher prior use of the curricula could be made. The decision of which curriculum a district or school will use during a particular school year is typically made before the end of the previous school year. As such, recruiting districts and schools that would begin study participation during the 2006-2007 school year needed to be completed before the end of the previous (2005-2006) school year. At that time, schools were not aware of all the teacher turnover that might occur before the start of the next school year, which made it impossible to confirm that no teachers had prior experience with the study’s curricula. Nevertheless, as described in Chapter II, the study team was fairly successful at identifying new users of the curricula. Nine percent of study teachers had used their assigned curriculum at the K-3 level at some point in the past, and the effects of the curricula reported in Chapter III have been adjusted for this prior use.

## **2. Recruiting Districts and Schools**

Recruiting districts and schools typically involved three main activities. The first activity included identifying sites that met the criteria above followed by initial outreach to assess district interest. Various sources were used to identify sites that met the criteria above, including national district data sets, the hundreds of districts MPR has worked with on previous studies, publisher nominations of districts that had expressed interest in using their curricula, and announcements about the study in national publications.

---

<sup>13</sup> Although only four schools are needed in a district to support implementation of the study’s four curricula, the goal was to recruit districts with at least eight elementary schools, so at least two schools could be assigned to each curriculum in each district. Having at least two schools per curriculum in each district helps maintain each curriculum’s presence in each district if some schools stop using their assigned curriculum, and helps reduce the potential confounding of school and curriculum effects when examining district-level results.

National district data sets—including both the Common Core of Data (CCD) and data from [www.SchoolMatters.com](http://www.SchoolMatters.com)—were used to rank districts by their schools’ free/reduced-priced meals eligibility.<sup>14</sup> The ranking was done from highest to lowest and only included districts with at least four elementary schools. Data from [www.SchoolMatters.com](http://www.SchoolMatters.com) were then used to examine math achievement of the districts on the list, to further winnow it down to those with math proficiency scores that were below the state average. The goal was to include schools with a range of low math proficiency (for example, those just below the state average and those significantly lower than the state average), so the study team could examine how the relative effects of the curricula are related to the extent to which students are struggling in math.

Two letters were then sent to each potential district: one to the district superintendent and another to the curriculum director. The letters briefly described the study and the benefits of participating. The study team followed up the letters with phone calls to assess each district’s interest.

The second recruiting activity involved site visits to interested districts that did not object to three critical elements of the study: (1) piloting all four of the study’s curricula, (2) random assignment of curricula to participating schools, and (3) the study’s data collection plan. Recruiters met with district administrators. If the administrators considered it appropriate at this stage of the recruiting process, the initial meeting also included principals and teachers from elementary schools that might be interested in study participation. In some districts where a small number of individuals were part of the initial meeting, recruiters sometimes were asked to make additional site visits to describe the study to other district or school staff. Sometimes, several follow-up visits were required, so recruiters could describe the study to all individuals who would be involved if the district participated.

During the visits, questions about the study’s curricula often arose. Because recruiters were not experts on the curricula, they answered only basic curriculum questions and relayed detailed questions to the appropriate publisher after the visit. If there was advance notice that detailed curriculum questions would arise, publisher representatives attended the meeting so questions could be answered immediately.

The third and final recruitment activity was to enroll schools, teachers, and any other relevant staff in districts that were interested in study participation. Enrollment began by confirming that schools interested in participation clearly understood the study’s parameters. Most importantly, recruiters confirmed that schools were willing to use any of the study’s four curricula and would support the study’s data collection.

A school was considered a participant when the study team received consent forms for all first grade teachers in the school. Recruiters provided consent forms to principals to distribute to teachers in interested schools. Signing the consent form meant that a teacher agreed to attend training on whatever curriculum was assigned to the school, implement the curriculum to the

---

<sup>14</sup> Both of these sources were used because, collectively, they contain several pieces of information that were useful for identifying sites.

best of his or her ability, and cooperate with student testing conducted by the study team. The study team also asked teachers to agree to several other data collection efforts, including teacher surveys and classroom observations. Although the other data collection efforts were not a requirement for study participation, response rates to these efforts were high (see Appendix A).

All 131 first grade teachers in the 39 cohort-one schools agreed to participate in the study.<sup>15</sup> Other staff that schools or publishers indicated were important for successful curriculum implementation also were encouraged to attend training. These other staff typically included math coordinators, math coaches, and supplemental teachers.

### **3. Characteristics of Participants**

Tables I.1 and I.2 present information that is useful for understanding the types of districts and schools that began study participation during the 2006-2007 school year. Table I.1 presents several key characteristics of all U.S. districts and those that agreed to participate. Table I.2 presents similar information for U.S. elementary schools and those that agreed to participate.

As the tables show, the characteristics of districts and schools that agreed to participate are consistent with the study team's recruitment goals. The study team contacted 118 districts from March to June 2006, and 4 of them agreed to participate beginning in the 2006-2007 school year—a district recruitment rate of 3.4 percent. The four districts are geographically dispersed in four states (Connecticut, Minnesota, New York, and Nevada), in three regions of the country (Northeast, Midwest, and West). The districts also are in areas with different levels of urbanicity—two districts are in urban areas, the third is in a suburban area, and the fourth is in a rural area. When compared to the average U.S. district, those that agreed to participate have a higher fraction of school-wide Title I eligible schools, students eligible for free/reduced-price meals, and minority students. A similar pattern exists when comparing U.S. elementary schools with those that agreed to participate.

### **D. RANDOM ASSIGNMENT AND STATISTICAL POWER**

Random assignment of curricula to schools was conducted separately for each participating district, and only after all teacher consent forms for all participating schools in a district were received. Obtaining teacher consent before random assignment helps to identify schools that are willing to participate in the study, regardless of the curriculum assigned to each school.

The study team used a “blocked” random assignment procedure that allocates similar numbers and types of schools, teachers, and students to each curriculum. The procedure divides schools in each district into blocks, where each block contains from four to seven schools with similar baseline characteristics. Random assignment of curricula to schools is then conducted within each block. This procedure helps to minimize chance differences in school characteristics and sample sizes across curriculum groups, which helps to increase the face validity and

---

<sup>15</sup> Only two first grade teachers in two separate schools were not included in the study, because those teachers worked with classrooms of high-needs students who were not eligible for testing.

TABLE I.1  
CHARACTERISTICS OF U.S. DISTRICTS AND COHORT-ONE  
PARTICIPATING DISTRICTS

	U.S. Districts	Cohort-One Participating Districts
Number of Schools	6	60
Title I Eligible Schools (percentage) <sup>a</sup>	60.1	56.3
School-Wide Title I Eligible Schools (percentage) <sup>a</sup>	23.4	32.3
Student Enrollment	3,045	22,102
Students Eligible for Free/Reduced-Price Meals (percentage)	38.0	55.3
Student Gender (percentage)		
Male	52.1	51.2
Female	47.9	48.8
Student Race/Ethnicity (percentage)		
White	74.8	48.9
Black	10.0	24.8
Hispanic	10.2	18.7
Asian	1.8	4.7
American Indian/Alaskan Native	3.2	2.9
<b>Sample Size</b>	<b>16,653</b>	<b>4</b>

Source: Author calculations using the 2003-2004 Common Core of Data (CCD). When free/reduced-price meals data were missing in the CCD, data were obtained from [www.GreatSchools.net](http://www.GreatSchools.net).

Note: Data include districts with at least one school with at least one student.

<sup>a</sup>The Title I program provides financial assistance to schools with high numbers/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 school-wide programs that serve all children in the school. Title I eligible schools have at least 35 percent of students from low income families.

TABLE I.2  
CHARACTERISTICS OF U.S. ELEMENTARY SCHOOLS AND COHORT-ONE  
PARTICIPATING SCHOOLS

	U.S. Elementary Schools	Cohort-One Participating Schools
Title I Eligible (percentage) <sup>a</sup>	71.9	74.4
School-Wide Title I Eligible (percentage) <sup>a</sup>	41.0	53.8
Student Enrollment (average)		
First Grade	70	73
Second Grade	68	71
Students Eligible for Free/Reduced-Price Meals (percentage)	47.4	68.7
Student Gender (percentage)		
Male	51.9	51.5
Female	48.1	48.5
Student Race/Ethnicity (percentage)		
White	59.3	46.2
Black	16.4	26.1
Hispanic	18.3	19.9
Asian	3.8	3.8
American Indian/Alaskan Native	2.2	4.0
<b>Sample Size</b>	<b>53,265</b>	<b>39</b>

Source: Author calculations using the 2003-2004 Common Core of Data (CCD). When free/reduced-price meals data were missing in the CCD, data were obtained from [www.GreatSchools.net](http://www.GreatSchools.net). The sample excludes one Math Expressions school (with 3 classrooms and 32 students) that participated during part of the school year and then stopped using its assigned curriculum and did not allow the study to collect follow-up data.

Note: Data include elementary schools with at least one first or at least one second grade student.

<sup>a</sup>The Title I program provides financial assistance to schools with high numbers/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 school-wide programs that serve all children in the school. Title I eligible schools have at least 35 percent of students from low income families.

statistical power of the evaluation design. Agodini, Deke, Atkins-Burnett, Harris, and Murphy (2008) provides more details about the blocked random assignment procedure used by the study. The way in which the procedure was implemented with the current sample is described in Appendix A. Chapter III shows that the four curriculum groups are comparable along important baseline characteristics.

The study's main results are based on students who were tested in both the fall and spring, and fall and spring class rosters were collected to identify students who should be tested at both points in time. The fall rosters were used to identify the students to whom parent consent forms should be distributed, and to select the student sample. The 39 schools included in the analysis contain a total of 131 first grade classrooms, as mentioned above. To detect the study's target effect size, a sample of 1,525 students—an average of about 11.5 students in each of the first grade classrooms—was randomly selected in the fall for study participation. Fall tests were administered to 1,457 students (96 percent) of the student sample. Parent refusals accounted for two-thirds of student nonresponse. In the spring, the goal was to test all sampled students still in a study school—the study did not track students who were not in a study school in the spring. Of the 1,525 students sampled in the fall, the study team was able to test 1,330 (87 percent) in the spring. Attrition (that is, transfers outside of a study school) accounted for most nonresponse. Of the 1,330 students tested in the spring, 1,309 also were tested in the fall (about 10 students per classroom) and the analysis is based on this sample.<sup>16</sup> This analysis sample represents 86 percent of students sampled in the fall for study participation. See Appendix A for more details about the sampling procedure and testing response.

Table I.3 presents the number of schools, classrooms, and students included in the analysis, in total and by curriculum group. Each of the four curricula was randomly assigned about 10 schools and a total of 33 classrooms and 325 students.

The effect size that can be detected with the study's current sample is as small as 0.22, where effect size is defined as a fraction of the standard deviation of the test score.<sup>17</sup> The minimum effect size that can be detected depends on sample size, how the sample is distributed across the curriculum groups, and the extent to which students are clustered in schools and classrooms according to their baseline achievement, after adjusting for other baseline student, teacher, and school characteristics included in the HLM analysis. As described above, the study's random assignment procedure allocated a similar sample size to each of the four curriculum groups—an equal allocation provides the greatest statistical power. In the current sample, the school- and classroom-level intraclass correlation coefficients (ICC) equal 0.00 and 0.08,

---

<sup>16</sup> Given the number of schools and classrooms included in the study, the statistical power benefits of pre- and post-testing more than 10 students per classroom are minimal, though the costs are significant because the study used an individually administered assessment, as described below.

<sup>17</sup> The effect size equals the difference between average student math scores of any two curriculum groups, divided by the pooled standard deviation of the score for the two curricula being compared.

TABLE I.3  
NUMBER OF COHORT-ONE SCHOOLS, CLASSROOMS,  
AND STUDENTS, BY CURRICULUM

	Curriculum				
	All	Investigations	Math Expressions	Saxon	SFAW
Schools	39	10	9	9	11
Classrooms	131	33	31	31	36
Average # of classrooms/school	3.4	3.3	3.4	3.4	3.3
Students Both Fall and Spring Tested	1,309	332	314	304	359
Average # of students/classroom	10	10	10	10	10

Note: Author calculations. The sample excludes one Math Expressions school (with 3 classrooms and 32 students) that participated during part of the school year and then stopped using the curriculum and did not allow the study to collect follow-up data.

respectively, after adjusting for student, teacher, and school characteristics.<sup>18</sup> The calculation is based on a three-level clustered design and accounts for the six unique pair-wise comparisons of effects that can be made with the study’s four curricula, as described above.

This minimum detectable effect represents about 15 percent of the one-year math achievement gain made by the average first grader from a low socioeconomic background—the type of students that largely are part of this evaluation.<sup>19</sup> Put differently, when comparing two curriculum groups, student achievement must differ by at least 15 percent of the gain made by the average first grader from a low income family to be able to detect those differences with the four districts and 39 schools that are examined in this report.

---

<sup>18</sup> There is clustering at the school level because, if random assignment were repeated, a different set of classrooms would be assigned to the study’s curricula. There also is clustering at the classroom level because a sample of students in each classroom was tested, so a different set of students would be tested if the sampling were repeated.

<sup>19</sup> This statistic is based on data from the national Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K) (Rathburn and West 2004). On average, children in the ECLS-K who were in the bottom quintile of socioeconomic status (a composite measure based on an equal weighting of children’s parents’ education, occupation, and household income) gained about 16 scale points in math during the first grade. The standard deviation for these children’s fall scores was 10.9. Therefore, an effect size of 0.22 equals 2.18 scale points ( $0.22 \times 10.9 = 2.40$ ) during first grade, which, in turn, equals 15 percent of the average math gains made by the average first grader [ $(2.40/16) \times 100 = 15\%$ ].

## **E. DATA COLLECTION**

Figure I.1 illustrates the timing of the data collection efforts during the 2006-2007 school year. Table I.4 lists the study's research questions and the data collection efforts used to gather information that supports answers to each question. The research question about the sustained effects of the curricula is not included in the table because this issue will be examined in a follow-up report.

### **1. Outcome Measure**

To measure the relative effects of the curricula, the study team assessed student math achievement using the assessment developed for the National Center for Education Statistics' Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K). The goal was to use an assessment that had already been developed, and that assesses the knowledge and skills mathematicians and math educators feel are important for early elementary school students to develop. The ECLS-K assessment meets these study requirements, as well as accepted standards of validity and reliability. The assessment also meets other important requirements, including individual administration, being nationally normed, ability to measure achievement gains over the study's grade range (which ultimately will include the first, second, and third grades), and accuracy in capturing achievement of students from a wide range of backgrounds and ability levels.

Another important feature of the ECLS-K assessment is that it is an adaptive test, which is an approach used to measure achievement that is tailored to a student's achievement level. In particular, the test begins by administering to each student a short, first-stage routing test used to broadly measure each examinee's achievement level. Depending on the score on the routing test, the student is then assigned to one of three longer second-stage tests: (1) an easy test, (2) a middle-difficulty test, or (3) a difficult test. Some of the items on the second-stage tests overlap, and this overlap is used by item response theory (IRT) techniques (Lord 1980) to place scores on the different tests on the same scale. IRT estimates the number of items students would have answered correctly if they had taken all of the questions on all three of the second-stage tests. The analysis is based on these scale scores, which, according to the test developers, are the correct scores to analyze for our purposes (Rock and Pollack 2002). Adaptive tests are useful for measuring achievement because they limit the amount of time children are away from their classrooms and reduce the risk of ceiling or floor effects in the test score distribution—something that can have adverse effects on measuring achievement gains.

FIGURE I.1

DATA COLLECTION TIMELINE DURING THE 2006-2007 SCHOOL YEAR

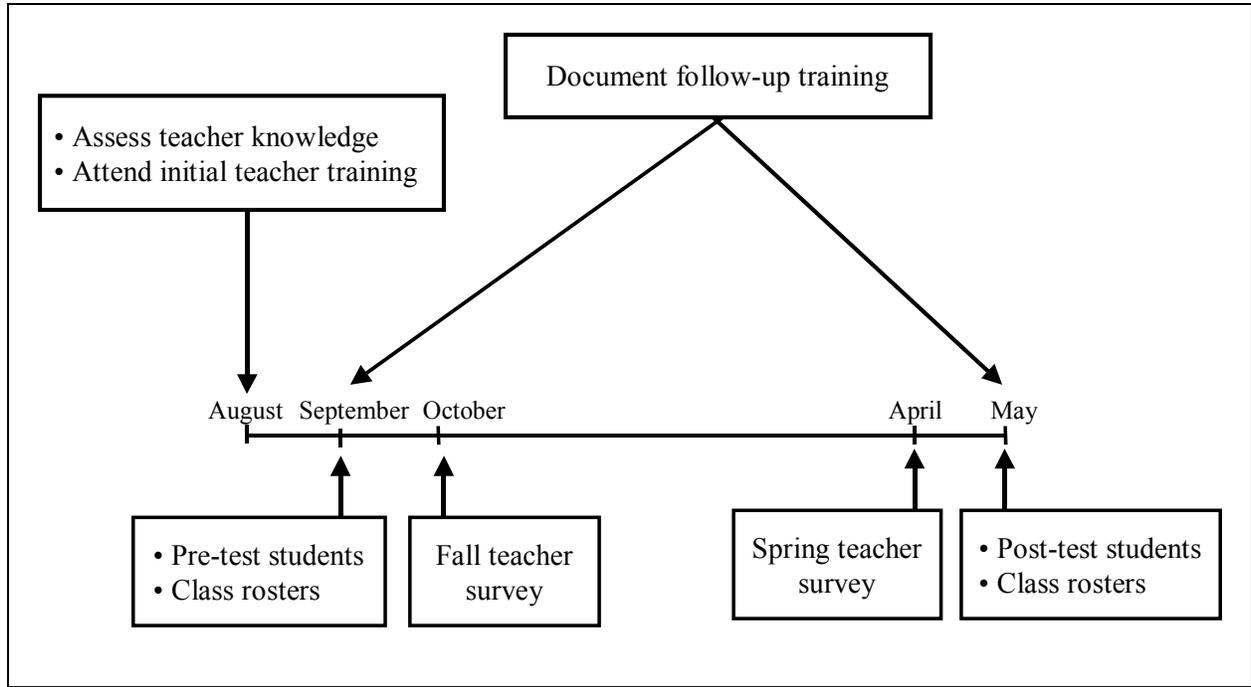


TABLE I.4

RESEARCH QUESTIONS AND SUPPORTING DATA COLLECTION EFFORTS:  
2006-2007 STUDY PARTICIPANTS

Research Question	Supporting Data Collection Effort
1. What are the relative effects of different early elementary math curricula on student math achievement in disadvantaged schools?	► Fall and spring math tests of first grade students. Student roster data and teacher characteristics from the survey are used in the analysis.
2. Under what conditions is each math curriculum most effective?	► Conditions are defined using school- and teacher-level characteristics, such as school fall achievement and teacher education.
3. What is the relationship between teacher knowledge of math content/pedagogy and the effectiveness of the curricula?	► Teacher scores on the study-administered assessment of math content and pedagogical knowledge.

The assessment includes questions in the five math content areas used in the *Mathematics Framework for the 1996 National Assessment of Educational Progress* (National Assessment Governing Board 1996):

1. Number Sense, Properties, and Operations
2. Measurement
3. Geometry and Spatial Sense
4. Data Analysis, Statistics, and Probability
5. Patterns, Algebra, and Functions

The items in each of the second-stage tests administered to the first graders can primarily be classified as Number Sense, Properties, and Operations, with the remainder from the other areas. The easy test contained only a few items from each of the remaining areas, whereas the middle-difficulty and difficult tests contained more such items. On the middle-difficulty test, the remaining items were mainly about Patterns, Algebra, and Functions, whereas those on the difficult test were mainly about Data Analysis, Statistics, and Probability.<sup>20</sup>

The study team administered the student assessment. The baseline (fall) test was administered as close to the beginning of the school year as possible, and the follow-up (spring) test as close to the end of the school year as possible. Testers pulled students from their classrooms one at a time and took them to a quiet place (such as the school library) to administer the assessment. The total time required for pulling a student from the classroom, testing, and bringing the student back was about 45 minutes.

Student answers on the assessment were sent to the Educational Testing Service for scoring.<sup>21</sup> A three-parameter IRT model was used to place scores from the different tests students took on the same scale. Reliabilities for the study's sample equal 0.93 for the fall score and 0.94 for the spring score, and are consistent with the national ECLS-K sample (Rock and Pollack 2002, pp. 5-7 through 5-9).<sup>22</sup> Also, there were no floor or ceiling effects observed in either the fall or spring scores.

## **2. Other Data Collection**

To help interpret measured effects, the following other data collection efforts were conducted by the study team:

---

<sup>20</sup> See Rock and Pollack (2002) for more information about the process used to develop the ECLS-K assessment.

<sup>21</sup> Educational Testing Service was a developer of the ECLS-K Mathematics Assessment.

<sup>22</sup> Reliabilities are based on the internal consistency (alpha) coefficients.

- ***Assessment of Teacher Knowledge of Math Content and Pedagogy.*** Teacher math content/pedagogical knowledge was assessed at the initial teacher training sessions before the curricula were introduced, using an assessment developed by researchers at the University of Michigan.<sup>23</sup> Scores on the test are included in the analysis of student achievement to examine the relationship between teacher math content/pedagogical knowledge and the effects of the curricula.
- ***Curriculum Training Received by Teachers.*** The study team took attendance at the initial teacher trainings the publishers conducted before the start of the school year. Attendance at the follow-up trainings that occurred during the school year was recorded and provided by the publishers and was also collected from teachers through the surveys described below.
- ***Teacher Surveys.*** Two surveys were administered to teachers. The first survey was administered in the fall and focused on teacher background information, classroom characteristics, curriculum training provided by the publishers up to that point, and math instruction approaches used before joining the study. The second survey was administered in the spring and gathered information on follow-up training provided by the publishers, usage of the assigned curriculum and any other math curricula, and math instructional practice used during the year. Information on the spring survey was used to assess teacher adherence to the study's curricula.
- ***Student Characteristics from Class Rosters.*** The study team collected rosters for each classroom in the study to select the student sample. Student demographic information was requested as part of the roster collection, so the demographics could be included in the analysis to help increase the study's statistical power. The request included student gender, date of birth, race/ethnicity, free/reduced-price meals eligibility, whether the student had limited English proficiency or was an English language learner, and whether the student had an individualized education plan or received special services for students with a disability.

Appendix A reports response rates to the data collection efforts. The data collection forms are contained in Agodini et al. (2008), with the exception of the student math assessment and teacher knowledge assessment because those instruments are copyrighted.<sup>24</sup>

---

<sup>23</sup> The teacher assessment includes items about teacher pedagogical content knowledge in two major domains: (1) knowledge of mathematics for teaching and (2) knowledge of students and mathematics. Items focus on numbers, operations, and patterns, functions and algebra—the three mathematics content areas most frequently covered in the elementary grades. Mathematicians, math educators, professional developers, former teachers and the authors themselves (who had experience teaching and observing elementary classrooms) wrote items. Hill, Schilling, and Ball (2004) provides details about the assessment's development process. The reliability of the teacher test score for the study's sample equals 0.81.

<sup>24</sup> The study team is also conducting classroom observations to assess curriculum implementation, and each classroom in the current sample was observed once during the 2006-2007 school year. The observation data are not presented in this report because the reliability of those data cannot be assessed until observations have been

## F. FUTURE PUBLICATION PLANS

Another 71 schools joined the study during the 2007-2008 school year (the year after the 39 schools examined in this report joined), and curriculum implementation occurred in both the first and second grades in all participating schools. A follow-up report is planned that will present results based on all 110 schools participating in the evaluation, and for both the first and second grades.<sup>25</sup> The study also is supporting curriculum implementation and data collection during the 2008-2009 school year in a subset of schools, in which implementation will be expanded to the third grade. A third report is planned that will present those results.

---

*(continued)*

completed in all the study schools. The plan is to present the observation data in future reports described in the next section.

<sup>25</sup> The study's full sample size (of 12 districts and 110 schools) is consistent with the study's target of 12 districts and 108 schools, which was selected so an effect size as small as 0.20 could be detected (see Agodini et al. 2008). This effect size calculation for the study's full sample was conducted before any data for the evaluation were collected. As described above, the minimum effect size that can be detected with the 4 districts and 39 schools examined in this report is 0.22 and close to the value for the full sample. When calculating statistical power during the design phase for the full sample, assumptions had to be made for some parameters in the calculation and those assumptions turned out to be conservative for the current sample. In particular, the extent of school- and classroom-level clustering and the explanatory power ( $R^2$ ) of the statistical model (HLM) used to calculate relative curriculum effects were based on estimates from previous studies, but are conservative estimates for the study's current sample.

## II. CURRICULUM IMPLEMENTATION

The study's goal was to evaluate the effects of the curricula based on the type of implementation that occurs in a typical district that purchases the materials. Results based on this level of implementation indicate the effects of the curricula from typical use, and would be more informative to districts that are considering which curriculum to purchase than results based on some level of implementation that only the study could achieve.

To meet this goal, it was important to consider how a district adopts a new curriculum to ensure that implementation of the study's curricula occurs as it would outside of the study's context. When a district adopts a new curriculum to implement, many activities need to occur for the implementation to be successful. For example, adoptions can include discussions among many staff—ranging from district administrators to teachers—and the curriculum ultimately selected may depend on buy-in from a majority of these individuals. In addition, the district orders curriculum materials far in advance of the start of the school year, makes decisions about how to allocate teacher time during in-service days to provide opportunities for curriculum training, and establishes supports within the district that can help resolve issues surrounding implementation—such as ensuring that curriculum coordinators are knowledgeable about the curriculum being adopted.

To ensure that all the activities districts typically undertake during a curriculum adoption occurred in the context of the study, the study team provided some basic implementation support. The support began during the site recruitment process. Before random assignment began, the study team sought buy-in for all four of the study's curricula from all key district- and school-level staff. After the study team conducted random assignment, the team introduced the participating districts to the publishers. Publishers then worked with the districts and schools to deliver curriculum materials when study participants needed them. Publishers also worked with schools and teachers to establish training days, and the study team provided logistical and financial support for the trainings. When teachers received training during noncontract time (during summer, evenings after school, or weekends) they were compensated for their time at district salary rates as required by teacher unions.<sup>26</sup>

Although the study team provided some basic supports, they were not responsible for implementation; instead, implementation ultimately reflects what publishers and districts achieved. For example, some schools notified the study team about a variety of implementation issues, ranging from long-term substitutes in study classrooms who needed curriculum training to teachers encountering challenges using their assigned curriculum. Addressing issues such as these was not the responsibility of the study team, so they immediately brought these issues to the attention of the publishers who were responsible for following up with the schools.

---

<sup>26</sup> The study team sought to support implementation in ways that are consistent with typical district and publisher practices. However, it is unclear if the support provided by the study differed from typical support, or if the study's support affects the generalizability of the study's results.

To characterize implementation, the study team administered two surveys to teachers, and this chapter summarizes the data from those surveys. The first (fall) survey was administered during October and November 2006. The survey collected information about teacher characteristics (such as their education and experience) and some preliminary information about implementation to date, such as participation in training provided by publishers and whether teachers were using their assigned curriculum. The second (spring) survey was administered during April and May 2007. The survey asked teachers about training on their assigned curriculum, use of the curriculum, and math instruction in their classroom throughout the school year. The number of teachers who responded to the two surveys differed by 3 due to teacher turnover during the school year.

For all of the survey data reported, statistical tests were conducted to determine whether there were any differences across the curriculum groups. Two types of tests were performed using hierarchical linear modeling (HLM) techniques, depending on the data measure. For baseline characteristics that do not change over time (such as gender and race), statistical tests examined whether characteristics differed across the curricula.<sup>27</sup> For characteristics that could change over time (such as time spent on math instruction and content covered), statistical tests controlled for classroom and school characteristics when examining whether there are differences across the curricula.<sup>28</sup> The statistical tests examine the joint equality of each item across the curriculum groups, and only those items that are significantly different at the 5 percent level of confidence are discussed below.<sup>29</sup>

## **A. CONTEXT FOR CURRICULUM IMPLEMENTATION**

In Chapter I, we saw that the four study districts examined in this report are located in four different states—two districts are located in urban areas, the third in a suburban area, and the fourth in a rural area. In addition, participating schools, on average, have higher poverty levels than the nation’s schools (Table I.2).

---

<sup>27</sup> These statistical tests were conducted using two-level HLMs. The first (teacher-level) equation regressed each teacher measure on an intercept and a teacher-level error term. The second (school-level) equation regressed the intercept from the first equation on a school-level intercept, binary indicators for three of the four curricula, binary indicators for all but one of the blocks to which the schools were assigned during random assignment, and a school-level error term. By including indicators for the blocks, the degrees of freedom used to calculate the statistical significance of the results are adjusted to reflect the information (number of blocks constructed) used when conducting random assignment. HLMs that are appropriate for continuous, binary, and categorical variables were used accordingly.

<sup>28</sup> These statistical tests were conducted using two-level HLMs. The first (teacher-level) equation regressed each teacher measure on teacher race, education, experience, score on the content/pedagogical test, class size, prior use of the assigned curriculum at the K-3 level, average class fall math achievement, variance of the class fall score, and skewness of the class fall score. The second (school-level) equation regressed the intercept from the first equation on school free/reduced-price meals participation, Title I status, binary indicators for three of the four curricula, and binary indicators for all but one of the blocks to which the schools were assigned during random assignment. HLMs that are appropriate for continuous, binary, and categorical variables were used accordingly.

<sup>29</sup> The 5 percent level of confidence means there is no more than a 5 percent chance that any finding discussed could have occurred by chance.

These district and school characteristics are important for setting the context for curriculum implementation, as are the characteristics of teachers who were assigned to use the curricula. In terms of basic demographics, 92 percent of the teachers are white, 92 percent are female, and the average age of the teachers is 39 years old (Table II.1).<sup>30</sup> Along other important dimensions, we see that:

- ***Teacher Experience and Education.*** Teachers have an average of 13 years of teaching experience.<sup>31</sup> All teachers have a bachelor's degree, and 68 percent also have a master's degree or higher.<sup>32</sup> Sixty-nine percent of the bachelor's degrees are in an education field (elementary education, early childhood education, or K-12 education); the rest are in a variety of subject areas, none of which are mathematics. Looking across both bachelor's and master's degrees, 94 percent of teachers reported education as a major field of study for any of the degrees they earned. While no teachers have a degree in mathematics, 60 percent took at least one advanced math course; 97 percent took at least one advanced course in math education.
- ***Prior Professional Development.*** During the 12 months before the start of the school year, 32 percent of teachers participated in non-study math professional development. Types of professional development included math instruction, math content, and performance standards in math education.
- ***Teacher Knowledge.*** At each initial curriculum training session (described below), the study team administered an assessment of math content and pedagogical knowledge to teachers as the first activity of the day.<sup>33</sup> The test covers kindergarten through fifth grade knowledge. On the pedagogical knowledge subscale, study teachers on average correctly answered questions that involved identifying how students might use base 10 blocks to represent a 2-digit number and common errors students make when estimating. Teachers with above average scores were also able to identify student errors in working with fractions. On the content knowledge subscale, teachers on average correctly identified the number halfway between two decimals, and teachers with above average scores also could identify the product of two fractions represented on a number line.

---

<sup>30</sup> The standard deviation for teacher age is 11.2 years.

<sup>31</sup> The standard deviation for teacher experience is 10.5 years.

<sup>32</sup> Six percent of teachers held a degree higher than a master's degree. These higher degrees were advanced certificates in a subject area or Ph.Ds.

<sup>33</sup> The teacher assessment is included in the analysis of student achievement. As mentioned in Chapter I, the reliability of the teacher test score for the study's sample equals 0.81, and the reliability of the two (pedagogical and content knowledge) subscales equal 0.73 and 0.80, respectively.

TABLE II.1  
TEACHER BASELINE CHARACTERISTICS, BY CURRICULUM  
(Percentage Unless Stated Otherwise)

	Teachers by Curriculum					<i>p</i> -value
	All Teachers	Investigations	Math Expressions	Saxon	SFAW	
<b>Demographics</b>						
Average Age	39.4	41.7	39.2	39.0	37.9	0.42
Gender						
Male	7.9	--	--	--	--	0.44
Female	92.1	--	--	--	--	
Race*						
White	92.3	--	--	--	--	0.00
Other	7.7	--	--	--	--	
<b>Experience</b>						
Average Years of Teaching Experience	13.2	14.2	13.2	13.1	12.3	0.93
Type of Teaching Certificate Held						
Regular or standard	85.7	--	--	--	--	0.52
Other	14.3	--	--	--	--	
Content Area of Teaching Certificate						
Elementary education	92.0	--	--	--	--	0.58
Early childhood or K-12 education	8.0	--	--	--	--	
Grade Level for Teaching Certificate						
Elementary grades	86.2	--	--	--	--	0.30
Elementary and secondary grades	13.8	--	--	--	--	
<b>Education</b>						
Highest Degree Earned						
Bachelor's degree	32.5	28.1	35.7	27.6	38.2	0.21
Master's degree or higher	67.5	71.9	64.3	72.4	61.8	
Field for Bachelor's Degree						
Elementary education	57.8	68.7	55.6	62.1	45.5	0.43
Early childhood or K-12 education	10.8	0.0	11.1	10.3	21.2	
Mathematics	0.0	0.0	0.0	0.0	0.0	
Other	31.4	31.3	33.3	27.6	33.3	
Have a Second Major Field of Study	39.3	51.6	50.0	33.3	23.5	0.07
Second Field of Study (among those with a second field)						
Early childhood, elementary, or special education	40.8	50.1	46.2	14.3	37.5	0.38
Mathematics	0.0	0.0	0.0	0.0	0.0	
Other	59.2	49.9	53.8	85.7	62.5	

TABLE II.1 (continued)

	All Teachers	Teachers by Curriculum				<i>p</i> -value
		Investigations	Math Expressions	Saxon	SFAW	
Number of Advanced Math Courses Taken						
None	40.2	48.3	29.6	27.6	53.1	0.26
1 or 2	45.3	34.5	55.6	58.6	34.4	
3 or more	14.5	17.2	14.8	13.8	12.5	
Number of Advanced Math Education Courses Taken						
None	2.6	--	--	--	--	0.40
1 or 2	62.4	--	--	--	--	
3 or more	35.0	--	--	--	--	
<b>Math Content/Pedagogical Knowledge</b>						
Teacher Assessment (Scale Score)						
Total	-0.08	0.06	-0.06	-0.06	-0.24	0.48
Content knowledge	-0.62	-0.50	-0.62	-0.59	-0.77	0.44
Pedagogical knowledge	-0.31	-0.24	-0.30	-0.30	-0.40	0.80
<b>Professional Development in the 12 Months Prior to the 2006-2007 School Year</b>						
Participated in the Following Types of Professional Development (PD)						
Math instruction	22.0	25.0	21.4	20.7	20.6	0.91
Math content	18.9	18.8	25.0	14.3	17.6	0.46
Performance standards in math education	17.4	21.9	25.0	14.3	9.1	0.37
Other math-focused PD	18.9	21.2	25.0	17.9	12.1	0.66
Participated in Any of the Above Activities	32.3	42.4	32.1	31.0	23.5	0.27
<b>Sample Size</b>	<b>127</b>	<b>33</b>	<b>29</b>	<b>29</b>	<b>36</b>	

Source: Author calculations using data from the fall 2006 teacher survey, and the study-administered assessment of teacher math content and pedagogical knowledge. The sample excludes one Math Expressions school (with 3 classrooms and 32 students) that participated during part of the school year and then stopped using the curriculum and did not allow the study to collect follow-up data.

\*Statistically significant at the 5 percent level. The statistical tests were conducted using two-level HLMs (see the text at the beginning of Chapter II for details). A single *p*-value is reported for binary and multinomial variables, and indicates whether the fraction of teachers in each category of the variable differs across the curriculum groups.

-- Value suppressed to protect respondent confidentiality.

- ***Prior Use of the Assigned Curriculum.*** Nine percent of study teachers used their assigned curriculum in a primary grade (K-3) at some point prior to the study (Table II.2). Among teachers who taught in a primary grade in the prior school year, teachers most commonly reported using Silver Burdett Ginn Mathematics,<sup>34</sup> Saxon Math, and Everyday Math and as their core math curriculum. Teachers had five years of experience with these prior curricula, on average.

## **B. TEACHER CURRICULUM TRAINING**

A key component of curriculum implementation involved training teachers to use their assigned curriculum. The publishers provided both initial trainings that occurred before the start of the school year and follow-up training and support during the year. The publishers established a plan for follow-up training and support in their proposals to the study during the curriculum selection process and, in some cases, modified those plans after the initial training sessions in response to study teacher needs. In some cases, districts or schools asked publishers for additional training. The study team provided logistical and financial support for any level of training the publishers indicated was appropriate.

### **1. All Teachers Attended Initial Training on Their Assigned Curriculum**

Teachers received initial training on their assigned curriculum in summer 2006. The trainings were group sessions held at a location within each participating district, and separate trainings were held for each curriculum. Training typically occurred two to four weeks before the first day of school.<sup>35</sup>

Two sources of data were used to document attendance at the initial trainings. The first source was attendance forms collected by study team members, who attended each initial training and took attendance. The attendance forms documented each attendee's name, school affiliation, position, and arrival and departure time. The second source of data was the fall teacher survey, which asked teachers about their attendance at the initial training session. The survey provided an opportunity to document attendance of any teachers who may have filled an open position at a study school after initial training occurred and who may have attended a make-up training session.

The two sources of data on initial training are consistent and show that all study teachers attended initial training on their assigned curriculum (Table II.3). The publishers of Math Expressions provided two days of initial training, whereas the publishers of Investigations in Number, Data, and Space (Investigations), Saxon Math (Saxon), and Scott Foresman-Addison Wesley Mathematics (SFAW) provided one day each. The survey did not ask teachers if they

---

<sup>34</sup> Early editions were published by Silver, Burdett, Ginn, Inc. and later versions were published by Pearson Scott Foresman. Teachers did not indicate in their survey responses which edition they had previously used.

<sup>35</sup> Training dates were selected on the basis of district schedules, teacher availability, and trainer availability.

TABLE II.2  
CURRICULA PREVIOUSLY USED BY TEACHERS  
(Percentages)

	All Teachers	Teachers by Curriculum				<i>p</i> -value
		Investigations	Math Expressions	Saxon	SFAW	
Used the Assigned Curriculum at the K-3 Level At Some Point Prior to the Study	8.9	--	--	--	--	0.51
Taught Math in K-3 Last Year	90.2	90.6	89.3	89.7	91.2	0.99
Curriculum Used Last Year (among those who taught K-3 last year) <sup>a,*</sup>						
Everyday Math	17.4	--	--	--	--	0.00
Excel Math	9.2	--	--	--	--	
Harcourt Math	5.5	--	--	--	--	
Houghton Mifflin Math	6.4	--	--	--	--	
Saxon Math	23.9	--	--	--	--	
Silver Burdett Ginn Math	28.4	--	--	--	--	
Other	9.2	--	--	--	--	
Number of Years Used Last Year's Curriculum (among those who taught K-3 last year)	5.0	4.7	5.5	5.2	4.8	0.90
<b>Sample Size</b>	<b>127</b>	<b>33</b>	<b>29</b>	<b>29</b>	<b>36</b>	

Source: Author calculations using data from the fall 2006 teacher survey. The sample excludes one Math Expressions school (with 3 classrooms and 32 students) that participated during part of the school year and then stopped using the curriculum and did not allow the study to collect follow-up data.

Note: None of the statistics are significantly different across the curriculum groups, at the 5 percent level. The statistical tests were conducted using two-level HLMs (see the text at the beginning of Chapter II for details). A single *p*-value is reported for multinomial variables and indicates whether the fraction of teachers in each category of the variable differs across the curriculum groups.

<sup>a</sup>A small fraction reported more than one curriculum and were instructed to indicate the curriculum used most frequently, which is what is reported above.

\*Statistically significant at the 5 percent level. The statistical tests were conducted using two-level HLMs (see the text at the beginning of Chapter II for details). A single *p*-value is reported for multinomial variables and indicates whether the fraction of teachers in each category of the variable differs across the curriculum groups.

-- Value suppressed to protect respondent confidentiality.

TABLE II.3

TEACHER TRAINING ON THE ASSIGNED CURRICULUM  
(Percentage Unless Stated Otherwise)

	Teachers by Curriculum					<i>p</i> -value
	All Teachers	Investigations	Math Expressions	Saxon	SFAW	
<b>Initial Training</b>						
Attended Initial Training	100.0	100.0	100.0	100.0	100.0	1.00
Publisher-Specified Training Length	1-2 days	1 day	2 days	1 day	1 day	
Number of Days Attended*	1.2	1.0	2.0	1.0	1.0	0.00
How Well Prepared After Training*						
Very well	46.7	40.6	15.4	82.1	47.1	0.02
Adequate	38.3	50.0	38.5	17.9	44.1	
Somewhat or not at all	15.0	9.4	46.1	0.0	8.8	
<b>Follow-Up Training Reported on Fall Survey</b>						
Training Available as of Fall 2006*	69.4	100.0	14.3	51.7	100.0	0.00
Participated in Follow-Up Training*	62.9	100.0	10.7	27.6	100.0	0.00
<b>Sample Size</b>	<b>127</b>	<b>33</b>	<b>29</b>	<b>29</b>	<b>36</b>	
<b>Follow-Up Training Reported on Spring Survey</b>						
Training Available as of Spring 2007	97.5	96.7	100.0	96.7	96.9	1.00
Participated in Follow-Up Training	95.7	96.7	96.2	93.3	96.8	0.89
Number of Days Attended Follow-Up Training* (among those who attended)	1.5	2.9	0.5	0.4	2.2	0.00
<b>Sample Size</b>	<b>118</b>	<b>30</b>	<b>26</b>	<b>30</b>	<b>32</b>	

Source: Author calculations using data from the fall 2006 teacher survey, spring 2007 teacher survey, and study records on training attendance. The sample excludes one Math Expressions school (with 3 classrooms and 32 students) that participated during part of the school year and then stopped using the curriculum and did not allow the study to collect follow-up data.

Notes: Initial training was conducted by the publishers right before or soon after the first day of school. Follow-up training was conducted during the school year. On the spring survey, teachers were asked to report all follow-up training. Therefore, the spring information reflects all follow-up training, and the fall information reflects follow-up training that occurred by October or early November.

\*Statistically significant at the 5 percent level. The statistical tests were conducted using two-level HLMs (see the text at the beginning of Chapter II for details). A single *p*-value is reported for multinomial variables and indicates whether the fraction of teachers in each category of the variable differs across the curriculum groups.

attended the full amount of initial training provided, but study records tracked the length of attendance at training and indicate that 97 percent of teachers attended the full amount of their initial training.<sup>36</sup>

On the fall survey, teachers were asked how well the initial training prepared them to use their assigned curriculum with their students. More than 90 percent of teachers assigned to Investigations, Saxon, and SFAW indicated they felt either adequately or very well prepared to use their assigned curriculum, whereas 54 percent of the teachers assigned to Math Expressions felt similarly prepared.

## 2. Ninety-Six Percent of Teachers Attended Follow-Up Training on Their Assigned Curriculum

Each publisher also provided follow-up training and support to study teachers during the school year. Most trainers attempted to provide the first round of follow-up support within the first six weeks of school. Additional support was provided at different intervals for each curriculum. Trainers for Investigations and SFAW met with teachers every four to six weeks, Math Expressions trainers met with teachers up to two times during the school year, and Saxon trainers typically met with teachers once in the fall.

Unlike the initial trainings, the follow-up trainings were frequently provided to one school or one teacher at a time, and the structure of the training differed across and within the curricula. Each publisher provided information about the in-person support.

- ***Investigations.*** Trainers offered group training sessions prior to the start of each unit (about every four to six weeks). Sessions were typically three to four hours long and were held after school.
- ***Math Expressions.*** Trainers attempted to meet with teachers twice during the school year—once in the fall and again in the spring. Most follow-up support consisted of classroom observations followed by a short feedback session for teachers. Occasionally, trainers met with teachers as a group.
- ***Saxon.*** Trainers provided one follow-up session in the fall tailored to meet the needs of each district’s teachers. One district asked trainers to conduct demonstration lessons, after which the trainers met with teachers to debrief. Another district asked trainers to observe teachers and provide them with feedback, and yet another district asked trainers to provide teacher workshops.
- ***SFAW.*** Trainers offered group sessions about every four to six weeks throughout the school year. Sessions were typically three to four hours long and were held after school.

---

<sup>36</sup> Study records indicate that four teachers left training early.

These trainings typically were spread across many representatives from each publisher. In addition to in-person support, trainers were available for email and phone support throughout the school year.

Two sources of data about teacher participation in follow-up training were collected. Unlike the initial training, the study team did not attend each follow-up training session. Instead, each publisher received attendance forms to use at follow-up training sessions and was asked to return the completed forms to the study team soon after each follow-up training.<sup>37</sup> The study team was aware of all follow-up trainings that required study support, but may not have known about those that did not. The fall and spring teacher surveys provided an opportunity to obtain comprehensive information about follow-up training. On each survey, teachers were asked to report whether they had participated in any follow-up training to date, and the number of hours spent participating.

Attendance records of follow-up training provided by the publishers are consistent with teacher self-reports on the surveys. On the spring survey, 96 percent of teachers reported attending follow-up training, and the number of days attended varied by curriculum (Table II.3). Investigations and SFAW teachers reported attending 2.2 to 2.9 days of follow-up training, whereas Math Expressions and Saxon teachers reported attending 0.4 to 0.5 days.<sup>38</sup>

### **3. Other Sources of Professional Development**

On the spring survey, teachers were asked to report about non-study professional development received during the school year. Twenty percent of teachers reported receiving additional (non-study) professional development in math from other sources (Table II.4). Teachers participated in professional development related to math instruction, math content, performance standards in math education, and other math-focused professional development.

## **C. SCHOOL-BASED INSTRUCTIONAL SUPPORT**

The study team encouraged all math specialists and any other staff that districts, schools, or publishers indicated were important for curriculum implementation to participate in training. The study schools employed math specialists, such as math coaches and pull-out program teachers. Pull-out program teachers typically worked directly with students, whereas math

---

<sup>37</sup> The study team asked publishers about attendance if a form was not received within one week after a known follow-up training date.

<sup>38</sup> Information about follow-up training on the fall survey reflects follow-up training that occurred by October or early November.

TABLE II.4

NON-STUDY MATH PROFESSIONAL DEVELOPMENT DURING THE 2006-2007 SCHOOL YEAR  
(Percentages)

	Teachers by Curriculum					<i>p</i> -value
	All Teachers	Investigations	Math Expressions	Saxon	SFAW	
Participated in Any Non-Study Math PD	19.5	16.7	26.9	16.7	18.8	0.90
<b>Sample Size</b>	<b>118</b>	<b>30</b>	<b>26</b>	<b>30</b>	<b>32</b>	

Source: Author tabulations using data from spring 2007 teacher survey. The sample excludes one Math Expressions school (with 3 classrooms and 32 students) that participated during part of the school year and then stopped using the curriculum and did not allow the study to collect follow-up data.

Note: None of the statistics are significantly different across the curriculum groups, at the 5 percent level. The statistical tests were conducted using two-level HLMs (see the text at the beginning of Chapter II for details).

coaches provided support to teachers but typically did not work directly with students.<sup>39</sup> Other staff who were encouraged to participate included anyone who either directly or indirectly could be important for curriculum implementation.<sup>40</sup> While it was not the study team's responsibility to achieve a particular level of implementation, the team's goal was to establish a supportive environment that could facilitate the level of implementation publishers set out to achieve.

### 1. Seventy-Three Percent of Teachers Had Access to a Math Coach

Seventy-three percent of teachers reported that they had a school math coach or district math specialist available to help them with math instruction (Table II.5).<sup>41</sup> Among teachers who had a math coach or specialist available, 86 percent reported that these individuals were accessible either sometimes or almost always. Study records of training attendance also suggest that the math coaches were knowledgeable about the school's assigned curriculum. As mentioned in Chapter I, a total of 131 teachers participated in the study during the 2006-2007 school year. In addition to these teachers, study records indicate that 70 additional individuals attended the

<sup>39</sup> Math coaches typically serve as resources to classroom teachers for various tasks such as lesson planning, helping teachers stay informed about the curriculum resources available for use with students, or helping teachers with questions about math content, pedagogy, curriculum pacing, and preparation for standardized testing.

<sup>40</sup> In some study schools principals or assistant principals attended training, and in at least one district, a district-level curriculum coordinator attended the initial training.

<sup>41</sup> Four percent of teachers did not know if a math coach or district specialist was available.

TABLE II.5  
INSTRUCTIONAL SUPPORT AT STUDY SCHOOLS  
(Percentages)

	Teachers by Curriculum					<i>p</i> -value
	All Teachers	Investigations	Math Expressions	Saxon	SFAW	
Math Coach/Specialist Available at School	73.0	53.1	85.7	77.8	77.1	0.10
Accessibility of Math Coach/Specialist (among those with one available)						
Almost always	31.8	--	--	--	--	0.08
Sometimes	54.5	--	--	--	--	
Rarely	10.3	--	--	--	--	
Not at all	0.0	--	--	--	--	
Don't know	3.4	--	--	--	--	
Another Teacher Routinely Assists with Math Instruction <sup>a</sup>	17.3	21.2	13.8	20.7	13.9	0.56
Another Adult Routinely Assists with Math Instruction	32.5	30.3	24.1	34.5	40.0	0.52
<b>Sample Size</b>	<b>127</b>	<b>33</b>	<b>29</b>	<b>29</b>	<b>36</b>	

Source: Author calculations using data from the fall 2006 teacher survey. The sample excludes one Math Expressions school (with 3 classrooms and 32 students) that participated during part of the school year and then stopped using the curriculum and did not allow the study to collect follow-up data.

Note: None of the statistics are significantly different across the curriculum groups, at the 5 percent level. The statistical tests were conducted using two-level HLMs (see the text at the beginning of Chapter II for details). A single *p*-value is reported for multinomial variables and indicates whether the fraction of teachers in each category of the variable differs across the curriculum groups.

<sup>a</sup>Other teachers include pull-out program teachers such as resource, special education, and English language learner teachers.

-- Value suppressed to protect respondent confidentiality.

initial or follow-up trainings provided by the publishers. Sign-in sheets collected at the trainings indicate that some of the additional attendees included math coaches and math specialists.<sup>42</sup>

Teachers also had other supports, such as resource and special education teachers. Seventeen percent of teachers indicated that they had resource or special education teachers who routinely helped during math lessons (Table II.5). In addition to the support staff, 33 percent of teachers had another adult who routinely helped with math instruction. These other adults included teaching aides or assistants who routinely helped study teachers during their math lessons. The publishers invited resource teachers, special education teachers, and other adults to attend training sessions offered on the assigned curriculum. Sign-in sheets collected at each training indicate that these support staff also attended curriculum training sessions.

## **2. Teachers Reported Having a Supportive Instructional Environment**

Ninety-two percent of teachers agreed or strongly agreed that they felt supported by other teachers to try out new ideas in teaching math (Table II.6). Eighty-six percent of teachers agreed or strongly agreed that administrators promote innovations in math education. Approximately 76 percent of teachers agreed or strongly agreed that teachers regularly share ideas about math instruction and that teachers regularly work with one another on math curriculum and instruction. In addition, 80 percent of teachers reported that all or most teachers within their school share ideas on teaching, and 78 percent of teachers reported that all or most teachers within their school offer advice or help one another.

## **D. SOME BASICS ABOUT TEACHER USE OF THE ASSIGNED CURRICULUM**

On the fall and spring surveys, teachers were asked about use of their assigned curriculum. This included questions related to general curriculum use and math instruction in their classrooms (such as, Are you using the math curriculum assigned to your school?), and permit comparisons across curriculum groups and across the school year. The spring survey also included a set of curriculum-specific questions that are useful for assessing teacher adherence to the curricula. This section presents the basic information about teacher use of their assigned curriculum, and the next section presents more detailed information about teacher adherence.

### **1. Nearly All Teachers (99 percent in the fall and 98 percent in the spring) Reported Using Their Assigned Curriculum**

Even though teachers agreed to participate in the study regardless of the curriculum they were assigned, and were trained on their assigned curriculum and received new curriculum

---

<sup>42</sup> Specific information on the number and type of non-primary classroom teachers who attended training is not provided because these data were collected only for staff who required payment for attending training. Not all math coaches or other support staff (such as resource teachers, special education teachers, or teacher's aides) were eligible for payment for attending training, and attendance data on teachers who did not require payments were not systematically collected.

TABLE II.6  
INSTRUCTIONAL CLIMATE AT STUDY SCHOOLS  
(Percentages)

	Teachers by Curriculum					<i>p</i> -value
	All Teachers	Investigations	Math Expressions	Saxon	SFAW	
Teachers Agree or Strongly Agree with the Following Statements Regarding the Conditions for Teaching Math in Their School:						
Supported by other teachers to try out new ideas in teaching math	91.9	93.9	96.3	89.7	88.2	0.56
Administrators promote innovations in math education	86.2	93.9	92.6	75.9	82.4	0.12
Teachers regularly share ideas about math instruction	76.6	84.8	75.0	69.0	76.5	0.38
Teachers disagree about how to teach math	10.7	--	--	--	--	0.73
Teachers regularly work with one another on math curriculum and instruction	76.4	75.8	63.0	82.8	82.4	0.40
A specialist in math education regularly works with teachers in this school	24.4	12.1	25.9	27.6	32.4	0.39
Most curriculum changes introduced at this school gain little support among teachers	15.6	12.5	14.8	10.3	23.5	0.38
Most or All Teachers Within a School Interact the Following Ways:						
Work together to develop curriculum and instructional materials	51.6	42.4	57.1	58.6	50.0	0.44
Offer advice or help to each other	78.2	75.8	75.0	82.8	79.4	0.95
Share ideas on teaching	79.8	81.8	71.4	79.3	85.3	0.83
Promote new or innovative teaching practices	57.3	60.6	50.0	55.2	61.8	0.87
<b>Sample Size</b>	<b>127</b>	<b>33</b>	<b>29</b>	<b>29</b>	<b>36</b>	

Source: Author calculations using data from the fall 2006 teacher survey. The sample excludes one Math Expressions school (with 3 classrooms and 32 students) that participated during part of the school year and then stopped using the curriculum and did not allow the study to collect follow-up data.

Note: None of the statistics are significantly different across the curriculum groups at the 5 percent level. The statistical tests were conducted using two-level HLMs (see the text at the beginning of Chapter II for details).

-- Value suppressed to protect respondent confidentiality.

materials to use with their students, the question remains: Did teachers use their assigned curriculum? According to the teacher survey responses, the answer is yes.

Ninety-nine percent of teachers reported using their assigned curriculum as their core curriculum on the fall survey, and ninety-eight percent reported doing so on the spring survey (Tables II.7 and II.8). Early in the school year, Investigations teachers reported spending more time preparing to teach math than teachers using the other three curricula. On the fall survey, Investigations teachers reported spending 3.2 hours per week preparing to teach math. Teachers in the other curriculum groups spent 2.0 to 2.7 hours. Toward the end of the school year, the difference in prep time disappeared, and all four curriculum groups reported spending a similar amount of time (2.5 hours per week) preparing for math instruction.

## **2. Eighty-Eight Percent of Teachers Completed at Least 80 Percent of Their Curriculum**

In addition to using the curricula at the two points in time, the data suggest that teachers regularly used their curriculum throughout the school year (Table II.8). Eighty-eight percent of teachers reported completing 80 to 100 percent of their assigned curriculum on the spring survey. In each district, the school year lasted 10 months, and teachers completed the spring surveys 8 months into the school year—that is, 80 percent through the school year.

Teachers also had a favorable attitude toward their assigned curriculum. Eighty-two percent of teachers said they were very likely or likely to use their curriculum again, if they were given a choice (Table II.8).

## **3. One-Third of Teachers Supplemented with Other Materials**

Although nearly all teachers reported using their assigned curriculum as their core math curriculum, one-third reported supplementing with other materials (Tables II.7 and II.8).<sup>43</sup>

- ***Frequency of Supplementation.*** On the fall and spring surveys, 72 and 88 percent, respectively, reported supplementing at least once or twice a week.
- ***Reasons for Supplementation.*** Teachers reported various and multiple reasons for supplementation, including remediation, enrichment, and supplementing units or lessons in the assigned curriculum.

---

<sup>43</sup> Due to small sample sizes, statistical tests could not be performed across curriculum groups on the reasons for supplementation, frequency of supplementation, or materials used for supplementation.

TABLE II.7  
TEACHER INSTRUCTION AS REPORTED IN THE FALL  
(Percentage Unless Stated Otherwise)

	Teachers by Curriculum					<i>p</i> -value
	All Teachers	Investigations	Math Expressions	Saxon	SFAW	
Used the Curriculum Assigned by the Study as the Core Curriculum	99.2	97.0	100.0	100.0	100.0	1.00
Average Preparation per Week (hours)*	2.6	3.2	2.7	2.0	2.3	0.02
Supplemented the Assigned Curriculum with Other Materials	34.1	30.3	37.0	34.5	35.3	0.89
Frequency of Supplementation (among those who supplemented)						
Almost daily	36.1	--	--	--	--	
1–2 times per week	36.1	--	--	--	--	
1–2 times per month	27.8	--	--	--	--	
Reasons for Supplementation (among those who supplemented)						
Remediation with a small group	36.6	--	--	--	--	
Remediation with the entire class	24.4	--	--	--	--	
Enrichment with a small group	19.5	--	--	--	--	
Enrichment with the entire class	63.4	--	--	--	--	
Supplement to units or lessons	41.5	--	--	--	--	
Other	14.6	--	--	--	--	
Materials Used for Supplementation (among those who supplemented)						
Everyday Math	7.9	--	--	--	--	
Math Their Way	7.9	--	--	--	--	
Saxon Math	7.9	--	--	--	--	
Teacher Created	36.8	--	--	--	--	
Other	39.5	--	--	--	--	
<b>Sample Size</b>	<b>127</b>	<b>33</b>	<b>29</b>	<b>29</b>	<b>36</b>	

Source: Author calculations using data from the fall 2006 teacher survey. The sample excludes one Math Expressions school (with 3 classrooms and 32 students) that participated during part of the school year and then stopped using the curriculum and did not allow the study to collect follow-up data.

\*Statistically significant at the 5 percent level. The statistical tests were conducted using two-level HLMs (see the text at the beginning of Chapter II for details). Statistical tests could not be performed on the frequency of supplementation, reasons for supplementation, and materials used for supplementation due to small sample sizes.

-- Value suppressed to protect respondent confidentiality.

TABLE II.8  
TEACHER INSTRUCTION AS REPORTED IN THE SPRING  
(Percentage Unless Stated Otherwise)

	Teachers by Curriculum					<i>p</i> -value
	All Teachers	Investigations	Math Expressions	Saxon	SFAW	
Used the Curriculum Assigned by the Study as the Core Curriculum	98.3	93.1	100.0	100.0	100.0	1.00
Average Preparation per Week (hours)	2.5	2.6	2.7	2.6	2.2	0.75
Hours per Week of Math Instruction*	5.1	4.7	4.9	6.1	4.9	0.01
Completed at Least 80 Percent of the Lessons from the Assigned Curriculum	87.9	80.0	88.5	96.7	86.7	0.24
Supplemented the Assigned Curriculum with Other Materials	36.4	36.7	30.8	40.0	37.5	0.78
Frequency of Supplementation (among those who supplemented)						
Almost daily	39.6	--	--	--	--	
1–2 times per week	48.8	--	--	--	--	
Less than 1–2 times per week	11.6	--	--	--	--	
Reasons for Supplementation (among those who supplemented)						
Remediation with a small group	48.8	--	--	--	--	
Remediation with the entire class	39.5	--	--	--	--	
Enrichment with a small group	30.2	--	--	--	--	
Enrichment with the entire class	46.5	--	--	--	--	
Replacement for units or lessons	16.3	--	--	--	--	
Supplement to units or lessons	74.4	--	--	--	--	
Other	18.6	--	--	--	--	
Materials Used for Supplementation (among those who supplemented)						
Everyday Counts	11.6	--	--	--	--	
Everyday Math	9.3	--	--	--	--	
Excel Math	11.6	--	--	--	--	
Math Their Way	7.0	--	--	--	--	
Saxon Math	11.6	--	--	--	--	
Teacher Created	25.6	--	--	--	--	
Other	23.3	--	--	--	--	

TABLE II.8 (continued)

	Teachers by Curriculum					<i>p</i> -value
	All Teachers	Investigations	Math Expressions	Saxon	SFAW	
Likelihood of Using Assigned Curriculum Again, if Given a Choice						
Very likely	43.9	--	--	--	--	0.27
Likely	37.7	--	--	--	--	
Not at all likely	18.4	--	--	--	--	
<b>Sample Size</b>	<b>118</b>	<b>30</b>	<b>26</b>	<b>30</b>	<b>32</b>	

Source: Author calculations using data from the spring 2007 teacher survey. The sample excludes one Math Expressions school (with 3 classrooms and 32 students) that participated during part of the school year and then stopped using the curriculum and did not allow the study to collect follow-up data.

\*Statistically significant at the 5 percent level. The statistical tests were conducted using two-level HLMs (see the text at the beginning of Chapter II for details). A single *p*-value is reported for multinomial variables and indicates whether the fraction of teachers in each category of the variable differs across the curriculum groups. Statistical tests could not be performed on the frequency of supplementation, reasons for supplementation, and materials used for supplementation due to small sample sizes.

-- Value suppressed to protect respondent confidentiality.

- **Materials Used for Supplementation.** Supplemental materials used by teachers varied widely. The largest percentage of teachers reported using teacher-created supplemental materials. They also reported using an assortment of commercially available curriculum materials, such as Everyday Counts, Everyday Math, Excel Math, Math Their Way, and Saxon Math.<sup>44</sup> Everyday Counts and Math Their Way are supplemental programs, whereas Everyday Math, Excel Math, and Saxon Math are full curricula.

A 2008 national survey of the math market indicates that, among classroom teachers, teacher-created materials are the most commonly used supplemental materials—similar to what we observe in this study (Education Market Research 2008). The survey also found that teachers use a wide variety of commercially available supplemental materials, including materials from supplemental products and full curricula.

<sup>44</sup> In the fall (Table II.7), the largest percentage of teachers reported materials that are contained in the “other” category. These “other” materials contain numerous brand name products, but in most cases only one or two teachers reported each product and, therefore, the products are not reported separately to protect respondent confidentiality.

#### **4. Saxon Teachers Spent One More Hour on Math Instruction per Week**

Saxon teachers reported devoting one hour more per week to math, compared to teachers in the other three curriculum groups (Table II.8). In the spring survey, teachers reported the number of days per week and the number of minutes per day devoted to mathematics. An “hours per week” variable was constructed from this information. Investigations, Math Expressions, and SFAW teachers reported an average of 4.8 hours devoted to mathematics, while Saxon teachers reported an average of 6.1 hours.

#### **E. TEACHER ADHERENCE TO THE ESSENTIAL FEATURES OF THE CURRICULA**

This section examines the extent to which teachers adhered to the essential features of their assigned curriculum. To make this assessment, the study team had to determine what teachers should be doing with their curriculum and how often. Questions about “what” and “how often” would then be included on the spring teacher survey, and teachers would be asked to reflect back on the school year when answering the questions. Teachers would only receive questions for their assigned curriculum.

To define adherence, the study team reviewed each curriculum’s materials in depth to identify the essential and secondary features of each curriculum, and the recommended frequency with which each activity or practice should be implemented. Many of the essential and secondary activities are defined in Appendix C.<sup>45</sup>

Below, we summarize teacher responses to questions about their usage of the essential activities and practices of their assigned curriculum, and compare those responses to the expected frequency.<sup>46</sup> The final section in this chapter summarizes teacher reports of the number of lessons covered in 20 math content areas, and whether there were any differences across the curricula.

Three caveats are important to consider. First, the definition of adherence for each curriculum was specified by the study team after careful review of the curriculum materials. Conversations were held with the publishers to discuss draft definitions, and the publishers’ comments were considered as the study team developed final definitions.

Second, the conclusions about adherence are based on small teacher sample sizes for each curriculum group, and on analyses of individual adherence items. A more accurate assessment of implementation may require examining combinations of activities implemented on a given day or the relative frequency of activities (such as spending more time on teaching a new concept, rather than on fluency activities). The study’s follow-up report, described in Chapter I, will have

---

<sup>45</sup> Most terms that are self-explanatory are not included in Appendix C, with the exception of those that have curriculum-specific definitions.

<sup>46</sup> Appendix B contains tables that present teacher responses to the additional activities and practices of each curriculum.

a larger sample size that should permit the use of alternate analytic approaches for assessing implementation, and can be used to assess the sensitivity of the results presented here.

Third, the study's follow-up report will present information about adherence based not only on the spring teacher survey, but also on classroom observations. Chapter I mentioned that the study team is observing classrooms, and explained that the observation data are not presented in this report because the reliability of those data cannot be assessed until observations have been completed in all the study schools. The classroom observation protocol and spring teacher survey include some comparable items that will be used to examine the consistency of information collected through classroom observations and teacher surveys.

## **1. Descriptions of the Curricula and Teacher Adherence**

The study's curricula include a range of instructional approaches, from teacher-directed approaches to student-centered ones that are more aligned with social constructivist learning theory. While many of the curricula share common features, they are distinguished from one another in the emphasis placed on different instructional practices. The common features and differences are summarized in this section.

The curricula descriptions provided below are not comprehensive and exhaustive. The curricula have many features, some of which can vary across grade levels. The descriptions provided in this report focus on the features most evident in first grade classrooms, and are intended to be consistent with the way publishers describe their products and expect them to be used. The descriptions begin with abstracts provided by the publishers, followed by a summary generated by the study team.

After the description of each curriculum, a summary of teachers' responses to the essential features of their assigned curricula is presented. Teachers were asked to indicate how frequently they implemented their assigned curriculum's activities and practices on a six-point ordinal scale that included 0 (never), 1 (less than once a month), 2 (once or twice a month), 3 (one to two times per week), 4 (three to four times per week), and 5 (daily). Investigations and Math Expressions teachers also reported on the degree of success they had in facilitating the types of discussions called for in the respective curriculum. A four-point ordinal scale from 1 (not at all successful) to 4 (very successful) was used.

Tables II.9 through II.14 (which are included later in the text when discussed more fully) report the mean and median teacher responses for each curriculum's activities, along with the expected frequency of implementing the activities and the percentage of teachers who reported meeting the expected frequency.<sup>47</sup> For example, a mean of 4 (three to four times per week) for a particular activity indicates that it occurred three to four times a week in the average classroom,

---

<sup>47</sup> All teachers who responded to the survey are included in the analysis, including the small fraction who reported not using their assigned curriculum on the spring survey. Because we are currently working with a small sample size for each curriculum group that can affect the precision of the mean response, we also report the median response.

while a median of 5 (daily) indicates that at least half the teachers implemented the activity on a daily basis. The activities in each table are listed in order of average frequency, from highest to lowest.

Although daily implementation of most practices generally might be interpreted to mean stronger implementation, not all curricula encourage implementation of all activities on a daily basis and some activities (such as some types of assessments) should occur less frequently. The first column in Tables II.9, II.11, II.13, and II.14 indicates the expected frequency, and the discussion below compares the results to the expected frequency. That said, the curriculum materials are not always clear on how frequently activities or practices should be implemented. In addition, some activities and practices depend on the strengths and needs of individual students or the class as a whole. For example, implementing an error intervention is dependent on students making errors.

To assess adherence, we looked at each essential activity individually and examined the percentage of teachers who reported implementing each one with the expected frequency. Adherence is then defined as the percentage of teachers who implemented each essential activity with the expected frequency. In general, stronger adherence would be expected when a large percentage of teachers implement an essential activity with the expected frequency and when a large percentage of the essential activities are implemented as such. Activities without a clearly specified expected frequency were excluded from the assessment.

#### **a. Investigations**

*Curriculum Abstract.* Investigations is a K-5 mathematics curriculum developed by TERC under a grant from the National Science Foundation. Its four major goals are to:

- Offer students meaningful mathematical problems
- Emphasize depth in mathematical thinking rather than superficial exposure to a series of fragmented topics
- Communicate mathematics content and pedagogy to teachers
- Expand substantially the pool of mathematically literate students

Investigations offers in-depth experiences in number, data, geometry, and the mathematics of change. The following aspects of the curriculum ensure that all students are included in significant mathematical learning by:

- Spending time exploring problems in depth
- Finding more than one solution to many problems
- Developing their own strategies and approaches, based on their knowledge and understanding of mathematical relationships

- Choosing from a variety of concrete materials and appropriate technology, including calculators, as a natural part of their everyday mathematical work
- Expressing their mathematical thinking through drawing, writing, and talking

Each grade level is organized into units that involve students in the exploration of major mathematical ideas, and may revolve around two or three related areas—for example, addition and subtraction or geometry and fractions.

The curriculum is presented through a series of teacher books. Each book provides lesson plans, materials lists, reproducible student sheets for activities and games, a family letter, homework suggestions, opportunities for skill and practice, assessment activities, notes to the teacher about the mathematics students are encountering, and examples of classroom dialogues. Some units include software to extend students' experience with the mathematics being explored. In addition to the curriculum units, Student Activity Books, and Investigations at Home Booklets, and End of Unit Assessment Sourcebooks are also available for each unit in grades 1-5.

***Curriculum Description.*** Investigations uses a student-centered approach that implements instructional practices aligned with constructivist learning theory. The content is presented in thematic units, and activities within each unit include real-life problems that students are to solve in multiple ways. The curriculum emphasizes metacognition (thinking about one's own reasoning and the reasoning of one's peers) and communicating about mathematics in multiple ways rather than focusing on getting the correct answer. Students work on a smaller number of problems in a class session, may work on a single problem across multiple sessions, and regularly use manipulatives. A 10-minute set of routine activities that provide daily arithmetic and data analysis practice is recommended in each unit.

The Investigations curriculum is designed to have students work in pairs or small groups and talk to one another about their work. Teachers spend much of their time facilitating conversations among students, helping students express their thoughts, and guiding students to a deeper understanding of the mathematical concepts they are working on. Classroom activities often vary by day and depend on the length of the investigation. For example, during an investigation lasting one week, on the first day the teacher will introduce the investigation (new concept) to the class, often through large group hands-on activities with the students. During the next two to three days, students will work in pairs or small groups to explore the concept, by working on one or two in-depth problems each day, playing mathematical games, or working on choice time activities. At the end of each day, they frequently discuss as a group what they worked on that day. In the last session of the investigation, the students and teacher will discuss as a group what they learned during the investigation and the strategies they used to solve problems.

***Teacher Adherence.*** In classrooms using Investigations, we would expect to see manipulatives available to students and students discussing different ways of solving a problem on a daily basis (Table II.9). Activities such as choice time and writing about how to solve a problem would be expected one to three times a week. In addition to implementing a variety of

TABLE II.9

INVESTIGATIONS: TEACHER-REPORTED FREQUENCY OF IMPLEMENTING  
ESSENTIAL CURRICULUM ACTIVITIES (N = 31)

Activity (Scale)	Expected Frequency	Met Expected Frequency (Percentage)	Mean Response	Median Response
<b>Teacher Activities</b>				
Make manipulatives accessible to students at all times during the lesson	5	93.1	4.90	5
Conduct at least one activity from the current Investigation	5	72.4	4.55	5
Allow students to choose manipulatives for use during the activity	5	75.9	4.48	5
Invite students to use multiple strategies or solutions to a problem	5	32.3	4.23	4
Prompt students to explain their answers	5	32.3	4.16	4
Refer to the 100 Chart	NS	NS	4.07	5
Ask students to demonstrate a procedure or concept to other students	4	83.9	4.06	4
End each lesson by asking students to share their thinking	4-5	72.4	3.79	4
Do choice time activities	3	89.7	3.41	3
Ask students to explore a concept or procedure before it is modeled	3-4	83.9	3.19	4
Use Teacher Checkpoints and Embedded Assessments	2-3	82.8	2.52	2
<b>Student Activities</b>				
Use manipulatives, pictures, or diagrams to solve problems	5	83.9	4.81	5
Discuss different ways of solving a problem	4-5	80.6	4.42	5
Explain a math concept or procedure to other students	4-5	80.6	4.26	4
Do problems that have more than one correct solution	3-4	93.5	4.06	4
Write about how to solve a problem	3	80.6	3.23	3

Source: Author tabulations using data from the spring 2007 teacher survey. The sample excludes two Investigations teachers who did not complete the above items in the survey.

Note: Teachers were asked to indicate how frequently they implemented the activities on the following scale: 0 (never), 1 (less than once a month), 2 (once or twice a month), 3 (one to two times a week), 4 (three to four times a week), and 5 (daily). A mean of 4 indicates that teachers implemented an activity an average of three to four times a week.

NS indicates the expected frequency was not specified.

instructional activities, we would expect to see teachers facilitating discussions that call for metacognitive thinking and to elicit from students multiple solutions to problems (Table II.10).

Investigations recommends a frequency of implementation for 15 of the 16 essential activities listed in Table II.9. Adherence to the 15 items (that is, the percentage of teachers who met the expected frequency) ranged from 32 to 94 percent. For 13 of the 15 activities, at least 72 percent of teachers reported implementing them with the expected frequency. Among activities that involved manipulatives, at least 75 percent of teachers reported implementing the activity on a daily basis, as recommended. On average, Investigations teachers reported doing other key activities at least once a week, except for using teacher checkpoints and embedded assessments, which is not expected to occur as frequently.

Overall, Investigations teachers also reported being moderately successful in implementing discussions that asked students to explain reasoning, discuss concepts, and share multiple approaches to solutions. As shown in Table II.9, 81 percent of the teachers reported that students discussed different ways of solving problems with the expected frequency. In addition, teachers on average thought they were moderately to very successful at facilitating discussions that allow students to explain their thinking, and facilitating discussions that enable students to offer or share multiple approaches to solving a problem (Table II.10).

TABLE II.10

INVESTIGATIONS: TEACHER-REPORTED SUCCESS AT FACILITATING DISCUSSIONS FOCUSED ON PROCESS (N = 31)

Type of Discussion (Scale)	Mean Response	Median Response
Discussions that allow students to explain their answers	3.39	4
Discussions that enable students to offer or share multiple approaches to solving a problem	3.32	4
Discussions that enable students to raise mathematical questions and/or discuss mathematical concepts	3.00	3

Source: Author tabulations using data from the spring 2007 teacher survey. The sample excludes two Investigations teachers who did not complete the above items in the survey.

Note: Teachers rated their success at facilitating discussions on the following scale: 1 (not at all successful), 2 (somewhat successful), 3 (moderately successful), and 4 (very successful).

## b. Math Expressions

*Curriculum Abstract.* Math Expressions is a complete Kindergarten through Grade 5 curriculum based on the research results of the Children’s Math Worlds (CMW) project. The CMW project was conducted by Dr. Karen C. Fuson, now professor emerita of learning sciences at Northwestern University, Evanston, Illinois, and funded over a ten-year period by the National

Science Foundation. Both the program and the research combine a focus on conceptual understanding with opportunities to develop fluency with problem solving and computation. Math Expressions incorporates approaches from both reform and traditional mathematics programs while contributing new and effective teaching strategies to mathematics instruction. Key aspects of this curriculum include application of accessible algorithms that can be more easily understood and used by students; use of student math drawings and research-based visual representations to support student understanding and class discussion of mathematical thinking; an emphasis on in-depth sustained learning of core grade-level concepts (rather than a spiral curriculum) to support students' conceptual understanding and fluency; and a "learn by teaching" design to support teachers new to the curriculum. Embedded in the program are five core classroom structures—Building Concepts, Math Talk, Student Leaders, Quick Practice, and Helping Community—that support children from all backgrounds in developing mathematical understanding, competence, and confidence.

***Curriculum Description.*** Math Expressions is a relatively new curriculum, which uses both teacher-directed and student-centered instructional approaches. The curriculum encourages teachers to teach students efficient and effective procedures, while also promoting children's natural solution methods. Math Expressions is organized to provide sustained work on key concepts, rather than spiraling lessons.<sup>48</sup> The program emphasizes the development of student leaders, a collaborative classroom culture, and "math talk," which involves children talking about and representing their thinking.

The Math Expressions curriculum is designed to begin each day with a series of routines involving the calendar, money, a number chart, and counting. The math lesson often occurs later in the day, and begins with a 'quick practice' fluency activity. Afterwards, the teacher often conducts a whole class lesson in which new information is introduced and students are encouraged to discuss and demonstrate mathematical ideas (using math talk). The teacher fosters this discussion while introducing efficient procedures, and visual learning supports are used to help students link their knowledge to formal mathematical concepts. Students then practice the new skill or concept in pairs, small groups, or individually. Student leaders, math talk, and a helping community (where everyone is considered a teacher and a learner) are emphasized in all portions of the math lesson.

***Teacher Adherence.*** We would expect to see half of the activities listed in Table II.11 occurring daily. The other activities would be expected at least once a week or three to four times a week. For example, all of the math talk structures are not expected to occur daily. Math talk structures are specified activities and ways of interacting about mathematics, and they include *Solve and Discuss*, *Step-by-Step*, and *Scenarios*. Usually only one or two of these structures is present in a lesson, so we would not expect to see each of them daily. Other activities, such as administering quick quizzes, also are not expected to occur daily.

---

<sup>48</sup> Spiraling refers to the practice of introducing a concept or procedure in a lesson at an elementary level and then revisiting the concept later and bringing students to the next level of understanding. This spiraling can continue throughout the school year and/or across school years. Bruner (1960) first coined the phrase in 1960 as a way to structure curriculum around big ideas.

TABLE II.11

MATH EXPRESSIONS: TEACHER-REPORTED FREQUENCY OF IMPLEMENTING  
ESSENTIAL CURRICULUM ACTIVITIES (N = 27)

Activity (Scale)	Expected Frequency	Met Expected Frequency (Percentage)	Mean Response	Median Response
<b>Teacher Activities</b>				
Assign homework	5	63.0	4.48	5
Use Quick Practice activity	5	66.7	4.37	5
Complete the Daily Routines for the unit	5	61.5	4.23	5
Use proof drawings	4	88.5	4.19	4
Use student leaders during the Daily Routines	4-5	70.4	4.04	5
Ask students to demonstrate a procedure or concept to other students	5	18.5	3.96	4
Use Step-by-Step at the board	3-4	92.6	3.93	4
Use Solve and Discuss at the board	3-4	88.9	3.89	4
Use Scenarios	3-4	85.2	3.78	4
Use student leaders during the Quick Practice activity	4-5	59.3	3.70	4
Administer Quick Quizzes	3	33.3	2.00	2
<b>Student Activities</b>				
Use manipulatives, pictures, or diagrams to solve problems	5	59.3	4.37	5
Explain a math concept or procedure to other students	5	33.3	3.93	4
Ask mathematical questions of other students	5	29.6	3.15	3
Write about how to solve a problem	NS	NS	2.85	3

Source: Author tabulations using data from the spring 2007 teacher survey.

Note: Teachers were asked to indicate how frequently they implemented the activities on the following scale: 0 (never), 1 (less than once a month), 2 (once or twice a month), 3 (one to two times a week), 4 (three to four times a week), and 5 (daily). A mean of 4 indicates that teachers implemented an activity an average of three to four times a week.

NS indicates the expected frequency was not specified.

Of the 15 activities listed in Table II.11, a total of 14 have a recommended frequency of implementation. Adherence to the 14 activities ranged from 19 to 93 percent. For 10 of the 14 activities, at least 59 percent of Math Expressions teachers reported implementing them with the expected frequency (Table II.11). Teachers reported assigning homework, doing quick practice, completing the daily routines, using proof drawings, and using student leaders during the routine with their class an average of three to four times a week, with a median implementation of daily on all of these activities except proof drawings. Most of the other essential curriculum practices were done at least once a week on average, except for quick quizzes.

Similar to Investigations, Math Expressions teachers are expected to involve students in discussions that call for metacognitive thinking. In addition, Math Expressions teachers should encourage children to take leadership roles by posing questions to one another and commenting on the thinking of others. Metacognitive discussions and student leader use are expected to occur in Math Expressions classrooms, even in the early grades. On average, Math Expressions teachers thought they were moderately to very successful in implementing these discussions, but expressed less success with enabling students to ask mathematical questions and encouraging students to build on the ideas of classmates (Table II.12).

TABLE II.12

MATH EXPRESSIONS: TEACHER-REPORTED SUCCESS AT FACILITATING DISCUSSIONS FOCUSED ON PROCESS (N = 27)

Type of Discussion (Scale)	Mean Response	Median Response
Discussions that allow students to explain their answers	3.56	4
Discussions that enable students to offer or share multiple approaches to solving a problem	3.41	3
Discussions that enable students to raise mathematical questions and/or discuss mathematical concepts	2.89	3
Discussions that encourage students to reference other students' ideas in their comments	2.63	3

Source: Author tabulations using data from the spring 2007 teacher survey.

Note: Teachers rated their success at facilitating discussions on the following scale: 1 (not at all successful), 2 (somewhat successful), 3 (moderately successful), and 4 (very successful).

### c. Saxon

**Curriculum Abstract.** For almost 20 years, Saxon has been providing elementary math curriculum that uses a multisensory approach designed to enable all children to develop a solid foundation in the language and basic concepts of mathematics. The program is intended to align with how young children learn and build fluency with math skills. This is accomplished through

hands-on activities and mathematical conversations that actively engage students in the learning process. Concepts are developed, reviewed, and practiced over time supported by a philosophy that believes that understanding follows doing and discussing; mastery follows learning over time, and fluency follows practicing over time.

Saxon is an imprint of Harcourt Achieve, Inc. Harcourt Achieve produces learning solutions and content that fundamentally and positively change the lives of young and adult learners. Published under the Rigby, Saxon, and Steck-Vaughn imprints, its products are based on a developmental philosophy that assesses learners' skills, matches them to appropriate content, and accelerates them to meet and exceed expectations. The Rigby imprint offers progressive learning solutions for core reading and English language learner instruction that provide differentiated instruction to match each student's instructional level. The Saxon imprint offers the nation's bestselling and most thoroughly researched skills-based mathematics program for grades K-12, as well as popular phonics, K-3 spelling, and early learning programs. The Steck-Vaughn imprint offers easy-to-use, innovative learning solutions that accelerate content-area knowledge, reading skills, and preparation for standards-based tests, allowing learners to meet and exceed expectations.

***Curriculum Description.*** Saxon uses a teacher-directed instructional approach that provides scripted lesson plans for teachers. Each lesson integrates the mathematical strands and spirals them throughout the school year. New material is introduced gradually each day through explicit instruction and modeling by the teacher. Each lesson also includes daily distributed practice of previously learned concepts and procedures. The curriculum uses frequent and cumulative assessments to help teachers monitor student progress.

The Saxon curriculum is designed to begin each day with a Morning Meeting that lasts 15 to 20 minutes. The meeting is a whole class activity in which students practice skills related to the calendar, time, money, graphing, counting, place value, problem solving, and mental computation. The math lesson usually occurs later in the day, and begins with a whole class activity in which the teacher guides students to write the number of the day, and then explicitly teaches the new concept. Afterward, the teacher guides practice using a worksheet. At the end of each lesson, the teacher will ask a few students to summarize for the entire class what they learned that day. Independent practice is assigned as homework. In addition to the Morning Meeting and math lesson, students practice fluency of number facts (fact practice) on a daily basis, either orally or in writing with the support of self-correcting materials, manipulatives, fact cards, or worksheets. Fact practice can occur during the same time period as the math lesson, or at another time during the day. The curriculum also provides additional enrichment activities (journal writing topics, literature connections, computer technology activities) and activities for practicing test-taking strategies.

***Teacher Adherence.*** All 12 essential activities listed in Table II.13 have a recommended frequency of implementation, and adherence to the activities ranged from 37 to 93 percent. For 9 of the 12 activities, at least 63 percent of Saxon teachers reported implementing them with the expected frequency. Seven of the 12 activities are expected to occur daily, and the median frequency reported by teachers for 6 of those 7 activities is daily. The seventh activity (completing all activities specified in the lesson) had a median reported frequency of three to four times a week.

TABLE II.13

SAXON: TEACHER-REPORTED FREQUENCY OF IMPLEMENTING  
ESSENTIAL CURRICULUM ACTIVITIES (N = 30)

Activity (Scale)	Expected Frequency	Met Expected Frequency (Percentage)	Mean Response	Median Response
State the lesson's objective from the script	5	83.3	4.80	5
Ask students to complete the Guided Class Practice worksheet	5	86.7	4.80	5
Model completion of the Guided Class Practice chart	5	73.3	4.67	5
Use the manipulative and visual representations specified in the lesson	5	63.3	4.63	5
Ask students to respond to your questions as a whole group	4-5	86.7	4.43	5
Complete Fact Practice specified in the lesson	4-5	73.3	4.23	5
Adhere to the lesson script	4-5	76.7	4.17	5
Complete all activities specified in the lesson	5	36.7	4.07	4
Ask students at the end of the lesson to summarize what they learned	5	51.7	4.07	5
Complete all parts of the Meeting specified in the lesson	5	50.0	4.03	5
Complete Fact Assessment if specified in the lesson	3	86.7	3.60	3
Administer written assessments	3	93.3	3.47	3

Source: Author tabulations using data from the spring 2007 teacher survey.

Note: Teachers were asked to indicate how frequently they implemented the activities on the following scale: 0 (never), 1 (less than once a month), 2 (once or twice a month), 3 (one to two times a week), 4 (three to four times a week), and 5 (daily). A mean of 4 indicates that teachers implemented an activity an average of three to four times a week.

#### d. SFAW

**Curriculum Abstract.** SFAW promotes mathematical proficiency by focusing on the development of both mathematics skills and essential understandings. This is accomplished through:

- An articulation of essential outcomes and conceptual understandings for both the teacher and the student
- Questioning strategies that develop higher-order thinking skills embedded into the student and teacher materials
- Development of mathematical communication as a means of building a deep understanding of important mathematics

A hallmark of SFAW is explicit instruction of essential mathematics skills and concepts, using concrete manipulatives and pictorial and abstract representations. This approach helps to move all students forward in the development of mathematical proficiency. Ongoing assessment and diagnosis are coupled with strategic intervention to meet the individual needs of students, including frequent and timely student assessments integrated throughout the program to demonstrate student understanding and guide and monitor instruction. The authors of SFAW also recognize the importance of quality, ongoing professional development, and teacher support. Thus, professional development is provided daily within the teaching materials and is ongoing in multiple formats, including various uses of technology, to support the continued development of highly qualified teachers.

**Curriculum Description.** SFAW is a basal program with a teacher-directed instructional approach. The program offers a variety of optional materials for teachers to use, including problem-solving worksheets, literature connections, connections to other content areas, re-teaching activities, activities for English language learners, and computer programs.

The SFAW curriculum is designed around a consistent daily lesson structure including the following six activities: *Spiral Review* (a brief review of previously learned material), *Investigating the Concept* (hands-on exploration of the new concept), *Warm-Up* (a brief activity to activate prior knowledge and connect it to the new lesson), *Teach* (direct instruction of the new material), *Independent Practice* (typically using worksheets), and *Assessment* (a closure activity to check student understanding of the new concept). In the *Investigating the Concept* and *Teach* portions of the lesson, the teacher's manual includes questions that offer students the opportunity to verbalize their understanding.

**Teacher Adherence.** Of the 13 essential activities listed in Table II.14, a total of 8 have a recommended frequency of implementation. Adherence to the 8 activities ranged from 42 to 81 percent. For 6 of the 8 activities, at least 55 percent of SFAW teachers reported implementing them with the expected frequency (Table II.14).

In addition to the essential SFAW activities, teachers also reported using other curriculum-specific activities (see Table B.4 in Appendix B). Stating the lesson objective, step-by-step

TABLE II.14

SFAW: TEACHER-REPORTED FREQUENCY OF IMPLEMENTING  
ESSENTIAL CURRICULUM ACTIVITIES (N = 32)

Activity (Scale)	Expected Frequency	Met Expected Frequency (Percentage)	Mean Response	Median Response
Do the Investigating the Concept activity	5	65.6	4.47	5
Use manipulatives during the lesson	4-5	81.3	4.25	4
Differentiate math instruction for students at different ability levels	NS	NS	4.03	4
Do the Warm Up activity	5	41.9	3.81	4
Use the Think About It questions	4	59.4	3.75	4
Provide additional activities for “early finishers”	NS	56.3	3.69	4
Do the Spiral Review	5	48.4	3.65	4
Use the Talk About It questions	4-5	58.1	3.65	4
Ask students to complete the Learn! section of the student worksheets	4-5	54.8	3.52	4
Ask Students to complete the Test-Taking Practice	NS	NS	2.44	2
Provide the recommended Error Intervention for struggling students	NS	NS	2.38	3
Ask Students to complete the Journal Activity	NS	NS	2.28	2
Administer SFAW assessments	2	78.1	1.88	2

Source: Author tabulations using data from the spring 2007 teacher survey.

Note: Teachers were asked to indicate how frequently they implemented the activities on the following scale: 0 (never), 1 (less than once a month), 2 (once or twice a month), 3 (one to two times a week), 4 (three to four times a week), and 5 (daily). A mean of 4 indicates that teachers implemented an activity an average of three to four times a week.

NS indicates the expected frequency was not specified.

guidance on completing the practice page, providing reading assistance to students as they complete the practice page, and introducing the vocabulary specified in the lesson were the other frequently implemented aspects of the curriculum.

## 2. Content Coverage

How and when content is introduced to children has been a topic of discussion among educators. The recent report of the National Mathematics Advisory Panel (2008, p.11) called for emphasis on a “well-defined set of the most critical topics in the early grades.” The National Council of Teachers of Mathematics (NCTM 2006) released *Curriculum Focal Points* (CFP) to offer guidance on providing mathematics instruction that is more coherent and focused. However, the National Mathematics Advisory Panel (2008, p.21) noted that CFP calls for time devoted to some topics that do not receive emphasis in the early grades in the highest-achieving countries in the Trends in International Mathematics and Science Study (Ginsburg, Cooke, Leinwand, Noell, and Pollock 2005). The curricula in this study approach the introduction of content in varied ways, with some curricula (particularly Investigations) using a more focused, thematic approach to content and others (such as Saxon) spiraling content throughout the year.

We looked at the math content that was covered by teachers in each of the study’s curriculum groups, and whether the coverage differed by curricula. This information is drawn from the spring teacher survey, which asked teachers to indicate the number of lessons they taught in each of 20 math content areas using a scale of 0 (none-I did not teach this topic), 1 (1-5 lessons), 2 (6-10 lessons), 3 (11-15 lessons), or 4 (more than 15 lessons). A lesson is a set of activities that are intended to be completed in one math class, typically about an hour in length. Teachers reported the number of lessons taught in each content area, regardless of whether they used their assigned curriculum or other materials.

The mean emphasis for each content area is indicated in Table II.15 for each curriculum. The items are arranged from the topics most frequently taught when all the curriculum groups are pooled together, to those least frequently addressed. A mean of 3, for example, indicates that 11 to 15 lessons were focused on that content.

Across the curricula, teachers reported most frequently teaching lessons on adding and subtracting with whole numbers, counting with whole numbers, word problems, and addition and subtraction facts with whole numbers. In these areas, the average teacher taught 11 to 15 lessons. This is consistent with the recommendation of the National Mathematics Advisory Panel (2008) and with CFP, which lists “Developing understandings of addition and subtraction and strategies for basic addition facts and related subtraction facts” as the first focal point for grade one (NCTM 2006).

To explore whether emphasis on some topics varied across the curricula, we analyzed the average number of lessons in each topic area by curriculum, while controlling for classroom and school characteristics. Classroom characteristics included teacher education, experience, score on the content/pedagogical assessment, prior use of the assigned curriculum, timing of survey completion, class size, average fall class achievement, variance of the fall score, and skewness of the fall score. School characteristics included free/reduced-price meals eligibility, Title I status, and indicators for the curriculum groups.

TABLE II.15  
AVERAGE EMPHASIS IN VARIOUS MATH CONTENT AREAS

Number of Lessons on: <sup>a</sup>	Teachers by Curriculum					<i>p</i> -Value
	All Teachers	Investigations	Math Expressions	Saxon	SFAW	
Adding and subtracting with whole numbers	3.54	3.50	3.67	3.80	3.22	0.64
Counting with whole numbers	3.35	3.47	3.63	3.63	2.75	0.12
Word problems*	3.34	3.23	3.85	3.53	2.81	0.02
Addition and subtraction facts with whole numbers*	3.31	2.73	3.59	3.83	3.13	0.01
Creating, continuing, or predicting patterns	3.02	3.23	3.00	3.30	2.56	0.08
Understanding numbers less than 10	3.01	2.80	3.41	3.37	2.53	0.75
Collecting or analyzing data	2.68	3.13	2.48	2.77	2.34	0.34
Money*	2.65	1.94	3.11	3.30	2.34	0.00
Graphs	2.56	2.42	2.52	2.80	2.50	0.28
Place value with whole numbers*	2.46	1.61	2.70	3.10	2.50	0.04
Geometric shapes or spatial relationships	2.28	2.68	2.00	2.03	2.38	0.23
Time	2.13	1.81	1.63	2.27	2.75	0.18
Measurement with standard tools	1.67	1.29	1.67	2.20	1.53	0.14
Fractions*	1.58	0.94	1.59	1.87	1.94	0.02
Nonstandard measurement	1.25	1.19	1.11	1.47	1.22	0.55
Probability*	1.05	0.84	1.33	0.60	1.44	0.02
Multiplying and dividing with whole numbers	0.21	0.13	0.23	0.33	0.16	0.17
Decimals*	0.13	0.16	0.26	0.07	0.03	0.01
Multiplication and division facts with whole numbers	0.07	0.03	0.15	0.03	0.06	0.66
Percents*	0.04	0.00	0.15	0.00	0.03	0.03
<b>Sample Size</b>	<b>120</b>	<b>31</b>	<b>27</b>	<b>30</b>	<b>32</b>	

Source: Author tabulations using data from the spring 2007 teacher survey. The sample excludes one Math Expressions school (with 3 classrooms and 32 students) that participated during part of the school year and then stopped using its assigned curriculum and did not allow the study to collect follow-up data.

<sup>a</sup> Possible range from 0 (none), 1(1-5 lessons), 2 (6-10 lessons), 3 (11-15 lessons), to 4 (more than 15 lessons). A mean of 4 indicates that teachers covered at least 15 lessons in the content area.

\*Statistically significant at the 5 percent level. The statistical tests were conducted using two-level (classroom and school) HLMs with controls for classroom and school characteristics (see the text at the beginning of Chapter II for details). The *p*-values were not adjusted for the multiple outcomes (topics) tested.

The analyses were conducted using two different models. The main analysis did not control for instructional time. Earlier in this chapter, we saw that math instructional time differs across the curricula (Saxon teachers spent one more hour on math, per week, than the other groups), suggesting that the amount of time teachers can devote to math is not specified by at least some districts and/or schools. If math instructional time is not specified by the districts and/or schools, it could be affected by the curricula, in which case it should not be controlled for in the analysis. However, since this may not be the case in all districts/schools, we also looked at results using a model that controlled for instructional time (which assumes that math instructional time is set by a district or school independently of the curriculum). The results are robust across the analyses that did and did not include instructional time, and the results without a control for instructional time are presented in Table II.15.

Coverage in 8 of the 20 content areas is significantly different across the curriculum groups, including word problems, addition and subtraction facts with whole numbers, money, place value with whole numbers, fractions, probability, decimals, and percents (Table II.15). Given the small number of teachers in each curriculum group, however, these differences in content coverage should be interpreted with caution because of limited statistical power to detect differences. The study's follow-up report, which will be based on a sample nearly three times the size of the one examined in this report, will provide more precise information for curriculum differences in content coverage.

### **III. CURRICULUM EFFECTS ON FIRST GRADE ACHIEVEMENT**

The previous chapter presented several key findings from an analysis of curriculum implementation. Teachers received training from the publishers on their school's assigned curriculum, and 98 to 99 percent of teachers reported using it as their core curriculum in both a fall and spring survey. In the spring survey, 88 percent of teachers reported completing at least 80 percent of their assigned curriculum. Also, on average, Saxon Math (Saxon) teachers reported spending one more hour per week on math instruction than teachers in each of the other curriculum groups. The average number of lessons covered in 8 of the 20 math content areas examined differed by curriculum. However, because of the small number of teachers in each curriculum group examined, there is limited statistical power to assess which curricula differ from each other in terms of content coverage.

This chapter presents the relative effects of the curricula on first grade math achievement. The results are based on 4 districts with 39 schools, 131 first grade classrooms, and 1,309 students that participated in the study during the 2006-2007 school year (see Table I.3). Because students were divided into four curriculum groups (that is, there was no "control" group that, for example, continued to use the various curricula used by schools before joining the study), the effect of each curriculum is reported relative to the effect of each of the other three curricula. In particular, spring math achievement of students in each curriculum group is compared to spring achievement of students in each of the other three curriculum groups.

Before presenting effects on student math achievement, it is worth recalling the information that is and is not provided by the study. The relative effects of the curricula presented below reflect 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 the curricula, and implementation reflects what publishers and teachers achieved, not some level of implementation specified by the study. Also, the relative effects of the curricula are based only on the ECLS-K math assessment. Lastly, because the participating sites are not a representative sample of districts and schools, the design does not support making statements about effects for districts and schools outside of the study.

#### **A. METHODS USED TO CALCULATE CURRICULUM EFFECTS**

Results are based on a random sample of about 10 students in each of the 131 study classrooms who were tested in both the fall and spring (a "longitudinal" sample). Results were computed using a student-level weight that sums to the number of students in each classroom that was eligible for fall testing. For example, if 20 students in a classroom were eligible for testing and 10 students were sampled and tested, each tested student was assigned a weight of 2. The weight was not adjusted for the small fraction of students that were eligible for testing but

could not be tested (mainly because of parental nonconsent), because a nonresponse analysis showed that none of the available baseline characteristics were related to nonresponse.<sup>49</sup>

Valid estimates of relative curriculum effects can be calculated with the study's data, if random assignment achieved its objective of creating curriculum groups with similar baseline characteristics. If this objective was achieved, differences in outcomes of the curriculum groups can be attributed to differences in curriculum usage—that is, causal statements can be made about curriculum effects.

Table III.1 shows that random assignment created curriculum groups with similar school characteristics, as expected. School-wide Title I eligibility, free/reduced-price meals eligibility, first- and second-grade enrollments, student gender, and student race/ethnicity are not significantly different across the curriculum groups. These results were expected, because (as described in Chapter I) a blocked random assignment procedure was used to allocate the curricula to schools.<sup>50</sup>

Although the study team randomly assigned curricula to schools, it did not randomly assign teachers to schools, nor students to teachers; nevertheless, all but one teacher characteristic and all student characteristics examined are not significantly different across the curriculum groups. Table II.1 (in Chapter II) shows that nearly all measures of teacher demographics, education, experience, and scores on the teacher assessment administered by the study team are not significantly different across the curriculum groups, with one exception. At least 93 percent of Investigations in Number, Data, and Space (Investigations), Math Expressions, and Saxon teachers classified themselves as white, whereas 78 percent of Scott Foresman-Addison Wesley Mathematics (SFAW) teachers did so.<sup>51</sup> Statistical tests indicate that these racial differences across the curriculum groups are statistically significant. As described below, the approach for calculating curriculum effects adjusts for teacher race.

---

<sup>49</sup> Curriculum effects also were examined for a sample of students who were in a study school during spring testing, whether or not they were in one of the schools during fall testing (a “cross-sectional” sample). To support this analysis, students who enrolled in a study school after fall testing also were tested in the spring—an average of one student per classroom. These students were added to the longitudinal sample to create a sample that was representative of the students enrolled in the classrooms during spring testing. Results based on this sample help us understand the effects of the curricula along a measure (achievement of all students in the spring) often used to judge school performance, such as Title I Adequate Yearly Progress. Results based on the cross-sectional sample are reported in Appendix D, and the main conclusions based on these results are similar to those based on the longitudinal sample.

<sup>50</sup> With a large sample of schools, a straightforward random assignment procedure that simply assigns curricula to schools in each district (that is, without creating blocks that contain similar schools and conducting random assignment within the blocks) would produce curriculum groups with similar characteristics. With the relatively small number of schools (about 10) assigned to each curriculum in the current sample, the straightforward procedure could result in chance differences between curriculum groups. The blocked random assignment procedure used by the study helps to minimize this possibility.

<sup>51</sup> Teachers were asked whether they are Hispanic or Latino, but this characteristic was not examined or included in the analysis of curriculum effects below because the number of teachers that reported being Hispanic or Latino was too small to support the analysis.

TABLE III.1

## BASELINE CHARACTERISTICS OF COHORT-ONE SCHOOLS BY CURRICULUM

	Schools by Curriculum					<i>p</i> -value
	All Schools	Investigations	Math Expressions	Saxon	SFAW	
Title I Eligible (percentage)	74.4	80.0	66.7	88.9	63.6	0.16
School-Wide Title I Eligible (percentage)	53.8	50.0	44.4	55.5	63.6	0.14
Students Eligible for Free/Reduced-Price Meals (percentage)	68.7	66.9	65.0	68.1	73.5	0.23
Student Enrollment (average)						
First grade	73	76	73	74	70	0.56
Second grade	71	74	71	74	64	0.24
Student Gender (percentage)						
Male	51.5	52.3	50.6	52.4	50.6	0.22
Female	48.5	47.7	49.4	47.6	49.4	0.22
Student Race/Ethnicity (percentage)						
White	46.2	47.4	48.3	51.0	39.4	0.45
Black	26.1	27.2	26.1	18.4	31.4	0.74
Hispanic	19.9	19.0	18.3	26.2	16.9	0.31
Asian	3.8	5.4	5.6	2.6	2.0	0.37
American Indian/Alaskan Native	4.0	1.0	1.7	1.8	10.3	0.37
<b>Sample Size</b>	<b>39</b>	<b>10</b>	<b>9</b>	<b>9</b>	<b>11</b>	

Source: Author tabulations using the 2003-2004 Common Core of Data (CCD). When free/reduced-price data were missing in the CCD, data were obtained from [www.GreatSchools.net](http://www.GreatSchools.net). The sample excludes 1 Math Expressions school (with 3 classrooms and 32 students) that participated during part of the school year and then stopped using the curriculum and did not allow the study to collect follow-up data.

Note: The *p*-values are results from statistical tests that examine the joint equality of each school characteristic across the curriculum groups. The statistical tests were conducted using regression models. The model regressed each school characteristic on an intercept, binary indicators for three of the four curricula, binary indicators for all but one of the blocks to which the schools were assigned during random assignment, and an error term. By including indicators for the blocks, the degrees of freedom used to calculate the statistical significance of the results are adjusted to reflect the information (number of blocks constructed) during used when conducting random assignment.

Table III.2 shows no significant differences across curriculum groups along all of the student characteristics that were collected. The characteristics include student fall math test scores, age at fall test, gender, race/ethnicity, limited English proficiency (LEP) or English language learner (ELL), individualized education plan (IEP) or receipt of special services for students with a disability, and number of days between the fall and spring test.

TABLE III.2  
BASELINE CHARACTERISTICS OF COHORT-ONE LONGITUDINAL STUDENTS,  
BY CURRICULUM

	Students by Curriculum					<i>p</i> -value
	All Students	Investigations	Math Expressions	Saxon	SFAW	
Fall Score (average)	31.0	32.2	29.9	31.1	30.9	0.20
Age at Fall Test (average)	6.5	6.5	6.4	6.5	6.4	0.93
Female (percentage)	49.2	50.9	47.8	50.6	47.7	0.61
Race/Ethnicity (percentage) <sup>a</sup>						
Hispanic	20.5	16.6	17.5	18.3	29.3	0.11
Black non-Hispanic	19.3	19.3	25.0	22.7	10.5	
Other non-Hispanic	60.2	64.1	57.5	59.0	60.2	
LEP or ELL (percentage)	13.4	10.9	11.5	12.4	18.5	0.67
IEP/special services (percent)	5.6	6.2	8.1	5.2	3.2	0.14
Days Between Fall and Spring Test (average)	239	236	242	238	238	0.48
<b>Sample Size</b>	<b>1,309</b>	<b>332</b>	<b>314</b>	<b>304</b>	<b>359</b>	

Source: Author tabulations using data from the fall first grade ECLS-K math test administered by the study, and school records. The sample excludes 1 Math Expressions school (with 3 classrooms and 32 students) that participated during part of the school year and then stopped using the curriculum and did not allow the study to collect follow-up data.

Note: The *p*-values are results from statistical tests that examine the joint equality of each student characteristic across the curriculum groups. The statistical tests were conducted using three-level hierarchical linear models (HLM). The first (student-level) equation regressed each student characteristic on an intercept and a student-level error term. The second (classroom-level) equation regressed the intercept from the first equation on a classroom-level intercept and error term. The third (school-level) equation regressed the intercept from the second equation on a school-level intercept, binary indicators for three of the four curricula, binary indicators for all but one of the blocks to which the schools were assigned during random assignment, and a school-level error term. By including indicators for the blocks, the degrees of freedom used to calculate the statistical significance of the results are adjusted to reflect the information (number of blocks constructed) during used when conducting random assignment. HLMs that are appropriate for continuous, binary, and categorical variables were used accordingly. A single *p*-value is reported for binary and multinomial variables, and indicates whether the fraction of students in each category of the variable differs across the curriculum groups.

<sup>a</sup>Students classified as Hispanic on school records were coded as Hispanic regardless of race. Non-Hispanic students classified as Black, or Black and other races, were coded as Black non-Hispanic. All other students were coded as Other non-Hispanic.

Hierarchical linear model (HLM) techniques were used to calculate the relative effects of the curricula on student math achievement—that is, effects on the spring Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K) math scale score. This technique incorporates the nested structure of the data, which includes students clustered in classrooms and classrooms clustered in schools, when calculating the statistical significance of the results. Clustering tends to reduce the precision of the results because outcomes of students within the same classroom and within the same school often are similar. Baseline measures of several characteristics related to student achievement were included in the HLM to increase the precision of the results, thereby helping to offset the precision losses from clustering:

- **7 Student Characteristics:** fall ECLS-K math scale score, age at fall test, days between the fall and spring test, gender, race/ethnicity, LEP/ELL, and IEP/special services
- **8 Teacher/Classroom Characteristics:** teacher race; education; experience; prior use of the assigned curriculum at the K-3 level; score on the math content/pedagogical test administered before curriculum training; and three classroom characteristics that may affect teacher instruction—class size, variance of the fall student math score, and skewness of the score<sup>52</sup>
- **3 School Characteristics:** Title I eligible, percentage of students eligible for free/reduced-price meals, and curriculum assignment<sup>53</sup>

Appendix D presents the variables included in the HLM, data sources for the variables, and details related to model estimation. The appendix also presents average unadjusted fall and spring math achievement of students in each curriculum group, and the average gain (spring minus fall) score for each group (see Table D.3).<sup>54</sup>

---

<sup>52</sup> A classroom-level measure of the variance of the fall student math score was included in the HLM to account for the heterogeneity of students in each class, and a classroom-level measure of the skewness of the score was included to account for the types of students (lower or higher achievers) that primarily comprise each class.

<sup>53</sup> As mentioned in Chapter I, random assignment of curricula was conducted within blocks of schools. The degrees of freedom used to calculate the statistical significance of the results were adjusted to reflect the information (number of blocks constructed) used when conducting random assignment. Operationally, this was accomplished by including in the school equation of the HLM the block to which each school was assigned.

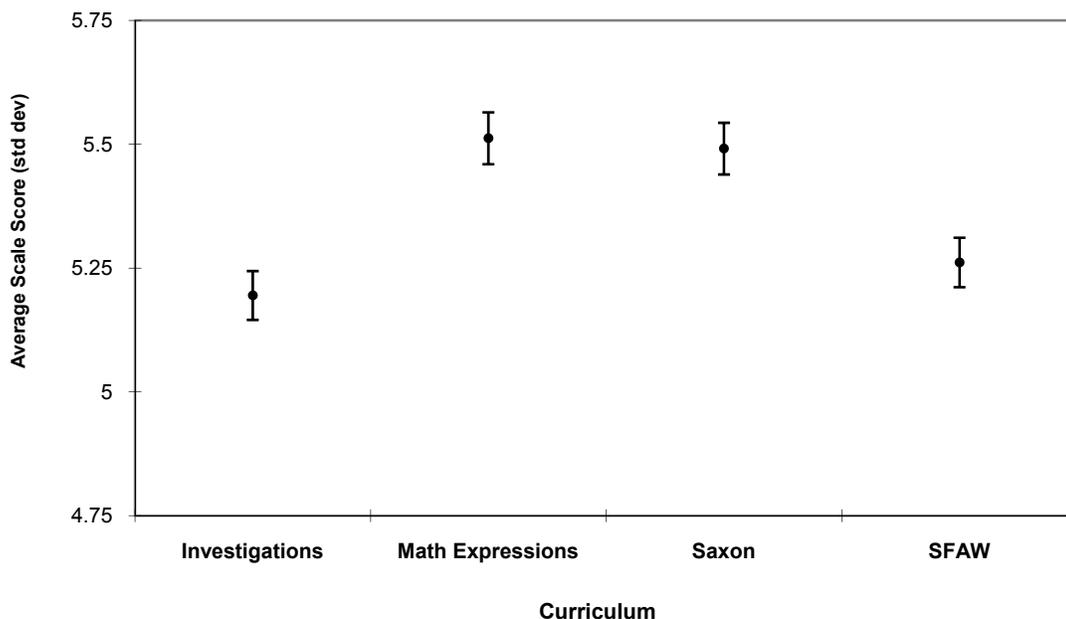
<sup>54</sup> Technically, only the outcome—the average spring score—of the four curriculum groups is needed to calculate relative effects. That is, using HLM techniques to adjust the spring score for the fall score and other baseline characteristics is not needed. However, adjusting for the fall score, in particular, helps to significantly improve the precision of the results, because the fall score accounts for a significant amount of variation in the spring score. In fact, the sample size needed to detect the study’s target effect size was calculated under the assumption that the fall score would be used in the analysis. Adjusting for the fall score also adjusts for random differences in starting points that can exist across curriculum groups when assigning a relatively small number of schools, and that could affect results if not accounted for. For example, the possibility exists that some curriculum groups are, by chance, assigned schools with higher fall scores than the other groups. These random differences in fall scores could persist in spring scores, which could lead to false conclusions about relative curriculum effects if only spring scores are used to calculate effects. The relative effects of the curricula described below are similar

## B. RELATIVE EFFECTS OF THE CURRICULA

Results based on the HLM are summarized in Figure III.1. It 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 student, teacher, and school characteristics listed above. The bars that extend from each dot represent the 95 percent confidence interval around each average score. Curricula with non-overlapping confidence intervals have average scores that are significantly different at the 5 percent level of confidence.<sup>55</sup> The results are presented in fractions of a standard deviation, which means that subtracting the average values for any two curricula indicates the effect size of using the first curriculum instead of the second.

FIGURE III.1

AVERAGE HLM-ADJUSTED SPRING MATH SCORE WITH CONFIDENCE INTERVAL,  
BY CURRICULUM (in Standard Deviations)



Note: The dots in each symbol represent the average HLM-adjusted spring math score (in standard deviations) 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 of confidence.

---

*(continued)*

when based on the simple averages, though the confidence intervals are wider than those on the HLM-adjusted averages, as expected.

<sup>55</sup> The 5 percent level of confidence means there is no more than a 5 percent chance that any finding discussed could have occurred by chance.

Table III.3 presents the magnitude of the results, in effect sizes, for each unique pair-wise curriculum comparison that can be made.<sup>56</sup> Effect sizes were calculated by dividing each pair-wise curriculum comparison by the pooled standard deviation of the spring score for the two curricula being compared, and Hedges' *g* formula (with the correction for small-sample bias) was used to calculate the pooled standard deviations. The table also presents the *p*-value for each result, and only results with *p*-values less than or equal to 0.05 are considered statistically significant and discussed below. The Tukey-Kramer method was used to adjust the statistical significance calculations for the six unique pair-wise curriculum comparisons that can be made (Tukey 1952, 1953; Kramer 1956).<sup>57</sup>

TABLE III.3  
AVERAGE DIFFERENCE BETWEEN PAIRS OF CURRICULA IN HLM-ADJUSTED SPRING STUDENT  
MATH ACHIEVEMENT, IN EFFECT SIZES  
(*p*-values Are in Parentheses)

	Effect of					
	Investigations Relative to			Math Expressions Relative to		Saxon Relative to
	Math Expressions	Saxon	SFAW	Saxon	SFAW	SFAW
Effect Size	-0.30*	-0.30*	-0.07	0.02	0.24*	0.24*
<i>p</i> -value	(0.00)	(0.00)	(0.80)	(0.99)	(0.02)	(0.05)

Source: Author tabulations using data from the spring first grade ECLS-K math test administered by the study, school records, fall 2006 teacher survey, and school-level data from the 2003-2004 Common Core of Data and www.GreatSchools.net. The sample excludes 1 Math Expressions school (with 3 classrooms and 32 students) that participated during part of the school year and then stopped using the curriculum and did not allow the study to collect follow-up data.

Note: Effect sizes were calculated by dividing each pair-wise curriculum comparison by the pooled standard deviation of the spring scale score for the two curricula being compared, and Hedges' *g* formula (with the correction for small-sample bias) was used to calculate the pooled standard deviations. The results were produced using a three-level hierarchical linear model (see Appendix D for details about the model). The Tukey-Kramer method was used to adjust the *p*-values for the six unique pair-wise curriculum comparisons that can be made.

\*Statistically significant at the 5 percent level.

<sup>56</sup> Results are reported only for the six unique pair-wise curriculum comparisons that can be made. For example, the table reports the difference in adjusted spring achievement between Investigations and Math Expressions, but not the opposite comparison (the difference between Math Expressions and Investigations) because the latter comparison equals the same magnitude as the former with the opposite sign.

<sup>57</sup> Appendix D describes the Tukey-Kramer method.

## **1. Student Math Achievement Was Significantly Higher in Math Expressions and Saxon Schools than in Investigations and SFAW Schools**

As the results in Table III.3 show, average 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. For a student at the 50th percentile in math achievement, these effect sizes mean that the student's percentile rank would be 9 to 12 points higher if the school used Math Expressions or Saxon, instead of Investigations or SFAW.

The results in Table III.3 also show that math achievement in schools assigned to the two more effective curricula (Math Expressions and Saxon) was not significantly different, nor was math achievement in schools assigned to the two less effective curricula (Investigations and SFAW). Student achievement differences of the two more effective curricula (Math Expressions and Saxon) equal 0.02 standard deviations and are not statistically significant. Similarly, student achievement differences of the two less effective curricula (Investigations and SFAW) equal 0.07 standard deviations and are not statistically significant.<sup>58</sup>

An important issue to consider is how the relative effects of Math Expressions and Saxon compare to the relative effects of other commonly used curricula not included in this study. Unfortunately, it is difficult to make such an assessment because of differences between this study's design and the designs of other curriculum studies.<sup>59</sup>

We can, however, consider other educational interventions research has shown to be effective, such as reducing class size, and Math Expressions' and Saxon's effects are at least as large (if not larger) than the effect of reducing first grade class sizes. Tennessee's Project STAR (Student-Teacher Achievement Ratio) is considered by many to be one of the few large-scale experimental studies in education with positive effects. The study compared student achievement of small classes (13-17 students), regular-sized classes (22-25 students), and regular-sized classes with both a teacher and teacher's aide. The effect size on first-grade math achievement of reducing class size from regular-sized to small ranged from 0.13 to 0.27 (Finn and Achilles

---

<sup>58</sup> We explored whether the results are sensitive to (1) the specification of the HLM used to estimate effects, (2) the one (Math Expressions) school that stopped using the curriculum and did not allow spring testing of students and, therefore, had to be excluded from the analysis, and (3) the few students that moved between study schools that used a different study curriculum. The results are robust to these sensitivity analyses—see Appendix D for more details.

<sup>59</sup> The What Works Clearinghouse (WWC) (2006) identified three other curricula not included in this study—Everyday Math, Houghton Mifflin Math, and Progress in Math 2006—that have been studied using methods that meet the WWC's evidence standards. The WWC concluded that Everyday Math has potentially positive effects on student math achievement with effect sizes ranging from  $-0.17$  to  $0.37$  (Carroll 1998; Waite 2001; Woodward and Baxter 1997; Riordan and Noyce 2001), whereas Houghton Mifflin Math and Progress in Math 2006 have no discernible effects (EDSTAR, Inc. 2004; Beck Evaluation & Testing Associates 2005). Direct comparisons of these results with the results for Math Expressions and Saxon (the two more effective curricula in this study) are difficult to make because the grade levels examined in the Everyday Math and Houghton Mifflin Math studies differ (the same grade level was examined in the Progress in Math 2006 study), and it is difficult to assess if the curriculum materials used and instructional practices of the comparison groups in these other studies are similar to the Investigations and SFAW comparisons made in this study.

1990).<sup>60</sup> As mentioned above, the effect sizes for Math Expressions and Saxon ranged from 0.24 to 0.30.

## 2. Some Curriculum Differentials Also Exist in Several Subgroups

The settings in which the curricula were used in this study vary, as they may when used in schools throughout the country. For example, although the study team’s goal was to recruit schools with students struggling in math, participating schools contain a range of low student math achievement. The effects of the curricula may differ among schools with lower and higher math achievement. Curriculum effects also may differ along other important characteristics that differentiate instructional settings.

To help educators understand the relative effects of the curricula in different school environments, we examined whether curriculum effects differ along six characteristics:

1. ***Participating Districts.*** We examined results for students in each of the four districts.
2. ***School Fall Achievement.*** We examined results for students in schools with average fall math scores in the lowest, middle, and highest third of the school-level score distribution.<sup>61</sup> Research indicates that math achievement in the earliest elementary grades is associated with achievement in the later elementary grades (Princiotta, Flanagan, and Hausken 2006). For example, the research indicates that students who scored in the lowest third in the fall of their kindergarten year scored lower than other students by the spring of fifth grade.
3. ***School Free/Reduced-Price Meals eligibility.*** We examined results for students in schools with up to 40 percent meals eligibility, and those with more than 40 percent eligibility. Schools that serve free or reduced-price lunches to more than 40 percent of their students qualify for higher “severe need” reimbursements.
4. ***Teacher Education.*** We examined results for students who have teachers with and without a master’s degree. All the teachers in our sample that do not have a master’s degree have a bachelor’s degree.

---

<sup>60</sup> Studies that reanalyzed data from Project STAR showed small class size benefits of 0.30 standard deviations (see Nye, Hedges, and Konstantopoulos 2000) or eight percentile points (see Krueger 1999) for first grade mathematics achievement. These more recent studies used HLM techniques to address clustering of students within schools and classrooms, examined actual class sizes which sometimes varied from intended class sizes assigned by the experiment, and examined effects for students who experienced small class sizes for more than one year.

<sup>61</sup> Curricula are typically implemented school wide, or at least grade/classroom wide. In other words, different curricula typically are not used with different students within a grade or classroom. Therefore, we created subgroups based on school- and teacher-level measures (of school, teachers, and student characteristics) because we suspect results based on these subgroups will be useful to educators undergoing a curriculum adoption decision. For example, effects based on schools with different average student achievement are likely to be more useful than effects based on individual students with different achievement because curriculum decisions are typically not made based on individual student achievement.

5. ***Teacher Experience.*** We examined results for students who have teachers with up to five years of experience, and with more than five years of experience. Research indicates that a significant portion of teachers leave the profession within five years of entering (Ingersoll 2002).
6. ***Teacher Math Content/Pedagogical Knowledge.*** We examined results for students who have teachers with scores in the first (lowest) quintile, and those with scores in the second through fifth quintiles. Research indicates that student achievement of first-grade teachers with scores in the lowest quintile is lower than student achievement among teachers with higher scores (Hill, Rowan, and Ball 2005). By examining these subgroups, we examined whether the relative effects of the curricula depend on teacher math knowledge for teaching.

Subgroup effects were estimated by including in the HLM described above interactions between the curriculum indicators and the characteristics. Separate HLMs were specified for each characteristic.<sup>62</sup> For example, to investigate the moderating effect of school fall achievement, we added variables to the HLM that interact the curriculum indicators with an indicator for students in schools with higher fall scores. Appendix D describes the model specifications in more detail, presents sample sizes for each subgroup, and the minimum detectable effect size for each subgroup.

Figures III.2. through III.7 report average adjusted spring achievement (in standard deviations) of the subgroups by curriculum. Subtracting the values for any two curricula within a subgroup indicates the effect size of using the first curriculum instead of the second.

Ignoring (for a moment) the statistical significance of curriculum differences, the pattern of results in most of the subgroups is consistent with the pattern observed for students overall. That is, in most subgroups, Math Expressions and Saxon students have higher average adjusted spring scores than Investigations and SFAW students. Moreover, there are no subgroups in which Investigations or SFAW students have higher average adjusted spring scores than Math Expressions or Saxon students.

Table III.4 reports the relative curriculum effects for each subgroup and the statistical significance of the results. The Tukey-Kramer method was used to adjust the  $p$ -values for the curriculum comparisons made within each characteristic, but the values were not adjusted for all the comparisons that can be made across all the subgroups. For example, the  $p$ -values for the district-level results were adjusted for the 24 curriculum comparisons made across the districts (that is, 6 curriculum comparisons were made in each of the 4 districts), but not for all 90 comparisons that can be made in the table. We did not adjust for all comparisons that can be made because the study was not designed to have sufficient statistical power for the subgroup analyses. Therefore, these results are best viewed as exploratory analyses that could raise policy-relevant questions that could be examined by other studies that are designed to have sufficient statistical power to address the questions.

---

<sup>62</sup> Interactions between curriculum indicators and teacher characteristics are cross-level interactions.

FIGURE III.2

AVERAGE HLM-ADJUSTED SPRING MATH SCORE IN STANDARD DEVIATIONS,  
BY DISTRICT AND CURRICULUM

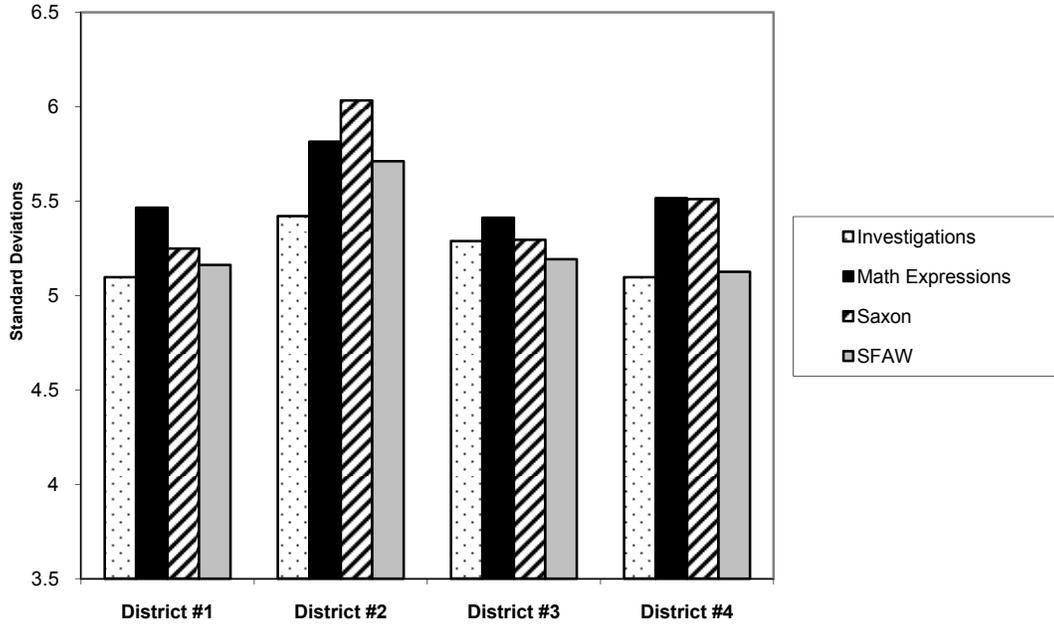


FIGURE III.3

AVERAGE HLM-ADJUSTED SPRING MATH SCORE IN STANDARD DEVIATIONS,  
BY SCHOOL FALL ACHIEVEMENT AND CURRICULUM

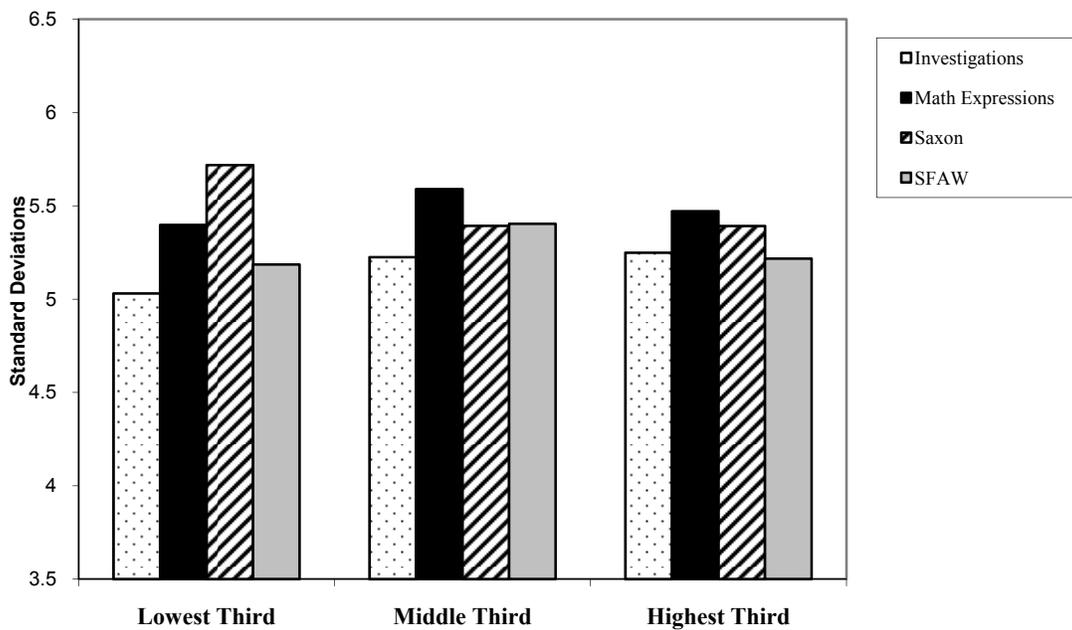


FIGURE III.4

AVERAGE HLM-ADJUSTED SPRING MATH SCORE IN STANDARD DEVIATIONS,  
BY SCHOOL FREE/REDUCED-PRICE MEALS ELIGIBILITY AND CURRICULUM

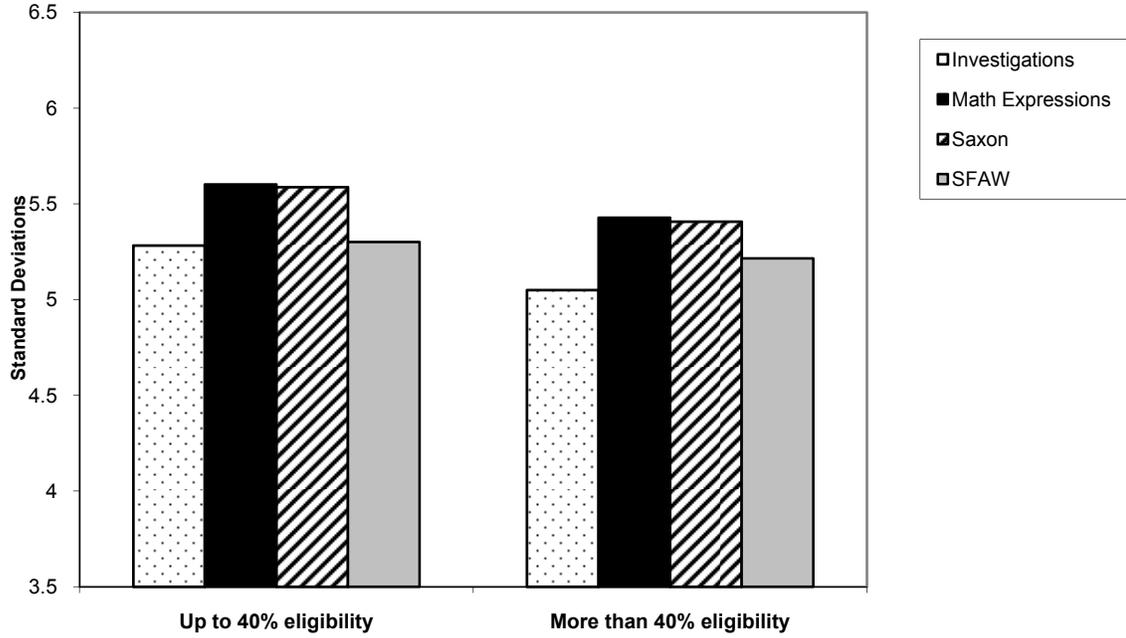


FIGURE III.5

AVERAGE HLM-ADJUSTED SPRING MATH SCORE IN STANDARD DEVIATIONS,  
BY TEACHER EDUCATION AND CURRICULUM

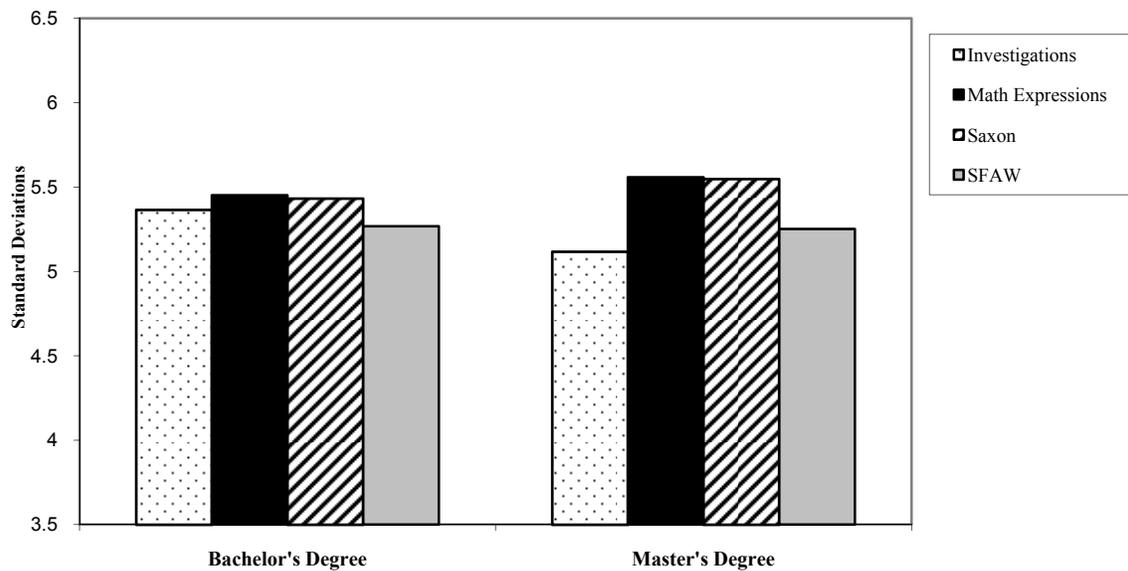


FIGURE III.6

AVERAGE HLM-ADJUSTED SPRING MATH SCORE IN STANDARD DEVIATIONS,  
BY TEACHER EXPERIENCE AND CURRICULUM

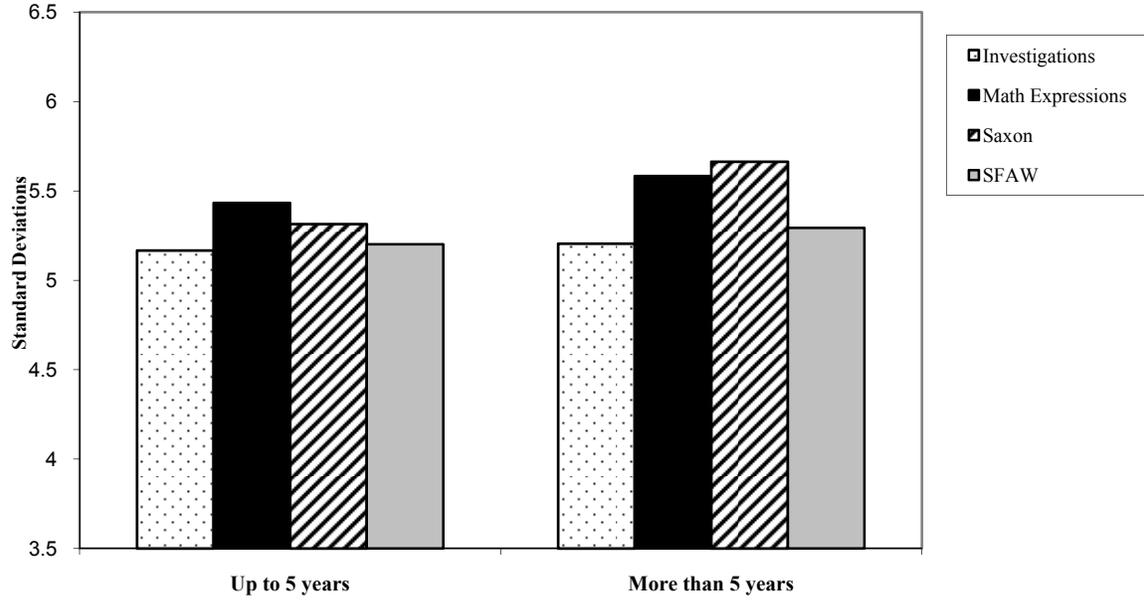


FIGURE III.7

AVERAGE HLM-ADJUSTED SPRING MATH SCORE IN STANDARD DEVIATIONS,  
BY TEACHER MATH CONTENT/PEDAGOGICAL TEST SCORE AND CURRICULUM

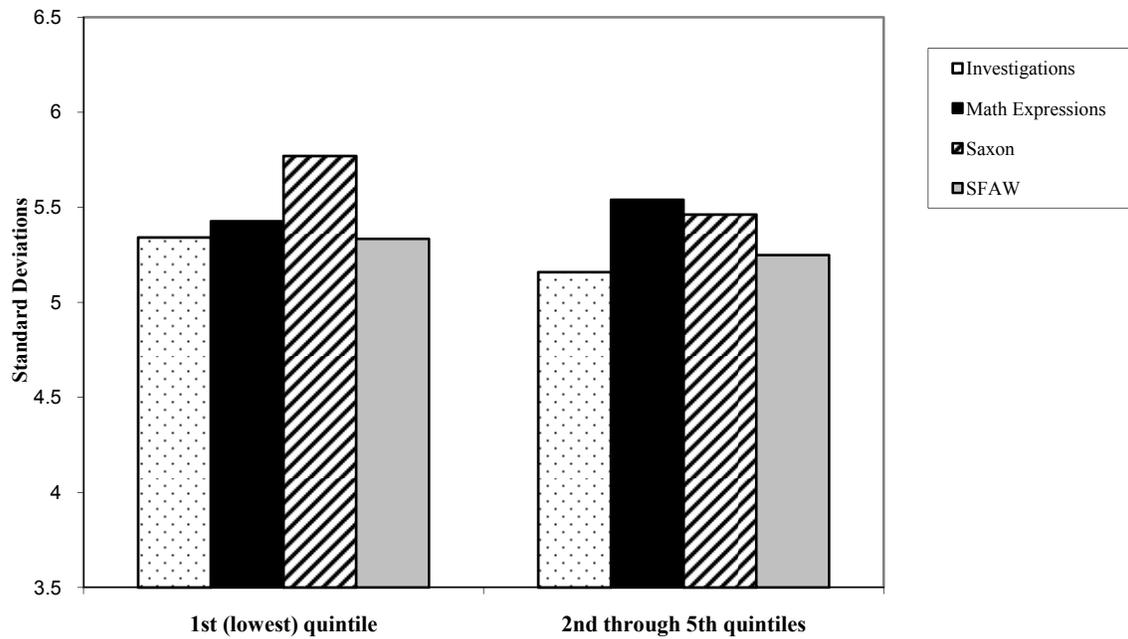


TABLE III.4

AVERAGE DIFFERENCE BETWEEN PAIRS OF CURRICULA IN HLM-ADJUSTED SPRING STUDENT  
MATH ACHIEVEMENT, BY SUBGROUPS AND IN EFFECT SIZES  
(*p*-values Are in Parentheses)

	Effect of					
	Investigations Relative to			Math Expressions Relative to		Saxon Relative to
	Math Expressions	Saxon	SFAW	Saxon	SFAW	SFAW
<b>Participating Districts</b>						
District #1	-0.35 (0.90)	-0.16 (0.99)	-0.06 (1.00)	0.22 (1.00)	0.30 (0.98)	0.09 (1.00)
District #2	-0.38 (0.37)	-0.63* (0.01)	-0.29 (0.78)	-0.22 (0.91)	0.10 (1.00)	0.34 (0.58)
District #3	-0.12 (1.00)	-0.01 (1.00)	0.10 (1.00)	0.12 (1.00)	0.21 (0.85)	0.11 (1.00)
District #4	-0.40* (0.01)	-0.43* (0.03)	-0.03 (1.00)	0.00 (1.00)	0.38* (0.02)	0.41 (0.15)
<b>School Fall Achievement</b>						
Lowest third	-0.35 (0.28)	-0.71* (0.00)	-0.15 (0.99)	-0.32 (0.29)	0.21 (0.83)	0.56* (0.01)
Middle third	-0.35 (0.15)	-0.17 (0.99)	-0.18 (0.86)	0.20 (0.81)	0.18 (0.87)	-0.01 (1.00)
Highest third	-0.21 (0.60)	-0.15 (0.91)	0.03 (1.00)	0.08 (1.00)	0.25 (0.46)	0.18 (0.92)
<b>School Free/Reduced-Price Meals Eligibility</b>						
Up to 40% eligibility	-0.30* (0.05)	-0.31 (0.08)	-0.02 (1.00)	0.01 (1.00)	0.29 (0.13)	0.30 (0.25)
Greater than 40% eligibility	-0.36* (0.03)	-0.37 (0.07)	-0.16 (0.82)	0.02 (1.00)	0.21 (0.44)	0.20 (0.67)
<b>Teacher Education</b>						
Less than master's degree	-0.08 (1.00)	-0.07 (1.00)	0.10 (0.99)	0.02 (1.00)	0.18 (0.72)	0.17 (0.85)
Master's degree or more	-0.42* (0.00)	-0.44* (0.00)	-0.13 (0.79)	0.01 (1.00)	0.30 (0.07)	0.31 (0.13)
<b>Teacher Experience</b>						
Up to 5 years	-0.25 (0.34)	-0.15 (0.90)	-0.03 (1.00)	0.12 (0.92)	0.23 (0.46)	0.12 (0.97)
More than 5 years	-0.36* (0.00)	-0.47* (0.00)	-0.09 (0.95)	-0.08 (0.98)	0.28* (0.04)	0.39* (0.01)

TABLE III.4 (continued)

	Effect of					
	Investigations Relative to			Math Expressions Relative to		Saxon Relative to
	Math Expressions	Saxon	SFAW	Saxon	SFAW	SFAW
<b>Teacher Math Content/Pedagogical Knowledge</b>						
First (lowest) quintile	-0.08 (1.00)	-0.44 (0.25)	0.01 (1.00)	-0.35 (0.44)	0.09 (1.00)	0.46 (0.17)
2nd through 5th quintiles	-0.36* (0.00)	-0.31* (0.02)	-0.09 (0.94)	0.08 (0.97)	0.28* (0.04)	0.22 (0.33)

Source: Author tabulations using data from the first-grade ECLS-K math tests administered by the study, school record, fall 2006 teacher survey, and school-level data from the 2003-04 Common Core of Data and www.GreatSchools.net. The sample excludes 1 Math Expressions school (with 3 classrooms and 32 students) that participated during part of the school year and then stopped using the curriculum and did not allow the study to collect follow-up data.

Note: The results were produced using a three-level hierarchical linear model (see Appendix D for details about the model). The Tukey-Kramer method was used to adjust the  $p$ -values for the six unique pair-wise curriculum comparisons that can be made.

\*Statistically significant at the 5 percent level.

Two of the four curriculum differentials that are statistically significant for students overall also are significant for several subgroups. Student math achievement was significantly higher in Math Expressions and Saxon schools than in Investigations schools in District #4, and in classes taught by teachers with master's degrees, by teachers with more than five years of experience, and by teachers with scores on the teacher test of math content and pedagogical knowledge that fall in the second through fifth quintiles of the score distribution. In District #4, the Math Expressions-SFAW differential also is positive and statistically significant and, in classes taught by teachers with more than five years of experience, the Math Expressions-SFAW and Saxon-SFAW differentials also are positive and statistically significant. The Math Expressions-SFAW differential also is positive and statistically significant in classes taught by teachers with scores on the teacher test that fall in the second through fifth quintiles of the score distribution.

One curriculum differential also is statistically significant in each of four subgroups not yet mentioned. In both subgroups based on school free/reduced-price meals eligibility, the Math Expressions-Investigations differential is positive and statistically significant. In District #2 and in schools with average fall math scores in the lowest third, the Saxon-Investigations differential is positive and statistically significant. The Saxon-SFAW differential also is positive and statistically significant in schools with average fall math scores in the lowest third.

### **C. NEXT STEPS FOR THE STUDY**

The study's follow-up report (described in Chapter I) will provide a more comprehensive look at the relative effects of the curricula. Some of the schools that will be added to the analysis in the follow-up report have classrooms taught entirely in Spanish and, therefore, used Spanish curriculum materials and were tested by the study team using Spanish-speaking testers—something that did not occur in the schools examined in this report. By adding these schools to the analysis, the follow-up report will provide broader information about the relative effects of the curricula. The follow-up report also will provide a more comprehensive look at effects for subgroups.

## REFERENCES

- Adelman, C. "Answers in the Toolbox: Academic Intensity, Attendance Patterns, and Bachelor's Degree Attainment." Washington, DC: U.S. Department of Education, 1999.
- Agodini, Roberto, John Deke, Sally Atkins-Burnett, Barbara Harris, and Robert Murphy. "Design for the Evaluation of Early Elementary School Mathematics Curricula." Report submitted to the U.S. Department of Education, Institute of Education Sciences. Princeton, NJ: Mathematica Policy Research, Inc., January 2008.
- Beck Evaluation & Testing Associates, Inc. "Progress in Mathematics<sup>®</sup> 2006: Grade 1 pre-post field test evaluation study." New York: Sadlier-Oxford Division, William H. Sadlier, Inc., 2005.
- Bruner, Jerome. *The Process of Education*. Cambridge, MA: Harvard University Press, 1960.
- Carnevale, A.P., and D.M. Desrochers. "Standards for What? The Economic Roots of K-16 Reform." Washington, DC: Educational Testing Service, 2003.
- Carroll, W.M. "Geometric knowledge of middle school students in a reform-based mathematics curriculum." *School Science and Mathematics*, vol. 98, no. 4, 1998, pp. 188-197.
- 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, 2005.
- Education Market Research. *Mathematics Market, Grades K-12, 2008: Teaching Methods, Textbooks/Materials Used and Needed, and Market Size*. Rockaway Park, NY: EMR, 2008.
- EDSTAR, Inc. "Large-scale evaluation of student achievement in districts using Houghton Mifflin." Raleigh-Durham, NC: Author, 2004.
- Erchul, William P., and Brian K. Martens. *School Consultation: Conceptual and Empirical Bases of Practice: Second Edition*. New York, NY: Springer, 2002.
- Finn, J.D., and C.M. Achilles. "Answers and Questions About Class Size: A Statewide Experiment." *American Educational Research Journal*, vol. 27, no. 6, 1990, pp. 557-577.
- Fuson, Karen C. *Math Expressions. Grade 1*. Boston, MA: Houghton Mifflin Company, 2006a.
- Fuson, Karen C. *Math Expressions Teacher's Guide Grade 1 Vol. 1*. Boston: Houghton Mifflin Company, 2006b.
- Fuson, Karen C. *Math Expressions Teacher's Guide Grade 1 Vol. 2*. Boston: Houghton Mifflin Company, 2006c.

- Ginsburg, A., G. Cooke, S. Leinwand, J. Noell, and E. Pollock. "Reassessing U.S. international mathematics performance: New findings from the 2003 TIMSS and PISA." Washington, DC: American Institutes for Research, 2005.
- Hill, Heather, Stephen G. Schilling, and Deborah Loewenberg Ball. "Developing Measures of Teachers' Mathematics Knowledge for Teaching." *The Elementary School Journal*, vol. 105, no. 1, 2004, pp. 11-30.
- Hill, Heather, Brian Rowan, and Deborah Loewenberg Ball. "Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement." *American Educational Research Journal*, vol. 42, no. 2, 2005, pp. 371-406.
- Ingersoll, Richard M. "Holes in the Teacher Supply Bucket." *The School Administrator*, vol. March 2002, p. 42.
- Institute of Education Sciences. "Final Report on the National Assessment of Title I: Summary of Key Findings." Washington, DC: U.S. Department of Education, National Center for Education Evaluation and Regional Assistance, 2007.
- Klein, David. "A Brief History of American K-12 Mathematics Education in the 20th Century" in James Royer (ed) *Mathematical Cognition: A Volume in Current Perspectives on Cognition, Learning, and Instruction*. Information Age Publishing, 2003, pp.175-225. <http://www.csun.edu/~vcnth00m/AHistory.html>. Accessed September 4, 2008.
- Klein, David. "A Quarter Century of U.S. 'Math Wars' and Political Partisanship." *Journal of the British Society for the History of Mathematics*, vol. 22, no. 1, 2007, pp. 22-33.
- Kliman, Marlene, Susan Jo Russell, Tracey Wright, and Jan Mokros. *Investigations in Number, Data, and Space, Introduction: Mathematical Thinking at Grade 1*. Glenview IL: Pearson Scott Foresman, 2006.
- Krueger, Alan B. "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics*, vol. 114, no. 2, 1999, pp. 497-532.
- Kramer C.Y. "Extension of Multiple Range Tests to Group Means with Unequal Numbers of Replications." *Biometrika*, vol. 12, 1956, pp. 307-310.
- Larson, Nancy. *Saxon Math 1*. Austin, TX: Harcourt Achieve, 2004.
- Larson, Nancy and Saxon Publishers. *Saxon Math 1 Lesson Sampler*. Austin, TX: Harcourt Achieve, 2006.
- Lee, J., W. Grigg, and G. Dion. *The Nation's Report Card: Mathematics 2007*. Publication No. NCES 2007-494. Washington, D.C: U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences, 2007.
- Lord, F.M. *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum Publishers, 1980.

- National Assessment Governing Board. *Mathematics Framework for the 1996 National Assessment of Educational Progress*. Washington, DC: Government Printing Office, 1996.
- National Council of Teachers of Mathematics (NCTM). *Curriculum Focal Points for Prekindergarten through Grade 8 Mathematics: A Quest for Coherence*. Reston, VA: NCTM, 2006.
- National Mathematics Advisory Panel. *Foundations for Success: The Final Report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education, 2008a.
- National Mathematics Advisory Panel. *Foundations for Success: Reports of the Task Groups and Subcommittees of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education, 2008b.
- National Research Council. *On Evaluating Curricular Effectiveness: Judging the Quality of K-12 Mathematics Evaluations*. Washington, DC: National Academies Press, 2004.
- National Research Council. *Adding it Up: Helping Children Learn Mathematics*. Washington, DC: National Academies Press, 2001.
- National Science Board. *Science and Engineering Indicators 2008*. Two Volumes (volume 1, NSB 08-01; volume 2, NSB 08-01A). Arlington, VA: National Science Foundation, 2008.
- Nye, B., L.V. Hedges, and S. Konstantopoulos. "Effects of Small Classes on Academic Achievement: The Results of the Tennessee Class Size Experiment." *American Educational Research Journal*, vol. 37, no. 1, 2000, pp. 123-151.
- Pearson Scott Foresman. "Scott Foresman-Addison Wesley Mathematics: Putting Research into Practice." Glenview, IL: Pearson Scott Foresman, undated.
- Princiotta, D., K.D. Flanagan, and E. Germino Hausken. *Fifth Grade: Findings from the Fifth Grade Follow-Up of the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99*. Publication no. NCES-2006-038. U.S. Department of Education. Washington, DC: National Center for Education Statistics.
- 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.
- Raudenbush, Stephen W. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage Publications, 2002.
- Riordan, J.E., and P.E. Noyce. "The impact of two standards-based mathematics curricula on student achievement in Massachusetts." *Journal for Research in Mathematics Education*, vol. 32, no. 4, 2001, pp. 368-398.

- Rock, Donald A., and Judith M. Pollack. "Early Childhood Longitudinal Study—Kindergarten Class of 1998-99 (ECLS-K), Psychometric Report for Kindergarten Through First Grade." Publication No. NCES 2002-05. Washington, DC: U.S. Department of Education, National Center for Education Statistics, 2002.
- Russell, Susan J., Karen Economopoulos, Jan Mokros, Marlene Kliman, Tracey Wright, Douglas H. Clements, Anne Goodrow, Megan Murray, and Julie Sarama. *Investigations in Number, Data, and Space. Grade 1*. Glenview, IL: Pearson Scott Foresman, 2006.
- Russell, Susan Jo, Karen Economopoulos, Megan Murray, Jan Mokros, and Anne Goodrow. *Implementing the Investigations in Number, Data, and Space Curriculum: Grades K, 1, and 2*. Glenview IL: Pearson Scott Foresman, 2004.
- Schoenfeld, Alan H. "The Math Wars." *Educational Policy*, vol. 18, no. 1, 2004, pp. 253-286.
- Slavin, Robert E., and Cynthia Lake. "Effective Programs in Mathematics: A Best-Evidence Synthesis." Working paper. Baltimore, MD: The Johns Hopkins University, February 2007.
- Tukey, J.W. "The Problem of Multiple Comparisons." "Allowances for Various Types of Error Rates." Unpublished IMS address, Chicago, IL, 1952.
- Tukey, J.W. "The Problem of Multiple Comparisons." Unpublished manuscript, 1953.
- Waite, R.D. "A study of the effects of Everyday Mathematics on student achievement of third-, fourth-, and fifth-grade students in a large north Texas urban school district." *Dissertation Abstracts International*, 61 (10), 2001, 3933A. (UMI No. 9992659).
- 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.
- Woodward, J., and J. Baxter. "The effects of an innovative approach to mathematics on academically low-achieving students in inclusive settings." *Exceptional Children*, vol. 63 no. 3, 1997, pp. 373-388.
- What Works Clearinghouse. "Elementary School Math." Washington, DC: U.S. Department of Education, 2006. [<http://www.whatworks.ed.gov/Topic.asp?tid=04&ReturnPage=default.asp>]. Accessed February 7, 2007.
- Whitehurst, G.J. "Research on Mathematics Instruction." Washington, DC: U.S. Department of Education, 2003. [[http://ies.ed.gov/director/speeches2003/02\\_06/2003\\_02\\_06.asp](http://ies.ed.gov/director/speeches2003/02_06/2003_02_06.asp)]. Accessed February 7, 2007.

## TABLE OF ACRONYMS

AYP	Adequate yearly progress
CCD	Common Core of Data
CFP	Curriculum Focal Points
CMW	Children’s Math Worlds
ECLS-K	Early Childhood Longitudinal Study-Kindergarten Class of 1998-99
ELL	English language learner
ETS	Educational Testing Service
HLM	Hierarchical linear modeling
IB	International Baccalaureate
ICC	Intracluster correlation coefficients
IEP	Individualized education plan
IES	Institute of Education Sciences
Investigations	Investigations in Number, Data, and Space
IRT	Item response theory
K-3	Kindergarten through third grade
K-5	Kindergarten through fifth grade
K-12	Kindergarten through twelfth grade
LEP	Limited English proficiency
MPR	Mathematica Policy Research, Inc.
NCLB	No Child Left Behind
NCES	National Center for Education Statistics
NCTM	National Council of Teachers of Mathematics
NS	Not specified
PD	Professional development
Saxon	Saxon Math
SFAW	Scott Foresman-Addison Wesley Mathematics
SRI	SRI International
WWC	What Works Clearinghouse



## **APPENDIX A**

### **DATA COLLECTION AND RESPONSE RATES**



This appendix provides an overview of the data collected from 2006-2007 school year participants. It also provides a detailed account of student sampling, and data collection procedures and response rates. The data collection instruments are contained in the study's design report (Agodini et al. 2008).

## **A. OVERVIEW OF SAMPLE, RANDOM ASSIGNMENT, AND DATA COLLECTION ACTIVITIES**

A total of 4 districts with 40 schools and 134 classrooms started the study's first (2006-2007) school year of curriculum implementation and data collection. District and school recruitment for this first cohort of participants was conducted from March to June 2006. As described in Chapter I, a blocked random assignment procedure was used to randomly assign schools within each district to one of the four curricula included the evaluation.

To illustrate the idea behind the random assignment procedure, consider a district with eight schools. Suppose the only difference between the schools is the number of first grade students, where four schools have a small number of first graders and the other four have a large number. The blocked random assignment procedure creates two blocks with four schools each, where the first block contains the four small schools and the second block contains the four large schools. The four curricula are then randomly assigned (without replacement) to the four schools in each block, which results in the same sample size and characteristics for each curriculum—two schools per curriculum, where one school contains a small number of students and the other a large number. The study team used a more complex procedure because several school characteristics were used to create the blocks, and the number of schools in some districts was not a multiple of four. For example, suppose the study includes two districts with 6 schools each—a total of 12 schools. To provide each curriculum with the same number of schools, three schools would be assigned to each curriculum across the two districts.

Two districts began the study with 8 schools each, and the other two districts with 12 schools each.

- In the two districts with 8 schools each, two blocks with 4 schools each were constructed in each district. The four curricula were then randomly assigned to the 4 schools in each block.
- In one of the districts with 12 schools, the district indicated that 4 groups with 3 schools in each fed into the district's four middle schools. It was important to the district that all students feeding into the same middle school used the same curriculum in the early grades, so the same curriculum was assigned to the schools in each feeder group.
- The other district with 12 schools initially confirmed that 13 schools would participate in the study. Of the 13 schools, 4 were magnets and were grouped into their own block, and 2 other blocks with 4 schools each were constructed. The 13th school became its own block. Investigations, Math Expressions, and Saxon were assigned 3 schools each, and Scott Foresman-Addison Wesley (SFAW) was assigned 4 schools. As schools were being notified of their curriculum assignments, we learned

that discussions about study participation in one of the schools that had not yet been notified of their curriculum assignment did not include some key school staff. Those staff explained that the school was applying to become an International Baccalaureate (IB) school and would adopt the IB Primary Years Programme. Because IB Primary Years Programme has its own curriculum, the school could not participate in the study, thereby dropping from 13 to 12 schools in the district. Since the school that could not participate was assigned Saxon, this left Saxon with 2 schools in that district.

All first grade mathematics teachers were recruited into the study in each of the study schools.<sup>63</sup> Teacher lists were provided by districts or individual schools, and teachers completed an agreement form acknowledging that they understood the data collection requirements and agreed to participate in the curriculum training provided by the publishers and to use the curriculum assigned to their school.

While the main focus of the study was on teachers who provide primary math instruction to a class of students, those providing supplemental math instruction and those assisting teachers with mathematics instruction were also asked to attend the curriculum training. During the initial curriculum training, which occurred shortly before the start of the school year, an assessment of math content and pedagogical knowledge was administered to teachers. Teachers also completed surveys in fall 2006 and spring 2007.<sup>64</sup>

Class rosters were collected for all the first grade classrooms in the study during the first two weeks of school and again in the spring. The fall rosters were used to identify all of the students to whom parent consent forms need to be distributed. In addition, the fall rosters were used to select the student sample. The student sample was randomly selected from all students enrolled in the 134 classrooms. Class rosters were collected again in the spring to identify new arrivers (those transferring into the classes after the fall 2006 test administration). Along with the spring rosters, student demographic information was also collected for all students enrolled in the study classrooms with parental consent.

Parent consent forms (to allow student testing) were distributed in the fall to parents of all students in study classrooms through the school or teachers. Parental consent was collected prior to testing (see the section on Obtaining Class Lists and Parent Consent for details). Students whose parents returned refusal forms, students who did not speak English, and those ineligible for testing due to cognitive or physical disabilities were excluded from testing. Parent consent forms were distributed in the spring to all new arrivers. All new arrivers that were both eligible for testing and had parental consent were included in the testing effort.

---

<sup>63</sup> Three special education classes did not participate in the study due to the nature of student disabilities.

<sup>64</sup> Each study classroom also was observed in spring 2007. All math instruction provided throughout the day was observed, including morning meetings or calendar time, the math lesson, and any additional practice or drill work provided at other times of the day. Classroom observation data are not part of this report.

While parental consent was not taken into consideration for sample selection, students ineligible for testing were not included in the sampling frame. Three students with testing barriers were sampled in fall 2006 and later identified as ineligible for testing.

The mathematics assessment from the Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K) was administered to students in fall 2006 and spring 2007 by the study's field testers. Testers attended a four-day testing and sampling training and were required to pass a certification test prior to data collection.

## **B. TEACHER SAMPLE AND DATA COLLECTION**

During the recruitment phase, districts or individual schools provided lists of first grade teachers and distributed study information packets containing teacher agreement forms to the first grade teachers. The agreement forms were signed by teachers indicating they understood the various data collection efforts and curriculum training activities in which they would be participating. Agreement forms were collected for all first grade teachers who provided math instruction in the sampled schools. This included 134 first grade teachers who provided the primary math instruction in the 134 classrooms that began study participation and 23 teachers who taught supplemental math instruction to first graders in pull-out resource or special education programs, or who assisted the primary math teacher during regular math instruction.<sup>65</sup> Self-contained special education classes were included in the study if their students were able to be tested.

Teachers attended an initial curriculum training provided by the publisher. A total of 20 trainings were held (each of the four publishers conducted at least one initial training session in each of the four districts) before the start of the school year. When teachers could not attend the main initial training date, publishers scheduled make-up training sessions for those teachers.

The first activity at the initial training session was a teacher assessment. All teachers who provide math instruction (either as primary classroom teachers or supplemental teachers) were asked to complete an assessment designed to measure their math content and pedagogical knowledge. The assessment was voluntary and administered by study team members. The study team took attendance at the initial training session and logged the number of training hours for each teacher.

As reported in Table A.1, one school withdrew from the study partway through the school year. After a few months of curriculum implementation, one of the 40 schools that began study participation indicated it was going to stop using its assigned curriculum (Math Expressions) and would not allow the study to test students in the spring. Because spring achievement is the outcome used to assess the relative effects of the curricula, the school—which contained three teachers—had to be excluded from the analysis, leaving 39 schools and 131 classrooms that could be included in the analysis. Figure A.1 shows the flow of schools through the study.

---

<sup>65</sup> An additional 2 team teachers also participated in training and the study with their teaching partners and 4 teachers left school or went out on leave during the year and their replacements were recruited into the study.

TABLE A.1

NUMBER OF SCHOOLS AND FIRST GRADE CLASSROOMS PARTICIPATING IN THE STUDY  
DURING THE 2006-2007 SCHOOL YEAR, BY CURRICULA

Curriculum	Schools		Classrooms	
	Fall 2006	Spring 2007	Fall 2006	Spring 2007
All	40	39	134	131
Investigations	10	10	33	33
Math Expressions	10	9	34	31
Saxon	9	9	31	31
SFAW	11	11	36	36

### Teacher Assessments

Teachers were asked to complete the math content and pedagogical assessment during their initial curriculum training. Ninety-six percent of the primary classroom math teachers completed the assessment (Table A.2). An additional 23 supplemental teachers and teaching assistants who provide math instruction to students and attended the teacher training also completed the assessment (not shown in tables).

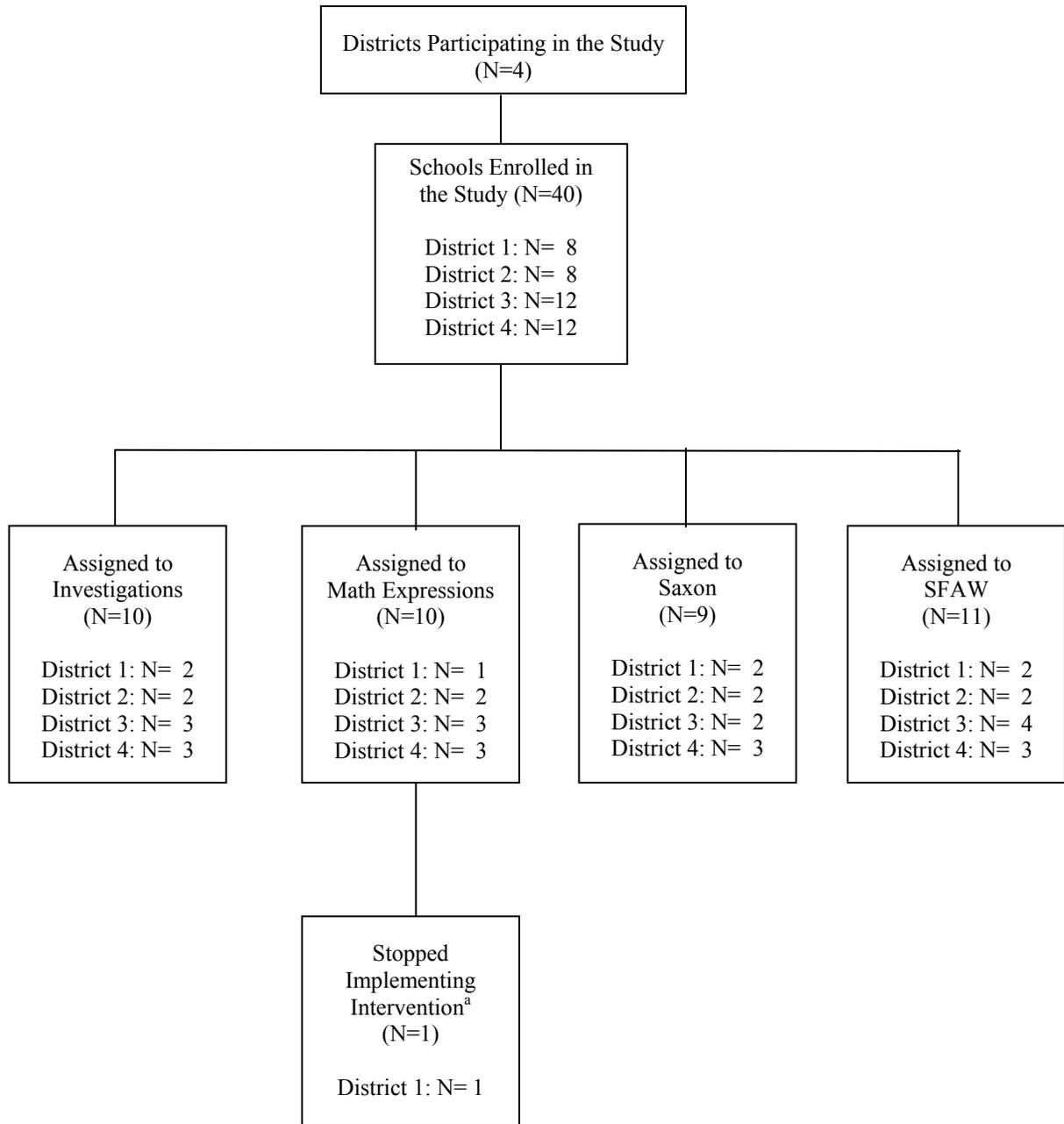
### Teacher Surveys

**Fall Survey.** In November 2006, the fall teacher questionnaire was mailed to teachers at their schools. Supplemental and assistant teachers were included in the mailing. A second mailing was sent to teachers' home addresses if they did not respond to the survey by December. A total of 130 primary math teachers completed the teacher survey, providing teacher data for 97 percent of classrooms (Table A.2).

**Spring Survey.** A spring follow-up survey was mailed in April 2007 to primary teachers in all classrooms still participating in the study (131 of the original 134 classrooms participated the entire academic year). Supplemental and assistant teachers were once again included in the mailing. A second mailing was sent to teachers who did not return a completed questionnaire within three weeks. Nonresponse follow-up was conducted using email prompts and field staff who were conducting spring student testing in the schools. A total of 88 percent of primary teachers in the original 134 classrooms participating at baseline completed the survey—a total of 90 percent of teachers in the 131 classrooms that remained in the study through spring 2007 (derived from Table A.2).

FIGURE A.1

FLOW OF SCHOOLS THROUGH THE STUDY



<sup>a</sup> One school in District 1 stopped implementing the intervention during the school year and did not permit follow-up data collection.

TABLE A.2

NUMBER AND PERCENTAGE OF CLASSROOMS IN WHICH THE PRIMARY MATHEMATICS TEACHER COMPLETED THE TEACHER KNOWLEDGE ASSESSMENT, AND THE FALL AND SPRING SURVEYS, BY CURRICULA: 2006-2007 SCHOOL YEAR PARTICIPANTS

Curriculum	Teachers*	Teachers Completing					
		Teacher Knowledge Assessment		Fall Teacher Survey		Spring Teacher Survey	
		Number	Percentage	Number	Percentage	Number	Percentage
All	134	129	96	130	97	118	88
Investigations	33	32	97	32	97	29	88
Math Expressions	34	31	91	33	97	28	82
Saxon	31	30	97	29	94	27	87
SFAW	36	36	100	36	100	34	94

\*Response rates presented in this table are based on the 134 classrooms that began the study in fall 2006, although there were only 131 classrooms remaining in the study for the spring 2007 teacher survey.

### Student Testing

The study team and a panel of experts in mathematics and math education reviewed several individually administered mathematics assessments and one group-administered assessment designed for first graders. Each panel member reviewed the curricula in the study and the assessments under consideration. The goal was to select a test that was not biased toward one or some of the curricula. The math assessment developed for the Early Childhood Longitudinal Study (ECLS-K, K1 Math Assessment) was selected for the study. It is an individually administered, nationally normed, and adaptive test.

The test was administered by the study's field testers, who attended a four-day testing and sampling training. Field staff were required to pass certification tests in sampling and field assessment prior to the fall testing effort, and only certified field assessors were used to collect data. Testing staff also received refresher training prior to the spring testing effort.

### Obtaining Class Lists and Parent Consent

Within one to two weeks of the first day of school, field staff obtained class rosters for each math teacher and reviewed with classroom teachers to ensure that all students enrolled in the class were included, to identify students listed but not enrolled in the class, and to identify students with language or other barriers that would deem them ineligible for testing. Just over 2,900 students were listed on the 134 class rosters. Of these, 2,770 were actually enrolled and eligible for student testing.

Parent consent forms to allow student testing were distributed to parents of all students in the study classrooms. Each of the four districts required passive consent. To obtain passive consent, permission forms were sent to parents along with a letter and brochure describing the study. Only parents who did not want their children to participate in the testing or to share their

student records were required to send in a signed refusal form. Parents were given at least one week to return refusal forms to the school before testing began.

A total of 45 refusals out of the 1,525 students randomly sampled for testing were received from parents (Table A.3). Thus, parent consent was obtained for 97 percent of the sampled students. The consent rate was the same for each of the four curriculum groups. Of the 148 new arrivers identified in spring 2007, parental consent was obtained for 140 students—a consent rate of 95 percent.

TABLE A.3

PARENT CONSENT RATES BY CURRICULA AND SAMPLED STUDENTS' ENTRY INTO THE STUDY:  
2006-2007 SCHOOL YEAR

Curriculum	Fall 2006 Student Sample			Spring 2007 New Arrivers		
	Total	With Parent Consent at Fall Testing		Total	With Parent Consent at Spring Testing	
		N	%		N	%
All	1,525	1,480	97	148	140	95
Investigations	379	367	97	34	34	100
Math Expressions	385	376	98	38	35	92
Saxon	352	340	97	34	31	91
SFAW	409	397	97	42	40	95

### Sampling Procedures

Once class rosters were collected, field staff reviewed the rosters with the teachers for accuracy and completeness. Teachers were asked to confirm that all children listed on the roster were enrolled in the class. Those who were not actually in the class were eliminated from the roster. Teachers were then asked to identify any students in their classes whose names were missing from the roster so that those names could be added. Field staff compared the total count of names on the final roster against the total class size to ensure accuracy.

Field samplers also worked with teachers to identify children who would not be able to participate in the study's individually administered math assessment given in English. Roughly 5 percent of the students on the original rosters were excluded because they were not actually enrolled in the class, did not speak English, or had physical or cognitive barriers that precluded testing. Of the 2,770 students enrolled and eligible for testing, a sample of 1,525 students was selected.

Student sampling was conducted for each classroom using a unique sampling matrix with a table of random numbers aligned to the class size. Matrices were developed to sample an average of 11 students per classroom. Field staff trained to sample students used a student tracking form (listing and numbering all eligible students) and a sampling matrix to randomly select the correct number of eligible students in each class. If a school had only one first grade teacher, the

matrices were designed to select all eligible students in the classroom. In schools with two first grade classes, up to 16 students per classroom were selected. In schools with three or more first grade teachers, up to 11 students per class were selected. Variations in the number of classrooms per school resulted in an average sample size of 11 students per classroom and 38 students per school in fall 2006.

In spring 2007, rosters were once again collected to identify students who had transferred into the class after baseline and a total of 148 new students were found. Consent packets were sent home to their parents and all new arrivers who were eligible for testing (that is, did not have a physical, cognitive, or language barrier) were added to the sample and those with parental consent were included in the spring assessment (140, or 95 percent of the 148 new arrivers were tested).

### **Student Testing Response Rates**

Student assessments were administered during the school day in the fall and spring, as close to the start and end of the school year as possible. Response rates were high. Fall tests were administered to 1,457 eligible students, or just under 96 percent of the fall student sample (Table A.4). Parent refusals accounted for almost two-thirds of student nonresponse. Almost all (98 percent) of students with parental consent were tested (derived from Table A.4).

At spring followup, tests were administered to 87 percent of the initial baseline sample of 1,525 students (Table A.4). Most nonresponse was due to sample attrition. Thirty-two students were lost when their school withdrew from the study. An additional 99 students moved or transferred out of the research schools between fall and spring testing. Five students still enrolled in a study school changed grades between fall and spring and were missed at testing, along with 14 other students we were unable to test during the spring follow-up data collection. Again, almost all (98 percent) of students with parental consent who were still enrolled in a study classroom were tested (derived from Table A.4). Ninety-two percent of students who transferred into the research classrooms after the fall data collection (new arrivers) were tested at spring follow-up. Student response rates for the baseline sample were similar across curricula, ranging from 95 to 96 percent at baseline and from 83 to 89 percent at spring follow-up (Table A.5). Figure A.2 summarizes the flow of students through the study.

### **Student Testing Response Rates of the Analysis Samples**

Tables A.3 through A.5 are based on the full study sample and account for all students enrolled in the study at baseline and their consent and testing status during both baseline and follow-up testing. These tables also provide consent and test data separately for new arrivers. The analyses in this report focus on two samples: (1) students who were enrolled in a

TABLE A.4

NUMBER AND PERCENTAGE OF SAMPLED STUDENTS TESTED AND TYPES OF  
NONRESPONSE: 2006-2007 SCHOOL YEAR

	Students Tested			Number of Non-Responders by Type				
	Total	Number	%	Parent Refusal	School Dropped	Moved	Changed Grade	Other Nonresponse
Fall 2006 Initial Sample	1,525	1,457	96	45	--	--	--	23
Spring 2007 Initial Sample	1,525	1,330	87	45	32	99	5	14
Spring 2007 New Arrivers	148	136	92	8	--	--	--	4

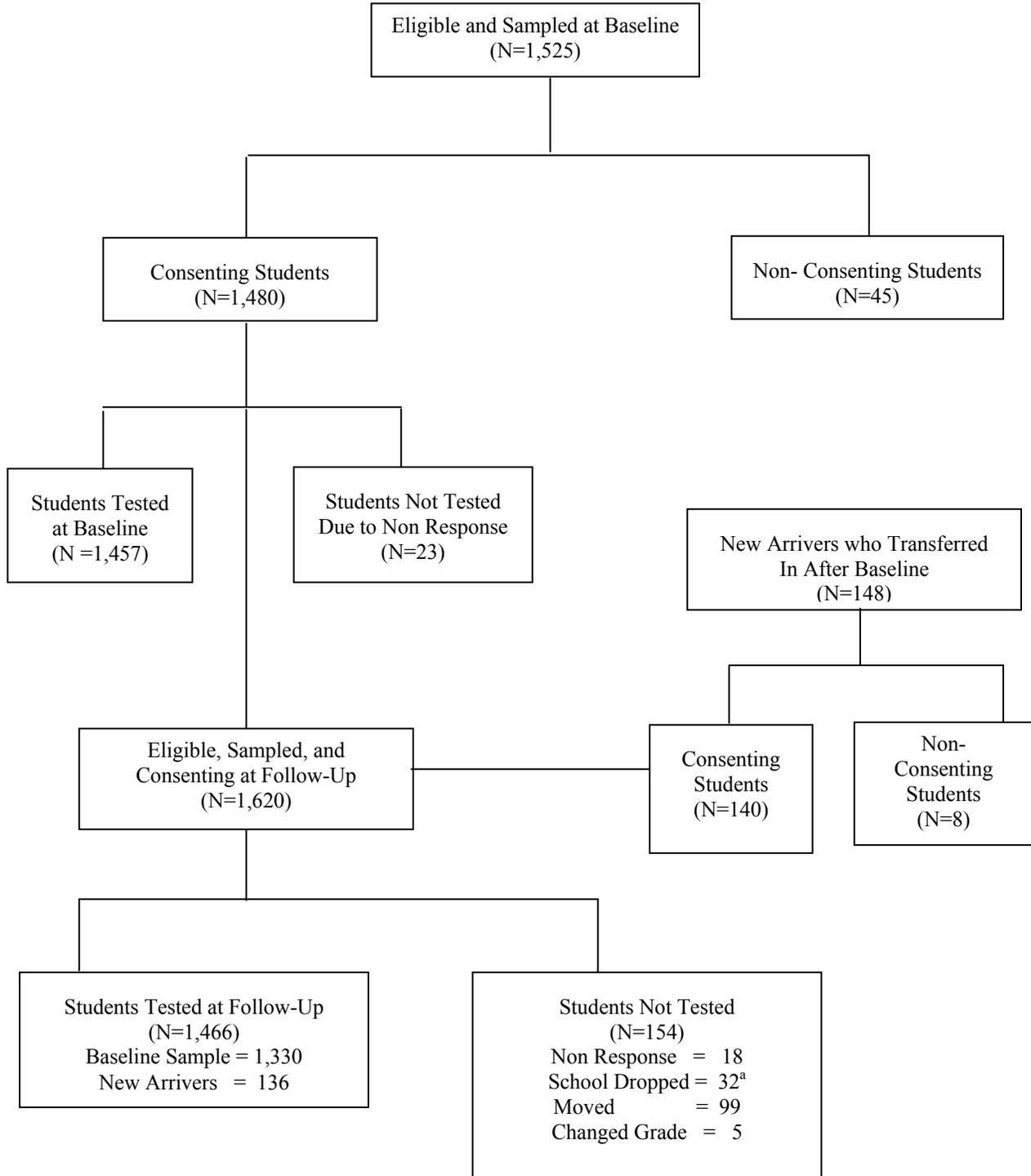
TABLE A.5

NUMBER AND PERCENTAGE OF BASELINE STUDENTS AND NEW ARRIVERS SAMPLED FOR  
TESTING, BY ROUND OF TESTING AND CURRICULA: 2006-2007 SCHOOL YEAR

Curriculum	Students Sampled at Baseline					Students Added as New Arrivers in Spring		
	Tested in Fall			Tested in Spring		Tested in Spring		
	Total	N	%	N	%	Total	N	%
All	1,525	1,457	96	1,330	87	148	136	92
Investigations	379	365	96	334	88	34	32	94
Math Expressions	385	368	96	321	83	38	34	89
Saxon	352	334	95	310	88	34	31	91
SFAW	409	390	95	365	89	42	39	93

FIGURE A.2

FLOW OF STUDENTS THROUGH THE STUDY



<sup>a</sup> One school dropped out of the study during the school year and did not permit follow-up student testing.

study classroom at both fall and spring testing (the longitudinal sample), and (2) students who were enrolled in a study classroom in the spring (the cross-sectional sample).

***Longitudinal Sample.*** Table A.6 reports the number of sampled students who were enrolled in the study classrooms in both fall and spring, and of these, the number and percentage tested. Of the 1,387 sampled students enrolled in the classrooms and eligible for testing, 94 percent were tested in both the fall and spring. This ranged from 93 to 95 percent across the curriculum groups.

***Cross-Sectional Sample.*** A spring cross-section sample may be relevant for policy as it may best reflect district and state annual testing programs (those actually enrolled in the classroom in the spring semester, regardless of when they arrived). Table A.7 provides the total number of students enrolled in the study classrooms in the spring, and the number and percentage tested of the 1,535 eligible students.

### **Timing of the Tests**

In the fall, tests were administered in each school within four weeks of the first day of classes. School start dates ranged from August 21 to September 11, and testing was conducted from September 13 through October 6. Spring tests were administered within one to six weeks of the end of the academic year. The goal was to keep the window for testing comparable across the curricula in the fall and spring. Spring assessments were administered from 210 to 244 days after fall testing (Table A.8).

### **Test Processing and Scoring**

Tests were administered using desk-top easels and laptop computers into which testers keyed student responses. The ECLS-K test begins with a routing section designed to assess a student's achievement level and to direct the child to the most appropriate test level (easy, middle-difficulty, or hard). The computer test program tracked the number of correct and incorrect responses during the routing section and automatically routed students to the appropriate math assessment, thus eliminating field assessor scoring errors.

Cleaned electronic test files were sent to Educational Testing Service (ETS) for item response theory (IRT) scoring. ETS was a developer of the ECLS-K mathematics assessment. For further information, see the methodology report prepared by ETS and the National Center for Education Statistics (NCES) describing in detail the psychometric properties of the ECLS-K mathematics assessment. The report is available through NCES and is posted on their website (Rock and Pollack 2002).

TABLE A.6

NUMBER OF SAMPLED STUDENTS ENROLLED AND ELIGIBLE FOR TESTING AT BOTH BASELINE AND SPRING FOLLOWUP, AND NUMBER AND PERCENTAGE TESTED IN BOTH THE FALL AND SPRING: 2006-2007 SCHOOL YEAR

Curriculum	Sampled Students Eligible for Testing		
	Total	Tested Fall and Spring	
		Number	%
All	1,387	1,309	94
Investigations	350	332	95
Math Expressions	329	314	95
Saxon	326	304	93
SFAW	382	359	94

Note: The sample excludes 1 Math Expressions school (with 3 classrooms and 32 students) that participated during part of the school year and then stopped using the curriculum and did not allow the study to collect follow-up data.

TABLE A.7

NUMBER OF SAMPLED STUDENTS ENROLLED AND ELIGIBLE FOR TESTING IN THE SPRING AND NUMBER AND PERCENTAGE TESTED, BY CURRICULUM AND TYPE OF SAMPLE—CROSS-SECTION SAMPLE: SPRING 2007

Curriculum	Sampled Students Eligible for Testing in Spring 2007								
	All			Longitudinal			New Arrivers		
	Total	Tested		Total	Tested		Total	Tested	
		N	%		N	%		N	%
All	1,535	1,466	96	1,387	1,330	96	148	136	92
Investigations	384	366	95	350	334	95	34	32	94
Math Expressions	367	355	97	329	321	98	38	34	89
Saxon	360	341	95	326	310	95	34	31	91
SFAW	424	404	95	382	365	96	42	39	93

Note: The sample excludes 1 Math Expressions school (with 3 classrooms and 32 students) that participated during part of the school year and then stopped using the curriculum and did not allow the study to collect follow-up data.

TABLE A.8

TESTING DATES AND NUMBER OF DAYS BETWEEN FALL AND SPRING TESTING START DATES  
AND END DATES, BY CURRICULA: 2006-2007 SCHOOL YEAR

Curriculum	Fall Baseline Testing Dates	Spring Follow-up Testing Dates	Number of Days Between Fall and Spring Testing Start Dates	Number of Days Between Fall and Spring Testing End Dates
Investigations	Sept. 13-Oct. 5	April 16-June 6	215	244
Math Expressions	Sept. 19-Oct. 5	April 17-June 6	210	244
Saxon	Sept. 13-Oct. 6	April 18-June 6	217	243
SFAW	Sept. 13-Oct. 6	April 24-June 4	223	241

TABLE A.9

NUMBER AND PERCENTAGE OF STUDENTS FOR WHOM STUDENT DEMOGRAPHIC RECORDS  
AND INDIVIDUAL DEMOGRAPHIC ITEMS WERE COLLECTED,  
BY TYPE OF SAMPLE AND ITEM: 2006-2007

Data Forms and Items	Longitudinal		New Arrivers	
	Total	Resp	Total	Resp
	N	%	N	%
Sample	1,387		148	
Forms	1,352	97	140	95
Items				
Age	1,306	94	138	93
Free/reduced-price meals	1,063	77	128	86
Gender	1,352	97	130	88
EP for disability/remediation	1,205	87	135	91
IEP for gifted/talented	1,205	87	135	91
LEP/ELL	1,239	89	140	95
Race/ethnicity	1,210	87	131	89

Note: Item response and item nonresponse are equal to the percentage of students for whom we have data for the individual item divided by all sampled students and thus incorporate missing items due to nonresponse.

## **Student Demographic Records**

The study team requested student demographic data for all students enrolled in the research classrooms with parental consent. A student demographic form was created and given to schools in late spring 2007. The data items obtained for individual students included gender, age, limited English proficient or an English language learner (LEP/ELL), eligibility for free or reduced-price meals, race/ethnicity, and special education plans or services. The study team obtained student records for 97 percent of the longitudinal sample and 95 percent of new arrivers (Table A.9). Eligibility for free or reduced-price meals was reported for 75 percent of the total baseline sample. The study team obtained item response rates of 85 percent or better for all other student characteristics.

**APPENDIX B**

**TEACHER-REPORTED FREQUENCY OF IMPLEMENTING  
OTHER CURRICULUM-SPECIFIC ACTIVITIES**



TABLE B.1

INVESTIGATIONS: TEACHER-REPORTED FREQUENCY OF IMPLEMENTING  
OTHER CURRICULUM-SPECIFIC ACTIVITIES (N = 31)

How Often Teachers Report Doing the Following Activities with the Target Class <sup>a</sup>	Expected Frequency	Mean Response	Median Response
Introduce the tasks for the session	5	4.66	5
Do the Classroom Routines	5	4.48	5
Use students' correct responses as a basis for discussion	4-5	4.00	4
Use students' incorrect responses as a basis for discussion	4-5	3.79	4
Use guidelines in the lesson for individualizing instruction for struggling students	NS	3.48	3
Introduce the homework	2-3	2.69	3
Use Teacher Checkpoints and Embedded Assessments	2-3	2.52	2
Communicate with parents about math activities	2-3	2.52	2
Review homework with the class	NS	1.79	1

Source: Author tabulations using data from the spring 2007 teacher survey. The sample excludes two Investigations teachers who did not complete the above items in the survey.

<sup>a</sup>Teachers were asked to indicate how frequently they implemented the activities on the following scale: 0 (never), 1 (less than once a month), 2 (once or twice a month), 3 (one to two times a week), 4 (three to four times a week), and 5 (daily). A mean of 4 indicates that teachers implemented an activity an average of three to four times a week.

NS indicates the expected frequency was not specified.

TABLE B.2

MATH EXPRESSIONS: TEACHER-REPORTED FREQUENCY OF IMPLEMENTING  
OTHER CURRICULUM-SPECIFIC ACTIVITIES (N = 27)

How Often Teachers Report Doing the Following Activities with the Target Class <sup>a</sup>	Expected Frequency	Mean Response	Median Response
Use teaching the lesson activities	5	4.67	5
Assign the remembering worksheet	5	4.19	5
Group students for each activity as recommended in the teachers' guide	5	3.46	4
Use differentiated instruction activities	NS	3.15	3
Use math writing prompts	3-4	2.85	3
Conduct ongoing assessment activities	4-5	2.81	3
Administer unit tests	2	1.78	2

Source: Author tabulations using data from the spring 2007 teacher survey.

<sup>a</sup>Teachers were asked to indicate how frequently they implemented the activities on the following scale: 0 (never), 1 (less than once a month), 2 (once or twice a month), 3 (one to two times a week), 4 (three to four times a week), and 5 (daily). A mean of 4 indicates that teachers implemented an activity an average of three to four times a week.

NS indicates the expected frequency was not specified.

TABLE B.3

SAXON MATH: TEACHER-REPORTED FREQUENCY OF IMPLEMENTING  
OTHER CURRICULUM-SPECIFIC ACTIVITIES (N = 31)

How Often Teachers Report Doing the Following Activities with the Target Class <sup>a</sup>	Expected Frequency	Mean Response	Median Response
Prepare all required materials in advance of the lesson	5	4.63	5
Group students for each activity as specified in the lessons	5	4.17	5
Preview the homework for students	5	3.87	5
Administer oral assessments and record student responses	2	1.90	2

Source: Author tabulations using data from the spring 2007 teacher survey.

<sup>a</sup>Teachers were asked to indicate how frequently they implemented the activities on the following scale: 0 (never), 1 (less than once a month), 2 (once or twice a month), 3 (one to two times a week), 4 (three to four times a week), and 5 (daily). A mean of 4 indicates that teachers implemented an activity an average of three to four times a week.

TABLE B.4

SFAW MATH: TEACHER-REPORTED FREQUENCY OF IMPLEMENTING  
OTHER CURRICULUM-SPECIFIC ACTIVITIES (N = 30)

How Often Teachers Report Doing the Following Activities with the Target Class <sup>a</sup>	Expected Frequency	Mean Response	Median Response
State the objective of the lesson	5	4.84	5
Provide step-by-step guidance on how to complete the practice page	NS	4.69	5
Provide reading assistance to students as they complete the practice page	NS	4.52	5
Introduce the vocabulary specified in the lesson	3-4	4.41	5
Provide additional activities for “early finishers”	4-5	3.69	4
Do the spiral review	5	3.65	4
Group students into small groups for collaborative activities	NS	3.63	3
Use the leveled practice provided for students at varying levels (below, on level, above)	NS	2.77	3
Use instant check mat	NS	2.25	2.5
Provide opportunities for students to use online materials or other supplemental materials provided by SFAW	NS	0.63	0

Source: Author tabulations using data from the spring 2007 teacher survey.

<sup>a</sup>Teachers were asked to indicate how frequently they implemented the activities on the following scale: 0 (never), 1 (less than once a month), 2 (once or twice a month), 3 (one to two times a week), 4 (three to four times a week), and 5 (daily). A mean of 4 indicates that teachers implemented an activity an average of three to four times a week.

NS indicates the expected frequency was not specified.

**APPENDIX C**

**GLOSSARY OF CURRICULUM-SPECIFIC TERMS**



## Investigations

**Activities** – “each Investigation is conducted through a series of activities that include pair and small-group work, individual tasks, and whole-class discussions.” “Activities are loosely grouped by one-hour class sessions.” (Russell et al. 2004, p.7)

**Allow Students to Choose Manipulatives for Use During the Activity** – during each activity, students should choose the manipulatives they wish to use to solve the problem. “A key part of the teacher’s job is to ensure that it becomes natural for students to use appropriate materials as they solve problems.” (Russell et al. 2004, p. 11)

**Choice Time Activities** – an “activity time when students work simultaneously on different activities focused on similar mathematical content. Individual students or pairs choose which activities to work on and make their own decisions about when to move from one to another.” (Russell et al. 2004, p. 7)

**Classroom Routines** – “activities in counting, exploring data, and understanding time and changes” that should be done on a regular basis. The activities are often incorporated into a daily schedule, such as a morning meeting. “Routines are short and can be done whenever you have a spare 10-15 minutes.” (Kliman et al. 2006, p. I-4)

**Embedded Assessments** – specific activities within a unit that are “designed to help teachers examine the work of individual students, figure out what it means, and provide feedback. From the student’s point of view ... [they] are no different from any others, they don’t look or feel like traditional tests. These activities sometimes involve writing and reflecting, ... a brief interaction between student and teacher, and ... the creation and explanation of a product.” (Russell et al. 2004, p.14)

**Homework** – “... an extension of classroom work. Sometimes it offers review and practice of work done in class, sometimes preparation for upcoming activities, and sometimes numerical practice that revisits work in earlier units.” (Kliman et al. 2006, pp. I-5 to I-6)

**Hundred (100) Chart** –



1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

**Introduce the Tasks for the Session** – a session is at least a one-hour math class. “Sessions are numbered consecutively through an investigation” and they are often grouped into multiple sessions that comprise a single activity. (Kliman et al. 2006, pp. I-2 to I-3)

**Investigations** – in-depth projects in which students collect, organize, represent, and analyze data. Investigations vary in length, from two to three days, to one to three weeks. Each investigation involves a small number of problems that students work on in depth, with

“students actively using mathematical tools and consulting with peers as they find their own ways to solve problems.” (Russell et al. 2004, pp. 6-7)

**Manipulatives** – “concrete materials,” including (but not limited to) “interlocking cubes, 100 charts, geometric shapes, play money, and rulers.” (Russell et al. 2004, pp. 11)

**Teacher Checkpoints** – “time for teachers to pause and reflect on their teaching plans, to observe students at work, and to get an overall sense of how the class is doing in the unit.” Checkpoints “offer tips on what teachers should be looking for and how they might adjust their pacing.” (Russell et al. 2004, pp. 13-14)

## Math Expressions

**Daily Routines** – a set of activities that are performed on a regular basis, not necessarily during the regular math lesson. The activities can include (but are not limited to) solving problems with money, number charts, counting, and calendars. (Fuson 2006b, pp. xxi-xxvi)

**Differentiated Instruction Activities** - “Every ... lesson includes intervention, on level, and challenge differentiation to support classroom needs.” (Fuson 2006b, p. x)

**Homework** – “children complete homework assignments every night.” Homework “develops and consolidates understanding of math concepts” and “help children become organized and self regulatory.” (Fuson 2006b, p. xviii-xix)

**Math Writing Prompts** – writing activities for students that are noted in the teacher’s guide to “provide opportunities for in-depth thinking and analysis.” (Fuson 2006b, pp. x-xi)

**Ongoing Assessment Activities** – questions are provided in every lesson for teachers to ask students, providing an informal assessment of student achievement (Fuson 2006b)

**Proof Drawings** – “A picture children create to show how to solve a problem including the solution.” (Fuson 2006c, p. T10)

**Quick Practice** – “The opening 5-10 minutes of each math period are dedicated to activities that allow students to practice newly-acquired knowledge. These consolidating activities help students become faster and more accurate with concepts ... activities are the same throughout each unit. In this way they become familiar routines ...” (Fuson 2006b, p. xviii)

**Quick Quizzes** – formal assessments that contain open response questions. They are administered after a set of lessons on a similar concept (for example, after a few lessons on Addition Stories with Unknown Partners). (Fuson 2006b)

**Remembering Worksheet** – “provide practice with important concepts covered in all the units to date.” They are intended for use when children are in need of a refresher of what they have learned and can be used as extra homework. (Fuson 2006b, p. xix)

**Scenarios** – “A group of students is called to the front of the classroom to act out a particular situation.” “The main purpose ... is to demonstrate mathematical relationships in a visual and memorable way.” (Fuson 2006b, p. xx)

**Solve and Discuss at the Board** – “The teacher selects 4 to 5 children ... to go to the classroom board and solve a problem, using any method they choose. Their classmates work on the same problem at their desks. Then the teacher picks 2 or 3 children to explain their methods. Students at their desks are encouraged to ask questions and assist their classmates in understanding.” (Fuson 2006b, p. xix)

**Step-by-Step at the Board** – “Several children go to the board to solve a problem ... a different student performs each step of the problem, describing the step before everyone does it. Everyone else at the board and at their desks carries out that step.” (Fuson 2006b, p. xx)

**Student Leaders** – students take on leadership roles to help other students learn. Students may lead practice or a discussion. (Fuson 2006b, p. ix)

**Teaching the Lesson Activities** – most lessons specify at least two activities for teachers to use to convey the day’s lesson to students. (Fuson 2006b)

**Unit Tests** – formal assessments that are administered at the end of each unit. Tests can contain open response questions or multiple choice questions. (Fuson 2006b, p. 1C)

## Saxon Math

**Fact Assessment** – formal assessments administered in every fifth lesson to “measure basic fact fluency.” (Larson and Saxon Publishers 2006, p. 34)

**Fact Practice** – “Students apply addition and subtraction strategies introduced during the lesson to develop automaticity of basic math facts.” This practice is often performed on worksheets. (Larson and Saxon Publishers 2006, p.20)

**Guided Class Practice Worksheet** – immediately after the math lesson, “students apply what they have learned and the teacher is able to provide further explanation ...” (Larson and Saxon Publishers 2006, p. 11). Students often apply what they have learned using worksheets that contain information learned in the day’s lesson along with previously taught concepts. (Larson and Saxon Publishers 2006, p. 18)

**Homework** – worksheets for students to complete independently. They “include practice of the new increment as well as previously taught concepts ...” (Larson and Saxon Publishers 2006, p. 11)

**Lesson Script** – each day’s lesson includes a “comprehensive script” for teachers to use through each part of the math lesson. (Larson and Saxon Publishers 2006, p.10)

**Manipulatives** – tangible materials such as linking cubes and pattern blocks that “promote student learning through engaging, hands-on math experiences.” (Larson and Saxon Publishers 2006, p. 16)

**The Meeting** – a daily whole-class activity. It “reinforces previously learned concepts and helps students develop the foundational skills needed to learn more advanced math concepts.” The Meeting includes a variety of activities, including calendars, counting, clocks, number patterns, graphs, problem solving, and mental computation. (Larson and Saxon Publishers 2006, p. 10)

**State the Lesson’s Objective from the Script** – each lesson script begins with a statement for the teacher to read to students informing them what they will learn that day. (Larson and Saxon Publishers 2004)

**Written Assessments** – cumulative assessments that occur in every fifth lesson “to assess students’ knowledge and understanding of concepts.” (Larson and Saxon Publishers 2006, p. 34)

## SFAW Math

**Investigating the Concept** – portion of the lesson that occurs towards the beginning of the lesson, during which the new concept for the day is often introduced often through the use of hands-on activities. (Pearson Scott Foresman, p.18)

**Instant Check Mat** – an 8” x 12” erasable blank worksheet that students can use to show their work. (Charles et al. 2005, p. T9)

**Manipulatives** – materials that can be used to represent mathematical concepts, such as counters and base-ten blocks. (Pearson Scott Foresman, p.17)

**Journal Activity** – in every lesson, a journal idea is provided as a means of ongoing assessment. “Journal tasks take on many forms, including drawing pictures and diagrams as well as writing explanations and descriptions.” (Pearson Scott Foresman, p.51)

**Learn! Section of Student Worksheets** –the very first question on student worksheets. It “introduces concepts and vocabulary clearly.” (Charles et al. 2005, p. T8)

**Leveled Practice Provided for Students at Varying Levels** – allows teachers to customize instruction to match student abilities. Each lesson provides suggestions for below level, on level, and above level students. (Charles et al. 2005, p. T9)

**Provide Additional Activities for “Early Finishers”** – each lesson specifies “instructional suggestions for students who complete their assignments early.” (Pearson Scott Foresman, p. 47)

**Provide the Recommended Error Intervention for Struggling Students** – In each lesson, the teacher’s guide provides an “If ... then” suggestion for how teachers should deal with particular student struggles. (Pearson Scott Foresman, p.50)

**State the Lesson Objective** – each daily lesson objective is clearly identified at the beginning of each lesson. (Charles et al. 2005, p. T9)

**Spiral Review** – each lesson includes a problem of the day and a set of “test prep” questions for students (on a worksheet or overhead transparency). The problem of the day and spiral review questions cover previously learned material. (Charles et al. 2005)

**Talk About It Questions** – questions provided in each lesson to give students “an informal assessment opportunity that lets them verbalize their understanding.” (Charles et al. 2005, p. T10)

**Test-Taking Practice** – at the end of each lesson a set of assessment questions can be completed by students on a worksheet or by using an overhead transparency. (Pearson Scott Foresman, p. 52)

**Think About It questions** – questions that teachers should ask students in each lesson to give students “a chance to verbalize and clarify understanding before practice begins.” (Charles et al. 2005, p. T8)

**Warm Up Activity** - An activity that is performed at the beginning of each lesson to activate “prior knowledge of skills your students will need in the upcoming lesson.” (Charles et al. 2005, p. T9)



**APPENDIX D**

**CONSTRUCTING THE ANALYSIS SAMPLES AND  
ESTIMATING CURRICULUM EFFECTS**



This appendix describes how the analysis samples used to estimate curriculum effects were constructed and provides more details about the approach for estimating the effects. The first section describes the students that were included in the analysis samples, and the student-, teacher-, and school-level measures for each student. It also describes the techniques used to impute any missing data and the weights that were developed for the analysis samples. The second section describes the statistical models that were used to estimate relative effects and presents the results for the models. It also describes the models used to estimate curriculum effects for the subgroups that were examined.

## **A. CONSTRUCTING THE ANALYSIS SAMPLES**

Two analysis samples were constructed to estimate the effects of the curricula on student achievement. The primary analysis sample was a longitudinal sample that includes the 1,309 students who were tested in both the fall and spring. For these students, their teacher and school characteristics were measured during the fall assessment. A secondary analysis sample (a cross-sectional sample) includes the 1,466 students who were tested in the spring. Included among these students are those who were tested in both the fall and spring (that is, the longitudinal sample), those who were eligible for testing in the fall but could not be tested then but were tested in the spring, and those who arrived in a study school after the fall assessments were administered. For the cross-sectional sample, teacher and school characteristics were measured during the spring assessment.

### **Measures Included in the Analysis Files**

Both the longitudinal and the cross-sectional analysis files contain student-, teacher-, and school-level measures. Student-level math test scores were obtained from a file provided by Educational Testing Service that included scores based on the fall and spring math assessments. Every student began the assessments with the same first-stage form and, depending on the score on the first stage, was assigned an easy, a middle-difficulty, or a hard second-stage form. Item response theory (IRT) techniques, which analyze patterns of correct and incorrect answers, were used to put scores from the different forms on the same scale to allow comparisons. An overall scale score was constructed that estimates the student's performance on the whole set of assessment questions.

School records were used to construct other student-level measures that were included in the analysis files. The measures include student demographics (age, gender, and race/ethnicity), whether the student is limited English proficiency (LEP) or an English language learner (ELL), and whether that student had an individualized education plan or service (IEP). The number of days between the fall and spring assessments also was constructed and included in the analysis files.

Teacher-level measures were obtained from the consent form, the assessment of math content and pedagogical knowledge, and the fall teacher survey. Teacher experience (in years) was obtained from the consent form that teachers completed before school random assignment occurred. Teachers were administered an assessment of their content knowledge and their pedagogical knowledge before the initial training on their school's assigned curriculum began.

An overall scale score and separate measures of content knowledge and pedagogical knowledge were included in the analysis files. Teacher education, race, and prior use of the assigned curriculum at the K-3 level were obtained from the fall teacher survey. Classroom size was obtained from school rosters and, to measure the heterogeneity of the students in the classroom, the classroom variance and skewness of the fall student math score were computed.

Two school-level measures extracted from the Common Core of Data (CCD) were included in the analysis files. Specifically, the file included the percentage of students receiving a free or reduced-price lunch and whether the school was Title I. In addition, the block that the school was placed into during the random assignment process, the curriculum assigned to the school, and the school district were included in the analysis files.

### **Imputing Missing Data**

Complete data were available for the school-level measures. Complete data also were available for the fall and spring student math test scores of the longitudinal file, and spring student scores of the cross-sectional file.

However, a small fraction of data were missing for some of the other student-level measures and for some of the teacher-level measures. For example, fall math scores were not available for the 7 percent of students in the cross-sectional file who arrived at a study school after the study team completed fall testing.

Tables D.1 and D.2 list the student- and teacher-level measures included in the longitudinal and cross-sectional analysis files. Measures that have a nonzero value in the “Number Missing” column are those student- and teacher-level measures with the small fraction of missing data.

Model-based imputations were used to replace missing data. With this technique, missing values on each measure are replaced with the predicted value of the measure from a regression model. Imputations were done separately for student- and teacher-level measures, and separately for the longitudinal and cross-sectional samples.

For the student-level measures of the longitudinal sample, only some demographic data were missing. Missing data were imputed using the fall math test score, the available demographic data, the school percentage of students receiving a free or reduced-price lunch, whether the school was Title I, and the school district.

Imputing missing student-level measures for the cross-sectional sample was more complex because fall test scores were systematically missing for students who enrolled in a study school after fall testing was complete. These scores were also missing for the small fraction of students who were eligible for testing in the fall but could not be tested. Students who arrived in a study school after the fall assessments were found to be more similar to students who were tested in the fall but left the study school before the spring assessment, than to students in the longitudinal sample (that is, those who were in a study school in both the fall and spring). To use this information, students who were tested in the fall but left the study school before the spring assessment were included in the imputation, and an indicator of whether the student was in a

TABLE D.1

## MODEL-BASED IMPUTATION OF MISSING DATA, LONGITUDINAL SAMPLE

Variable Name	N	Number Missing	Mean (Pre-Imputation)	Mean (Post-Imputation)
<b>Student Level</b>				
Fall math scale score	1,309	0	30.93	30.93
Age at fall test	1,257	52	6.46	6.46
Female	1,299	10	0.50	0.49
Race/ethnicity				
Hispanic	1,163	146	0.22	0.21
Non-Hispanic black	1,163	146	0.19	0.20
LEP/ELL	1,191	118	0.14	0.13
IEP/Special Services	1,158	151	0.07	0.06
<b>Teacher Level</b>				
Master's degree	120	11	0.67	0.67
Experience	131	0	11.78	11.78
Prior use of the assigned curriculum	123	8	0.09	0.11
Black	114	17	0.05	0.05
Assessment				
Overall IRT score	124	7	-0.08	-0.07
Content knowledge IRT score	124	7	-0.62	-0.62
Pedagogical knowledge IRT score	124	7	-0.31	-0.31

TABLE D.2

## MODEL-BASED IMPUTATION OF MISSING DATA, CROSS-SECTIONAL SAMPLE

Variable Name	N	Number Missing	Mean (Pre-Imputation)	Mean (Post-Imputation)
<b>Student Level</b>				
Fall math scale score	1,309	157	30.93	30.92
Age at spring test	1,403	63	7.12	7.12
Female	1,451	15	0.50	0.50
Race/ethnicity				
Hispanic	1,302	164	0.22	0.22
Non-Hispanic black	1,302	164	0.21	0.20
LEP/ELL	1,340	126	0.15	0.14
IEP/Special Services	1,300	166	0.07	0.07
<b>Teacher Level</b>				
Master's degree	122	9	0.66	0.66
Experience	131	0	11.48	11.48
Prior use of the assigned curriculum	120	11	0.09	0.09
Black	116	15	0.05	0.06
Assessment				
Overall IRT score	119	12	-0.10	-0.10
Content knowledge IRT score	119	12	-0.64	-0.64
Pedagogical knowledge IRT score	119	12	-0.32	-0.32

study school for only the fall or the spring was included in the regression model. The imputation model also included the other variables used for the longitudinal sample.

The number of days between the fall and spring assessments also was systematically missing for students who did not complete an assessment in the fall. Since the number of days between the fall and spring assessments is determined by the study's testing schedule and not by other student-level measures, the model-based imputation was not used to replace missing data for this measure. Instead, these students were assigned the average number of days among the students in the same classroom who had data.

Although imputations were conducted separately for the teacher-level measures of the longitudinal and cross-sectional samples, the same regression model was used for both samples. Missing teacher assessment measures and missing teacher survey measures were imputed using the available teacher assessment measures, the available teacher survey measures, teacher experience, the school percentage of students receiving a free or reduced-price lunch, whether the school was Title I, and the school district.

In addition to the number missing for the student- and teacher-level measures included in analysis files, Tables D.1 and D.2 list the means, both pre- and post-imputation, for these measures.

## **Weights**

Separate weights were developed for the longitudinal and cross-sectional samples. For the longitudinal sample, students who were tested in the fall and spring were weighted up to the number of students who were eligible to be tested in the fall, separately for each classroom. For example, if 20 students in a classroom were eligible to be tested in the fall but only 12 were tested in the fall and spring, each student who was tested in the fall and spring was assigned a weight of 1.67 (20/12). Similarly, for the cross-sectional sample, the number of students in each classroom who were tested in the spring were weighted up to the number of students in the classroom who were eligible to be tested in the spring.

No adjustment for nonresponse was included in the weights. Nonresponse rates for student testing were very low and did not differ by curriculum, as described above. In addition, the available characteristics (age, gender, race/ethnicity, LEP status, IEP status) of nonresponders did not differ from those of responders.

## **B. ESTIMATING CURRICULUM EFFECTS**

As described earlier, an experimental design was used to examine the relative effects of the study's four curricula on student math achievement. The design involved randomly assigning participating schools in each district to the study's four curricula. Because of random assignment, a simple and valid estimator of the relative effects of the curricula can be calculated by comparing the average spring math achievement of students in the four curriculum groups. Table D.3 presents average fall and spring math achievement of students in each curriculum group, and the average gain (spring minus fall) score for each group. However, the precision of

TABLE D.3

AVERAGE UNADJUSTED STUDENT MATH SCORES, BY CURRICULUM  
(Standard Deviations are in Parentheses)

Curriculum	Scale Score		
	Fall	Spring	Gain
Investigations	32.20 (8.73)	44.87 (8.64)	12.67 (6.06)
Math Expressions	29.94 (8.57)	45.45 (8.97)	15.51 (6.31)
Saxon	31.12 (8.64)	46.47 (7.62)	15.35 (6.82)
SFAW	30.89 (8.01)	44.28 (8.27)	13.39 (6.06)

Source: Author tabulations using data from the fall first grade ECLS-K math test administered by the study. The sample excludes 1 Math Expressions school (with 3 classrooms and 32 students) that participated during part of the school year and then stopped using the curriculum and did not allow the study to collect follow-up data.

these estimates can be increased by including in the analysis baseline values of measures that explain variation in the spring score. Also, when calculating the statistical significance of the results, the nested structure of the data must be incorporated into the calculations.

### Model for Estimating Curriculum Effects

A three-level hierarchical linear model (HLM) was used to estimate the relative effects of the study's curricula. For the longitudinal sample, the first (student) level of the HLM regressed the spring student scale score on the following student characteristics:

- **Fall score**—Student scale score on the fall assessment
- **Age**—Student age at the time of the fall assessment
- **Gender**—An indicator of whether the student is female
- **Race/ethnicity**—Indicators of whether the student is (1) Hispanic or (2) non-Hispanic black. Non-Hispanic white students and non-Hispanic students of other races serve as the reference category.
- **LEP/ELL**—Student is limited English proficient or an English language learner
- **IEP**—Student has an individualized education plan or service
- **Days between assessments**—The number of days between the student's fall and spring assessments

The second (classroom) level of the HLM regressed the intercept from the first-level equation on the following teacher characteristics:

- **Education**—Teacher has a master’s degree. Teachers who do not have a master’s degree, all of whom have a bachelor’s degree, serve as the reference category.
- **Experience**—Teacher experience, prior to the start of the school year, in years
- **Prior use of the assigned curriculum**—Teacher used the assigned curriculum at the K-3 level at some point before joining the study
- **Race**—Indicators of whether the teacher is black (white teachers and teachers of other races serve as the reference category) and whether the data were imputed. An indicator for imputed race was included because race information was missing for a larger fraction (13 percent) of teachers, than other teacher measures.
- **Class size**—Number of students in the classroom in the fall
- **Variance of the fall scale score for the classroom**—Calculated variance of the student scale score on the fall assessment for the classroom
- **Skewness of the fall scale score for the classroom**—Calculated skewness of the student scale score on the fall assessment for the classroom
- **Teacher assessment**—Teacher overall scale score on the assessment of math content and pedagogical knowledge

The third (school) level of the HLM regressed the intercept from the second-level equation on the following school characteristics:

- **Curricula**—Indicators of whether the school was assigned Investigations, Math Expressions, or Scott Foresman-Addison Wesley (SFAW). Schools assigned Saxon serve as the reference category.
- **Random assignment block**—Indicators for all but one of the blocks constructed for random assignment. Schools in the block without an indicator serve as the reference category.
- **Free/reduced-price meals**—The percentage of students eligible for free or reduced-price meals
- **Title I**—An indicator of whether the school was Title I

The same general model was estimated for the cross-sectional sample, but two measures—student age and class size—were constructed slightly differently. Student age was defined at the time of the spring assessment instead of the fall assessment, and class size was defined in the spring instead of the fall. In addition, the student-level weight for the cross-section sample was

constructed so that the students who were tested in the spring were weighted up to the number of students who were eligible to be tested, separately for each classroom.

### **Making Pair-Wise Comparisons**

With the four curricula included in the study, six unique pair-wise comparisons of effects 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. Because a Saxon indicator is not included in the model and thereby serves as the reference category, the coefficients on the Investigations, Math Expressions, and SFAW indicators indicate the effects of these curricula relative to Saxon. To make the pair-wise comparisons among Investigations, Math Expressions, and SFAW, the coefficients on the curriculum indicators are subtracted from one another. For example, to determine the effect of Investigations relative to Math Expressions, the coefficient on the Math Expressions indicator is subtracted from the coefficient on the Investigations indicator. Chapter III presents the results from the multiple curriculum comparisons, along with the statistical significance of each comparison.

To account for the multiple comparisons being made, the Tukey-Kramer method was used to adjust the estimated  $p$ -values. When performing several statistical tests, the chance of finding a significant effect that is actually due to chance increases. For example, with the four curriculum groups in this study, there are six unique pair-wise comparisons that can be made. If each comparison is made using a t-test with a 5 percent level of confidence, then the probability that one of those 6 tests will be statistically significant, even when there are no real differences between groups, could be as high as  $[1 - (1-0.05)^6] = 26$  percent. Put differently, the probability of mistakenly concluding that one curriculum is better than another is 26 percent, not the usual 5 percent. Tukey (1952) developed a method that specifically adjusts for pair-wise comparisons. The approach takes into account the dependencies between comparisons, while still maintaining a low probability of finding false effects. Tukey (1953) and Kramer (1956) independently developed a modification that is appropriate for unequal sample sizes.

### **Model Estimates Based on the Main (Longitudinal) Sample**

Table D.4 presents results based on the longitudinal sample for three specifications of the HLM: (1) a model that includes only the curriculum indicators and the block indicators used when conducting random assignment, (2) a model that adds to the first model the student fall score, and (3) a model that adds to the second model all the other student, teacher, and school controls. The results presented in the report are based on the third model. The pattern of results for the curriculum indicators is similar across the three model specifications. For each model, the table also presents the residual variances at the three levels (see the last three rows of the table).

The model was estimated with the SAS 9.1 software package, using the maximum likelihood estimation method of Proc Mixed. As a check, the model also was estimated with the HLM 6.06 software package and the results were consistent.

TABLE D.4

HIERARCHICAL LINEAR MODEL ESTIMATES FOR THE LONGITUDINAL SAMPLE  
(Outcome Is Spring Math Scale Score)

Variable Name	Model Using Only Block Dummies		Model Using Only Fall Scale Score		Full Model	
	Estimate	Standard Error	Estimate	Standard Error	Estimate	Standard Error
<b>Student Level</b>						
Intercept	48.06	0.98	25.92	0.80	57.16	18.28
Fall math scale score	--	--	0.69	0.02	0.67	0.02
Age at fall test	--	--	--	--	-0.82	0.39
Female	--	--	--	--	-0.08	0.30
Race/ethnicity						
Hispanic	--	--	--	--	-0.74	0.54
Non-Hispanic black	--	--	--	--	-0.96	0.57
LEP/ELL	--	--	--	--	-0.48	0.56
IEP/Special Services	--	--	--	--	-2.07	0.66
Days between assessments	--	--	--	--	-0.12	0.07
<b>Teacher Level</b>						
Master's degree	--	--	--	--	0.43	0.55
Experience	--	--	--	--	0.04	0.02
Prior use of the assigned curriculum	--	--	--	--	1.19	0.70
Race						
Black	--	--	--	--	-0.68	1.07
Race is imputed	--	--	--	--	0.39	0.61
Class size	--	--	--	--	0.16	0.08
Variance of the fall scale score	--	--	--	--	-0.01	0.01
Skewness of the fall scale score	--	--	--	--	-0.15	0.32
Teacher assessment overall score	--	--	--	--	-0.42	0.26
<b>School Level</b>						
Curricula						
Investigations	-1.37	1.15	-2.69	0.64	-2.49	0.62
Math Expressions	-0.25	1.18	0.05	0.65	0.18	0.61
SFAW	-1.95	1.12	-1.89	0.61	-1.93	0.70
Random assignment block						
Block 1	-4.59	1.66	-0.48	1.00	-2.99	2.88
Block 2	1.50	1.49	1.39	0.77	-2.35	2.71
Block 3	-2.13	1.47	-0.02	0.84	3.82	1.63
Block 4	-2.67	1.41	-0.87	0.79	2.22	1.53
Block 5	-6.09	1.74	-2.71	1.03	0.89	1.57
Block 6	-6.73	1.31	-3.69	0.73	-1.19	1.41
Block 7	-4.21	1.47	-4.65	0.83	-1.22	1.51
Free/reduced-price meals	--	--	--	--	-3.11	2.17
Title I	--	--	--	--	-0.96	0.66

TABLE D.4 (continued)

Variable Name	Model Using Only Block Dummies		Model Using Only Full Scale Score		Full Model	
	Estimate	Standard Error	Estimate	Standard Error	Estimate	Standard Error
<b>Residual Variance</b>						
Student Level	55.64		27.40		26.91	
Classroom Level	4.14		3.17		2.21	
School Level	3.06		0.08		0.00	

As mentioned earlier, model-based imputations were used to replace the small fraction of missing data with the predicted values of the measures from regression models based on the available data. Another approach would be to use multiple imputation techniques, which use a model-based approach as we did, but calculates a set of plausible values (as opposed to one value as we did) that represent the uncertainty about which value to impute. Model-based multiple imputations were not used as the primary imputation technique because it is extremely costly to implement the Tukey-Kramer method that adjusts for multiple comparisons when using multiple imputations. However, the HLM was estimated using the multiple imputation techniques, and conclusions based on the results from statistical tests of the curriculum effects are the same as those using the single imputation approach we used.

### Sensitivity Analyses

We explored whether the results are sensitive to (1) the specification of the HLM used to estimate effects, (2) the one (Math Expressions) school that stopped using the curriculum and did not allow spring testing of students and, therefore, had to be excluded from the analysis, and (3) the students that moved between study schools that used a different study curriculum.

***HLM Specification.*** The teacher assessment of math content and pedagogical knowledge (one of the controls included in the HLM) can be scored in four different ways. Item response theory (IRT) techniques can be used to create a single scale score based on all the items on the test, as well as two scale scores for each of the assessment's domains—content knowledge and pedagogical knowledge. Two separate HLMs were estimated with these scores, where one specification (which was reported in Chapter III) included the total scale score and the other included the two domain scale scores. To assess the sensitivity of using IRT techniques to calculate scores, two additional HLMs were estimated. One specification included the percentage of all items teachers answered correctly, and the other included two measures that decompose the score into the assessment's two domains—the percentage correct on the content items and the percentage correct on the pedagogical items.

Each of these four models was estimated with and without the student-level weight (for a total of eight models), to further assess the sensitivity of using the weight to calculate effects. Specifically, the following eight models were estimated:

1. Weighted with the overall scale score on the teacher assessment
2. Unweighted with the overall scale score on the teacher assessment
3. Weighted with the content knowledge scale score and pedagogical knowledge scale score on the teacher assessment
4. Unweighted with the content knowledge scale score and pedagogical knowledge scale score on the teacher assessment
5. Weighted with the overall percentage correct on the teacher assessment
6. Unweighted with the overall percentage correct on the teacher assessment
7. Weighted with the content knowledge percent correct and pedagogical knowledge percent correct on the teacher assessment
8. Unweighted with the content knowledge percent correct and pedagogical knowledge percent correct on the teacher assessment

Results for all of the models specified were very similar and showed nearly identical relative effects of the curricula.

***No Outcome Data for One School.*** We also explored whether the results are affected by the one (Math Expressions) school that stopped using the curriculum and did not allow spring testing of students and, therefore, had to be excluded from the analysis.<sup>66</sup> This issue was examined using two analysis. First, we examined whether students in the one Math Expressions school that did not allow spring testing are different from those in all the other Math Expressions schools that did allow spring testing. Ideally, the students in the one school are a random sample of students in all the Math Expressions schools. Since fall (baseline) achievement is a strong predictor of spring (follow-up) achievement, we made this assessment by comparing fall achievement of students in the Math Expressions school that dropped out of the study with students in the Math Expressions schools that stayed in.

This sensitivity analysis indicates that the results are not biased because outcome data were not available for one of the Math Expressions schools. Average fall achievement of dropouts (32.04) and stayers (29.81) is not significantly different. This analysis included 32 dropouts and 314 stayers. Because this sensitivity analysis is based on small sample sizes and therefore has

---

<sup>66</sup> As mentioned in Appendix A, one of the schools assigned to Math Expressions indicated after a few months of implementation that it was going to stop using its assigned curriculum and that it would not allow the study to test students in the spring. Because spring achievement is the outcome used to assess the relative effects of the curricula, the school—which contained 3 teachers and 32 students sampled for testing—had to be excluded from the analysis. A frequently used approach to address this type of data collection issue is to incorporate an adjustment for outcome nonresponse in the analysis. In practice, this approach accounts for students without outcome data by overweighting those with outcomes who have similar baseline characteristics. We did not use this approach because a fully justifiable nonresponse adjustment could not be calculated. In particular, students with outcome data who have similar baseline characteristics as those who could not be tested exist in the sample, but not in Math Expressions schools that possess similar school characteristics as the one that did not allow spring testing.

little statistical power, we further investigated the effects of no outcome data for the one Math Expressions school using a second sensitivity analysis.

The second sensitivity analysis exploits a property of random assignment. Because of random assignment, we can assume that the schools assigned to each of the curriculum groups are identical, within a known degree of statistical precision. Since one of the schools assigned to Math Expressions stopped using the curriculum and did not allow the study team to test students in the spring, it implies that one school in each of the other groups would have performed the same had they been assigned to Math Expressions. If we could identify those schools, we could exclude them from the analysis and recalculate the results. Since we cannot identify those schools, an alternative approach is to recalculate the results with two samples, one that excludes the lowest gaining Investigations, Saxon, and SFAW schools and another that excludes the highest gaining school in each of those curriculum groups. These two sets of results represent the upper and lower bound on the single set of results that we would calculate if we could identify the correct Investigations, Saxon, and SFAW schools to exclude from the analysis.

The pattern of results is robust to this sensitivity analysis. Table III.3 (in Chapter III) showed that both the Math Expressions-Investigations and Saxon-Investigations differentials equal 0.30 effect sizes. Results based on the sensitivity analysis described above indicate that these differentials lie between 0.27 and 0.33 effect sizes. Table III.3 also showed that both the Math Expressions-SFAW and Saxon-SFAW differentials equal 0.24 effect sizes. The sensitivity analysis indicates that these differentials lie between 0.17 and 0.25 effect sizes.

***The Small Number of Students That Crossed Over to Another Study Curriculum.*** Last, the results are not affected by “crossover.” In a study of this kind, where study schools in each district are using four curricula, the possibility exists that students move between study schools with different curricula during the school year. Five of the 1,309 students included in the longitudinal sample were in different study schools with different curricula between fall and spring testing. Analytic techniques can be used to correct results for crossovers, but those techniques cannot be used in this setting because the number of crossovers is too low to support the analysis. To explore whether the results are affected by the crossovers, we deleted them from the sample and reestimated the model. The results are identical to those reported in Table III.3.

### **Model Estimates Based on the Cross-Sectional Sample**

The results for the three-level HLM based on the cross-sectional sample are shown in Table D.5. The magnitude of the results for each unique pair-wise curriculum comparison that can be made, as shown in Table D.6, were comparable to those for the longitudinal sample. Average math achievement of Math Expressions and Saxon students was 0.30 to 0.32 standard deviations higher than Investigations and SFAW students. Also, the average adjusted spring scores of the two more effective curricula (Math Expressions and Saxon) are not significantly different, nor are the average adjusted scores of the two less effective curricula (Investigations and SFAW).

TABLE D.5

HIERARCHICAL LINEAR MODEL ESTIMATES FOR THE CROSS-SECTIONAL SAMPLE  
(Outcome Is Spring Math Scale Score)

Variable Name	Full Model	
	Estimate	Standard Error
<b>Student Level</b>		
Intercept	51.24	19.33
Fall math scale score	0.64	0.02
Age at spring test	-0.43	0.40
Female	-0.13	0.30
Race/ethnicity		
Hispanic	-1.48	0.52
Non-Hispanic black	-1.88	0.58
LEP/ELL	-0.27	0.57
IEP/Special Services	-2.12	0.65
Days between assessments	-0.10	0.08
<b>Teacher Level</b>		
Master's degree	0.00	0.58
Experience	0.05	0.02
Prior use of the assigned curriculum	0.91	0.78
Race		
Black	-0.33	1.00
Race is imputed	1.02	0.67
Class size	0.23	0.08
Variance of the fall scale score	-0.01	0.01
Skewness of the fall scale score	-0.38	0.37
Teacher assessment overall score	-0.31	0.27
<b>School Level</b>		
Curricula		
Investigations	-2.71	0.65
Math Expressions	0.07	0.65
SFAW	-2.51	0.73
Random assignment block		
Block 1	-3.50	3.03
Block 2	-2.77	2.86
Block 3	2.06	1.72
Block 4	1.68	1.61
Block 5	0.54	1.66
Block 6	-1.11	1.47
Block 7	-1.03	1.57
Free/reduced-price meals	-1.80	2.31
Title I	-1.47	0.68
<b>Residual Variance</b>		
Student Level	31.22	
Classroom Level	2.56	
School Level	0.00	

TABLE D.6

AVERAGE DIFFERENCE BETWEEN PAIRS OF CURRICULA IN HLM-ADJUSTED  
 SPRING STUDENT MATH ACHIEVEMENT FOR THE  
 CROSS-SECTIONAL SAMPLE, IN EFFECT SIZES  
 (*p*-values Are in Parentheses)

	Effect of					
	Investigations relative to			Math Expressions relative to		Saxon relative to
	Math Expressions	Saxon	SFAW	Saxon	SFAW	SFAW
Effect Size	-0.31*	-0.32*	-0.02	0.01	0.30*	0.31*
<i>p</i> -value	(0.00)	(0.00)	(0.90)	(1.00)	(0.00)	(0.01)

Source: Author tabulations using data from the spring first grade ECLS-K math test administered by the study, schools records, fall 2006 teacher survey, and school-level data from the 2003-2004 Common Core of Data and [www.GreatSchools.net](http://www.GreatSchools.net). The sample excludes 1 Math Expressions school (with 3 classrooms and 32 students) that participated during part of the school year and then stopped using the curriculum and did not allow the study to collect follow-up data.

Note: The results were produced using a three-level hierarchical linear model (see text for details about the model). The Tukey-Kramer method was used to adjust the *p*-values for the six unique pair-wise curriculum comparisons that can be made.

\*Statistically significant at the 5 percent level.

**Subgroup Analyses**

As described earlier, subgroup analyses were conducted to examine whether curriculum effects differ along six characteristics: (1) participating districts, (2) school fall achievement, (3) school free/reduced-price meals eligibility, (4) teacher education, (5) teacher experience, and (6) teacher math content/pedagogical knowledge.

Table D.7 presents school, teacher, and student sample sizes for each of the subgroups, along with the average value of the characteristic used to define each subgroup. For example, the cell for the “lowest third school fall achievement” subgroup indicates the average value of school fall achievement for the schools included in that subgroup. The table also presents the minimum detectable effect size for each subgroup. The effect sizes were calculated as described in Chapter I using the sample sizes reported Table D.7 and assuming that the sample is distributed evenly across the curricula.

TABLE D.7

## SAMPLE SIZES USED IN SUBGROUPS ANALYSES

Subgroup	Average Value of Subgroup Characteristic	Sample Size			Minimum Detectable Effect Size Between Any Pair of Curricula
		Schools	Teachers	Students	
Participating Districts				--	
District #1	--	7	22	232	0.94
District #2	--	8	23	212	0.81
District #3	--	12	34	348	0.52
District #4	--	12	52	517	0.43
School Fall Achievement <sup>a</sup>					
Lowest third	26.32	13	36	378	0.49
Middle third	30.22	13	43	411	0.47
Highest third	35.02	13	52	520	0.42
School Free/Reduced-Price Meals Participation					
Up to 40% eligibility	12.24%	17	72	712	0.34
Greater than 40% eligibility	70.32%	22	59	597	0.36
Teacher Education					
Bachelor's degree	--	26	43	429	0.41
Master's degree	--	35	88	880	0.28
Teacher Experience					
Up to 5 years	2.48	26	51	492	0.38
Greater than 5 years	17.70	36	80	817	0.29
Teacher Math Content/Pedagogical Knowledge <sup>a</sup>					
1st (lowest) quintile	-1.18	21	26	255	0.53
2nd through 5th quintiles	0.20	38	105	1,054	0.26

<sup>a</sup>School Fall Achievement and Teacher Math Content/Pedagogical Knowledge are expressed in scale score units.

Separately for each characteristic, the HLM estimated for the longitudinal sample was modified to include interactions between the curriculum indicators and the characteristic. For example, to examine whether curriculum effects differ along teacher education, the model was expanded to include eight third-level interactions:

1. Investigations interacted with teachers who had a master's degree
2. Investigations interacted with teachers who did not have a master's degree
3. Math Expressions interacted with teachers who had a master's degree
4. Math Expressions interacted with teachers who did not have a master's degree
5. SFAW interacted with teachers who had a master's degree
6. SFAW interacted with teachers who did not have a master's degree
7. Saxon interacted with teachers who had a master's degree
8. Saxon interacted with teachers who did not have a master's degree (serves as the reference category)

Similar models were used for the other characteristics.

Pair-wise comparisons to determine the relative curriculum effects for each subgroup were made using the process described earlier. If a subgroup had two levels, twelve pair-wise comparisons were made. For example, to examine if curriculum effects differ along teacher experience, the following pair-wise comparisons were made:

- Investigations among teachers who had eight or fewer years of experience relative to Math Expressions among teachers who had eight or fewer years of experience
- Investigations among teachers who had eight or fewer years of experience relative to SFAW among teachers who had eight or fewer years of experience
- Investigations among teachers who had eight or fewer years of experience relative to Saxon among teachers who had eight or fewer years of experience
- Math Expressions among teachers who had eight or fewer years of experience relative to SFAW among teachers who had eight or fewer years of experience
- Math Expressions among teachers who had eight or fewer years of experience relative to Saxon among teachers who had eight or fewer years of experience
- SFAW among teachers who had eight or fewer years of experience relative to Saxon among teachers who had eight or fewer years of experience
- Investigations among teachers who had more than eight years of experience relative to Math Expressions among teachers who had more than eight years of experience

- Investigations among teachers who had more than eight years of experience relative to SFAW among teachers who had more than eight years of experience
- Investigations among teachers who had more than eight years of experience relative to Saxon among teachers who had more than eight years of experience
- Math Expressions among teachers who had more than eight years of experience relative to SFAW among teachers who had more than eight years of experience
- Math Expressions among teachers who had more than eight years of experience relative to Saxon among teachers who had more than eight years of experience
- SFAW among teachers who had more than eight years of experience relative to Saxon among teachers who had more than eight years of experience

As described earlier, the Tukey-Kramer method was used to adjust the estimated  $p$ -values for the multiple comparisons being made. Chapter III presents the results from the multiple curriculum comparisons made for each subgroup, along with the statistical significance of each comparison.

