What Works Clearinghouse™

INSTITUTE OF EDUCATION SCIENCES

Elementary School Mathematics Updated May 2013

Scott Foresman–Addison Wesley Elementary Mathematics

Report Contents

Program Description[1]

Scott Foresman–Addison Wesley Elementary Mathematics is a core mathematics curriculum for students in prekindergarten through grade 6. The program aims to improve students' understanding of key math concepts through problem-solving instruction, hands-on activities, and math problems that involve reading and writing. The curriculum focuses on problem-solving skills, assessments, and exercises tailored to students of different ability levels. According to its developer, *Scott Foresman–Addison Wesley Elementary Mathematics* is aligned to the National Council of Teachers of Mathematics (NCTM) standards for the elementary grades.

Research[2]

The What Works Clearinghouse (WWC) identified three studies of *Scott Foresman–Addison Wesley Elementary Mathematics* that both fall within the scope of the Elementary School Mathematics topic area and meet WWC evidence standards.[3] All three studies meet WWC evidence standards without reservations, and together, they included 9,547 elementary students from grades 1–5 in 120 schools. These schools were located in a mix of urban, suburban, and rural settings in 15 states.

The WWC considers the extent of evidence for *Scott Foresman–Addison Wesley Elementary Mathematics* on the math performance of elementary school students to be medium to large for the mathematics achievement domain, the only domain examined for studies reviewed under the Elementary School Mathematics topic area.

Effectiveness

Scott Foresman–Addison Wesley Elementary Mathematics was found to have mixed effects on mathematics achievement for elementary school students.

Table 1. Summary of findings[4]

Outcome domain	Rating of effectiveness	Improvement index *(percentile points)*		Number of studies	Number of students	Extent of evidence
		Average	Range			
Mathematics achievement	Mixed effects	–2	–7 to +6	3	9,547	Medium to large

Program Information

Background

Scott Foresman–Addison Wesley Elementary Mathematics was developed and is distributed by Pearson Scott Foresman, a division of Pearson Education, Inc. Address: One Lake Street, Upper Saddle River, NJ 07458. Email: communications@pearsoned.com. Web: http://www.pearsonschool.com. Telephone: (201) 236-7000.

Program details

Scott Foresman–Addison Wesley Elementary Mathematics consists of teacher-led lessons that follow a check-learn-check-practice sequence, emphasizing key math concepts and skills. Teachers check students' skills prior to each lesson, introduce the lesson, and then check students' understanding during the lesson. Practice sections in the text permit students to further demonstrate their understanding of concepts and apply this knowledge to solving real-life problems. Lessons are typically 45–60 minutes in length and are organized into chapters. Each chapter extends over 2–8 weeks and uses texts, workbooks, transparencies, manipulatives, and technology through group and individual activities.

Cost

The cost of *Scott Foresman–Addison Wesley Elementary Mathematics* varies based on the grade and number of components included. Current prices for a single student edition textbook are $26.47 for kindergarten, $36.97 for grades 1–2, and $65.97 for grades 3–6. Student workbooks range from $3.97 to $7.47. A single teacher's edition textbook costs $209.97, and a manipulatives kit costs up to $434.97, depending on the contents.

Research Summary

The WWC identified 13 studies that investigated the effects of *Scott Foresman–Addison Wesley Elementary Mathematics* on the math performance of elementary school students.

The WWC reviewed seven of these studies against group design evidence standards. Three studies (Agodini, Harris, Thomas, Murphy, & Gallagher, 2010; Resendez & Azin, 2006; Resendez & Manley, 2005[6]) are randomized controlled trials that meet WWC evidence standards without reservations. These three studies are summarized in this report. Four studies do not meet WWC evidence standards.

The remaining six studies do not meet WWC eligibility screens for review in this topic area. Citations for all 13 studies are in the References section, which begins on p. 5.

Table 2. Scope of reviewed research[5]

Grade	1, 2, 3, 4, 5
Delivery method	Whole class
Program type	Curriculum

Summary of studies meeting WWC evidence standards without reservations

Agodini et al. (2010) presented results for 110 elementary schools that had been randomly assigned to one of four conditions: *Investigations in Number, Data, and Space®* (28 schools), *Math Expressions* (27 schools), *Saxon Math* (26 schools), and *Scott Foresman–Addison Wesley Elementary Mathematics* (29 schools). The analysis included 4,716 first-grade students and 3,344 second-grade students who were evenly divided among the four conditions. The study authors compared average spring math achievement of students in each condition after 1 school year of program implementation. Student outcomes were measured by the Early Childhood Longitudinal Study–Kindergarten (ECLS-K) math assessment.

Resendez and Azin (2006) randomly assigned 39 teachers of third- and fifth-grade students to either *Scott Foresman–Addison Wesley Mathematics* (20 teachers) or a comparison condition (19 teachers). The analysis included 837 to 863 students, depending on the outcome measure used, in the 39 classrooms.[7] The comparison curricula included two distinct basal curricula and a school-created math program that was based on a number of different math materials from various resources. The study compared average student math achievement outcomes of classrooms in the intervention condition with those of the comparison condition after 1 year of program implementation.

Resendez and Manley (2005) conducted a randomized controlled trial in which 35 teachers of second- and fourth-grade students were randomly assigned to either *Scott Foresman–Addison Wesley Elementary Mathematics* (18 teachers) or a comparison condition (17 teachers) using one of five different elementary math programs. The analysis included 491 to 624 students, depending on the outcome measure used. The teachers in the intervention condition were in their first year of implementing the *Scott Foresman–Addison Wesley Elementary Mathematics* program. The comparison programs included chapter-based basal curricula and strand/module-based investigative curricula. The study compared math achievement outcomes of students in the intervention condition with those of the comparison condition after 1 year of program implementation.

Effectiveness Summary

The WWC review of *Scott Foresman–Addison Wesley Elementary Mathematics* for the Elementary School Mathematics topic includes student outcomes in one domain: mathematics achievement. The findings below present the authors' estimates and WWC-calculated estimates of the size and statistical significance of the effects of *Scott Foresman–Addison Wesley Elementary Mathematics* on the outcomes of elementary school students. For a more detailed description of the rating of effectiveness and extent of evidence criteria, see the WWC Rating Criteria on p. 17.

Summary of effectiveness for the mathematics achievement domain

Three studies that meet WWC standards without reservations reported findings in the mathematics achievement domain.

Agodini et al. (2010) reported, and the WWC confirmed, statistically significant negative effects of the *Scott Foresman–Addison Wesley Elementary Mathematics* program on the Early Childhood Longitudinal Study–Kindergarten (ECLS-K) Math Assessment when compared to *Math Expressions* or *Saxon Math* in second grade. The study reported a statistically significant negative effect of the *Scott Foresman–Addison Wesley Elementary Mathematics* program on the ECLS-K Math Assessment when compared to *Math Expressions* in first grade, but the WWC found it no longer significant after adjusting for multiple comparisons. The study reports no significant effects of *Scott Foresman–Addison Wesley Elementary Mathematics* on the ECLS-K Math Assessment when compared to *Investigations in Number, Data, and Space*®. The average effect size across the curricula and two grades was not large enough to be considered substantively important according to WWC criteria (i.e., an effect size of at least 0.25). The WWC characterizes these study findings as a statistically significant negative effect.

Resendez and Azin (2006) reported no statistically significant effects of the *Scott Foresman–Addison Wesley Elementary Mathematics* program on either the TerraNova Comprehensive Tests of Basics Skills (CTBS) Basic Multiple Assessment (Math Total) or the TerraNova CTBS Basic Multiple Assessment Plus (Math Computation) scores. The average effect across the two outcome measures was not large enough to be considered substantively important according to WWC criteria. The WWC characterizes these study findings as an indeterminate effect.

Resendez and Manley (2005) reported no statistically significant effects of the *Scott Foresman–Addison Wesley Elementary Mathematics* program on either the TerraNova Comprehensive Tests of Basics Skills (CTBS) Basic Multiple Assessment (Math Total) or the TerraNova CTBS Basic Multiple Assessment Plus (Math Computation) scores. The average effect across the two outcome measures was not large enough to be considered substantively important according to WWC criteria. The WWC characterizes these study findings as an indeterminate effect.

Thus, for the mathematics achievement domain, one study showed statistically significant negative effects and two studies showed indeterminate effects. This results in a rating of mixed effects, with a medium to large extent of evidence.

Table 3. Rating of effectiveness and extent of evidence for the mathematics achievement domain

Rating of effectiveness	Criteria met
Mixed effects *Evidence of inconsistent effects.*	In the three studies that reported findings, the estimated impact of the intervention on outcomes in the *mathematics achievement* domain was negative and statistically significant in one study and indeterminate in two studies.
Extent of evidence	**Criteria met**
Medium to large	Three studies that included 9,547 students in 120 schools reported evidence of effectiveness in the *mathematics achievement* domain.

References

Studies that meet WWC evidence standards without reservations

Agodini, R., Harris, B., Thomas, M., Murphy, R., & Gallagher, L. (2010). *Achievement effects of four early elementary school math curricula: Findings for first and second graders* (NCEE 2011-4001). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/pubs/20114001/pdf/20114001.pdf

Additional sources:

Agodini, R., & Harris, B. (2010). An experimental evaluation of four elementary school math curricula. *Journal of Research on Educational Effectiveness, 3*(3), 199–253.

Agodini, R., Harris, B., Atkins-Burnett, S., Heaviside, S., Novak, T., & Murphy, R. (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.

Agodini, R., Harris, B., Atkins-Burnett, S., Heaviside, S., Novak, T., Murphy, R., & Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. (2009). *Achievement effects of four early elementary school math curricula: Findings from first graders in 39 schools*. Executive summary (NCEE 2009-4053). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Resendez, M., & Azin, M. (2006). *2005 Scott Foresman–Addison Wesley Elementary Math randomized control trial: Final report*. Jackson, WY: PRES Associates, Inc.

Additional source:

Resendez, M., & Sridiharan, S. (2006). *A study on the effectiveness of the 2005 Scott Foresman Addison–Wesley Elementary Math program: Technical report*. Jackson, WY: PRES Associates, Inc.

Resendez, M., & Manley, M. A. (2005). *Final report: A study on the effectiveness of the 2004 Scott Foresman–Addison Wesley Elementary Math program*. Jackson, WY: PRES Associates, Inc.

Additional source:

Resendez, M., & Sridharan, S. (2005). *Technical report: A study on the effectiveness of the 2004 Scott Foresman–Addison Wesley Elementary Math program*. Jackson, WY: PRES Associates, Inc.

Studies that meet WWC evidence standards with reservations

None

Studies that do not meet WWC evidence standards

Bhatt, R., & Koedel, C. (2012). Large-scale evaluations of curricular effectiveness: The case of elementary mathematics in Indiana. *Educational Evaluation and Policy Analysis, 34*(4), 391–412. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

Cambium Learning, Inc. (2006). *An evaluation of Voyages Mathematics, Fairview Public Schools 2005–2006: Technical report*. Natick, MA: Author. The study does not meet WWC evidence standards because the measures of effectiveness cannot be attributed solely to the intervention—there was only one unit assigned to one or both conditions.

Cummins-Colburn, B. J. L. (2007). Differences between state-adopted textbooks and student outcomes on the Texas Assessment of Knowledge and Skills examination. *Dissertation Abstracts International, 68*(06A), 168-2299. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

Gatti, G. G. (2004). *Scott Foresman–Addison Wesley Math national effect size study*. Upper Saddle River, NJ: Pearson Education. The study does not meet WWC evidence standards because it only includes outcomes that are overaligned with the intervention or measured in a way that is inconsistent with the protocol.

Studies that are ineligible for review using the Elementary School Mathematics Evidence Review Protocol

Jitendra, A. K., Deatline-Buchman, A., & Sczesniak, E. (2005). A comparative analysis of third-grade mathematics textbooks before and after the 2000 NCTM standards. *Assessment for Effective Intervention, 30*(2), 47–62. The study is ineligible for review because it does not examine the effectiveness of an intervention.

Klein, D. (2000). *High achievement in mathematics: Lessons from three Los Angeles elementary schools*. Washington, DC: Brookings Institution Press. The study is ineligible for review because it does not use a comparison group or single-case design.

Simpson, N. (2001). *Scott Foresman California Mathematics validation study pretest–posttest results*. Sacramento, CA: Pearson Scott Foresman. The study is ineligible for review because it does not use a comparison group or single-case design.

Slavin, R. E., & Lake, C. (2007). Effective programs in elementary mathematics: A best-evidence synthesis. *The Best Evidence Encyclopedia, 1*(2). Retrieved from http://www.bestevidence.org/word/elem_math_Feb_9_2007.pdf The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Additional source:

Effective programs for elementary mathematics: A best evidence synthesis. Educator's summary. (2009). Retrieved from http://www.bestevidence.org/word/elem_math_Mar_11_2009_sum.pdf

Triantos, L. M. (2005). *The aftermath of implementing a standards-based curriculum in a K-8 district: Is there a correlation between hands-on instruction and math scores?* (Unpublished master's thesis). Rowan University, Glassboro, NJ. The study is ineligible for review because it does not use a comparison group design or a single-case design.

WESTAT. (2003). *Analysis of field testing for Scott Foresman–Addison Wesley Mathematics 2004*. Rockville, MD: Author. The study is ineligible for review because it does not include a student outcome.

Appendix A.1: Research details for Agodini et al. (2010)

Agodini, R., Harris, B., Thomas, M., Murphy, R., & Gallagher, L. (2010). *Achievement effects of four early elementary school math curricula: Findings for first and second graders* (NCEE 2011-4001). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/pubs/20114001/pdf/20114001.pdf

Table A1. Summary of findings — **Meets WWC evidence standards without reservations**

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
Mathematics achievement	110 schools/8,060 students	–4	Yes

Setting

The study took place in elementary schools in 12 districts across 10 states, including Connecticut, Florida, Kentucky, Minnesota, Mississippi, Missouri, Nevada, New York, South Carolina, and Texas. Of the 12 districts, three were in urban areas, five were in suburban areas, and four were in rural areas.

Study sample

Following district and school recruitment and collection of consent from all teachers in the participating grades, 111 participating schools were randomly assigned to one of four curricula: (a) *Investigations in Number, Data, and Space*®, (b) *Math Expressions*, (c) *Saxon Mat*h, and (d) *Scott Foresman–Addison Wesley Mathematics*. Blocked random assignment of the schools was conducted separately within each district. In each district, participating schools were grouped together into blocks of four to seven schools based on characteristics such as Title I eligibility, free or reduced-price lunch eligibility status, grade enrollment size, math proficiency, and proportion of White and Hispanic students. Two districts had an additional blocking variable (magnet school status in one district and year-round school schedule in another district). One district required that all schools that fed into the same middle school receive the same condition. Schools in each block were randomly assigned among the four curricula. On average, 11 students were randomly sampled from each participating classroom for assessment. One school with three teachers and 32 students assigned to *Math Expressions* withdrew from the study and did not permit follow-up data collection.

The analysis sample included a total of 110 schools, 461 first-grade classrooms, 4,716 first graders, 328 second-grade classrooms, and 3,344 second graders. In the first grade sample, on average, 27 schools, 116 classrooms, and 1,180 students were assigned to each condition. In the second grade sample, on average, 18 schools, 82 classrooms, and 835 students were assigned to each condition.

Seventy-six percent of the schools in the study were eligible for Title I funding. Approximately half of the students in the sample were eligible for free or reduced-price lunch. Among students in the sample, 39% were White, 32% were non-Hispanic Black, 26% were Hispanic, 2% were Asian, and 1% were American Indian or Alaskan Native.

Intervention group

Students in the intervention group used *Scott Foresman–Addison Wesley Mathematics* as their core math curriculum. Study authors reported about nine out of 10 teachers self-reported completing at least 80% of the curriculum.

Comparison group

The study included three comparison groups: (a) *Investigations in Number, Data, and Space®*, (b) *Math Expressions*, and (c) *Saxon Math*. Each curriculum was implemented by comparison teachers for 1 school year.

Investigations in Number, Data, and Space® is published by Pearson Scott Foresman. It uses a student-centered approach that encourages reasoning and understanding and draws on constructivist learning theory. The lessons build on students' existing knowledge and focus on understanding math concepts rather than simply learning computational methods. The curriculum is organized in nine thematic units, each lasting 2–5.5 weeks. Study authors reported that about four out of five teachers self-reported completing at least 80% of the curriculum.

Math Expressions is published by Houghton Mifflin Harcourt and uses a blend of student-centered and teacher-directed instructional approaches. Students using the curriculum question and discuss mathematics and are explicitly taught problem solving strategies. There is an emphasis on using multiple specified objects, drawings, and language to represent concepts, and on learning through the use of real-world situations. Students are expected to explain and justify their solutions. Study authors reported that about nine out of 10 teachers self-reported completing at least 80% of the curriculum.

Saxon Math is published by Houghton Mifflin Harcourt and uses a teacher-directed approach that offers a script for teachers to follow in each lesson. It blends teacher-directed instruction of new material with daily practice of previously learned concepts and procedures. The teacher introduces concepts or efficient strategies for solving problems. Students receive instruction from the teacher, participate in guided practice, and then undertake individual practice. Frequent monitoring of student achievement is built into the program. Daily routines are extensive and emphasize practice of number concepts and use of methods (such as the use of number lines, counting on fingers, and diagrams) to represent mathematical concepts. Study authors reported that about six out of seven teachers self-reported completing at least 80% of the curriculum.

Outcomes and measurement

Mathematics achievement was measured using the mathematics assessment developed for the ECLS-K class of 1998–99. The assessment is individually administered, nationally normed, and adaptive. The assessment meets accepted standards of validity and reliability. Scale scores from an item response theory (IRT) model were used in the analysis. The test was administered in the fall of the implementation year (within 4 weeks of the first day of classes) to assess students' baseline math achievement. The test was also administered in the spring—that is, from 1–6 weeks before the end of the school year of program implementation. For a more detailed description of the outcome measure, see Appendix B.

Support for implementation

Teachers in all four groups were provided training by the curriculum publisher. Teachers assigned to *Scott Foresman–Addison Wesley Elementary Mathematics* received 1 day of initial training in the summer before the school year began. Follow-up training was offered about every 4–6 weeks throughout the school year. Follow-up sessions were typically 3–4 hours long and held after school.

Teachers assigned to *Investigations in Number, Data, and Space®* (comparison group 1) were provided 1 day of initial training in the summer before the school year began. Follow-up sessions were typically 3–4 hours long and held after school.

Teachers assigned to *Math Expressions* (comparison group 2) were provided 2 days of initial training in the summer before the school year began. Two follow-up trainings were offered during the school year. Follow-up sessions typically consisted of classroom observations followed by short feedback sessions with teachers.

Teachers assigned to *Saxon Math* (comparison group 3) were provided 1 day of initial training in the summer before the school year began. One follow-up training session, tailored to meet each district's needs, was offered during the school year.

Appendix A.2: Research details for Resendez & Azin (2006)

Resendez, M., & Azin, M. (2006). ***2005 Scott Foresman–Addison Wesley Elementary Math randomized control trial: Final report*****. Jackson, WY: PRES Associates, Inc.**

Table A2. Summary of findings — **Meets WWC evidence standards without reservations**

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
Mathematics achievement	39 classrooms/863 students	+3	No

Setting

Four schools located in urban and suburban settings participated in the study. Two schools were located in Ohio and two schools were located in New Jersey.

Study sample[8]

Third- and fifth-grade teachers were randomly assigned within schools to either the intervention or comparison condition. The baseline sample included 39 teachers (20 intervention and 19 comparison) and 915 students (468 intervention and 447 comparison). Twenty-three teachers taught third grade (13 intervention and 10 comparison), and 16 taught fifth grade (seven intervention and nine comparison). No teachers left the study, and student attrition was low. Between 837 and 863 students were tested at the end of the school year on the TerraNova Comprehensive Tests of Basics Skills (CTBS) Basic Multiple Assessment (Math Total) and TerraNova CTBS Basic Multiple Assessment Plus (Math Computation).[9] On average, participating schools had a lower percentage of Hispanic and African-American students, special education students, and students eligible for free or reduced-price meals than the national average. These schools had higher average percentages of Asian students and students with higher ability levels than the national average.

Intervention group

Students used the 2005 *Scott Foresman–Addison Wesley Elementary Mathematics* curriculum during the 2005–06 school year. The program was implemented according to the curricula guidelines. Implementation was monitored throughout the school year using online teacher logs and classroom observation. The study authors reported that teachers covered 79% of the curriculum on average.

Comparison group

Comparison students used three different math curricula. Students in two schools used a chapter-based, comprehensive basal program. Students in a third school used a different basal math program that placed greater emphasis on repetitive, sequential review and regular assessments. Students in a fourth school used a school-created math program that was based on a number of different math materials from various resources. The comparison curricula generally covered the same content as *Scott Foresman–Addison Wesley Elementary Mathematics*. The study authors reported that teachers covered 80% of the curricula on average.

Outcomes and measurement

The authors administered the TerraNova Basic Multiple Assessment Plus test (Level 13 in third grade and Level 15 in fifth grade). The math test provides two overall scores: the TerraNova Comprehensive Tests of Basics Skills (CTBS) Basic Multiple Assessment (Math Total) and the TerraNova CTBS Basic Multiple Assessment Plus (Math Computation) Total. The Math Total score is based on multiple choice and constructed response items that are predominantly word problems that measure basic, applied, and higher-order thinking skills. The Math Computation Total is based on the Plus test booklet, which contains only multiple-choice computational problems. Scale scores were used in the analysis. For a more detailed description of these outcome measures, see Appendix B.

Support for implementation

Teachers received 3 hours of initial training prior to implementing *Scott Foresman–Addison Wesley Elementary Mathematics* in their classes. At the initial training session, the trainer described the key components of the curriculum, reviewed the teacher's edition textbook and available ancillary resources, offered examples of when to use certain materials, provided an overview of the math technology available, and modeled a math lesson. The training focused on the components most vital to the program and those that were required for full implementation.

Two follow-up sessions were offered during the school year. The first was offered 4–8 weeks into the school year and lasted 2 hours. The session was informal and allowed teachers to discuss and ask questions about implementation issues. A second follow-up session, addressing pacing issues and further covering the technology available with the program, was provided to one school in March. The other three schools were offered the second follow-up session but chose not to receive it.

Appendix A.3: Research details for Resendez & Manley (2005)

Resendez, M., & Manley, M. A. (2005). ***Final report: A study on the effectiveness of the 2004 Scott Foresman–Addison Wesley Elementary Math program*****. Jackson, WY: PRES Associates, Inc.**

Table A3. Summary of findings **Meets WWC evidence standards without reservations**

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
Mathematics achievement	35 classrooms/624 students	–2	No

Setting

This study took place in six elementary schools in four states: Kentucky (two suburban schools), Virginia (one urban school), Washington (one urban school), and Wyoming (one rural and one suburban school).

Study sample

Second- and fourth-grade teachers were randomly assigned within schools to the intervention using *Scott Foresman–Addison Wesley Elementary Mathematics*. The baseline sample included 35 teachers (18 intervention and 17 comparison) and 742 students (389 intervention and 353 comparison). Of the 35 study teachers, 19 taught second grade (10 intervention and nine comparison) and 16 taught fourth grade (eight intervention and eight comparison). The analysis samples included 35 teachers (18 intervention and 17 comparison). The TerraNova CTBS Basic Multiple Assessment Plus (Math Computation) analysis sample included 491 students (264 intervention and 227 comparison) whereas the TerraNova CTBS Basic Multiple Assessment (Math Total) analysis sample included 624 students (347 intervention and 277 comparison). About one-third of participating students were minorities. At two of the six participating schools, more than 90% of students were eligible for free or reduced-price meals. The percentage of students eligible for free or reduced-price meals at the other four schools was similar to the national average of 37%.

Intervention group

Students in the intervention group used the 2004 *Scott Foresman–Addison Wesley Elementary Mathematics* curriculum during the 2004–05 school year. The teachers in the intervention group were implementing the intervention curriculum for the first time. The study authors reported that teachers covered 70% of the curriculum on average.

Comparison group

Students in the comparison group used five different comprehensive math curricula. These curricula are not identified in the study, but the study authors report that the comparison curricula covered the same content as *Scott Foresman–Addison Wesley Elementary Mathematics*. The study authors reported that teachers covered 75% of the curricula on average.

Outcomes and measurement

The primary outcome measure was the CTBS, Basic Multiple Assessment Plus test. The authors describe the TerraNova CTBS as a reliable, standardized test consisting of multiple-choice, constructed response, and computational problems. According to the authors, it offers broad coverage of mathematics content in most textbooks and reflects NCTM standards. The assessment provides two overall scores: the TerraNova CTBS Basic Multiple Assessment (Math Total) and TerraNova CTBS Basic Multiple Assessment Plus (Math Computation) Total. Normal curve equivalent scores were used in the analysis. For a more detailed description of these outcome measures, see Appendix B.

Support for implementation

Teachers in the intervention group met with a *Scott Foresman–Addison Wesley Elementary Mathematics* professional trainer for approximately 4 hours prior to implementing the curriculum in their classes. In the initial training session, the trainer described the key components of the curriculum, reviewed the materials provided, offered examples of when to use certain materials, and provided an overview of the math technology available. Two follow-up sessions, approximately 2 hours each, were offered. The first follow-up session occurred 4–8 weeks after teachers began implementation. The second follow-up session was provided to five of the six participating schools and occurred 10–18 weeks after implementation.

Appendix B: Outcome measures for the mathematics achievment domain

Mathematics achievement	
Early Childhood Longitudinal Study–Kindergarten (ECLS-K) Math Assessment	This assessment was developed for the ECLS-K class of 1998–99. The ECLS-K is a nationally normed adaptive test. The assessment measures understanding and skills in five content areas: (a) number sense, properties, and operations; (b) measurement; (c) geometry and spatial sense; (d) data analysis, statistics, and probability; and (e) patterns, algebra, and functions. On the first-grade test, approximately three-quarters of the items focused on number sense, properties, and operations, with the remaining items predominantly focused on statistics, algebra, and functions. An ECLS-K math assessment for the second grade did not exist, so the study authors worked with the developer of the ECLS-K, Educational Testing Service, to select appropriate items from existing ECLS-K math assessments (including the K–1, third-, and fifth-grade instruments). Half of the items on the second-grade test were related to number sense, properties, and operations, with the other half covering measurement; geometry and spatial sense; and patterns, algebra, and functions (as cited in Agodini et al., 2010).
TerraNova Comprehensive Tests of Basics Skills (CTBS) Basic Multiple Assessment (Math Total)	The TerraNova CTBS Basic Multiple Assessment is a standardized test that provides an overall score for mathematics (the Math Total score). Level 12 was administered to second-grade (34 questions), Level 13 to third-grade (38 questions), Level 14 to fourth-grade (43 questions), and Level 15 to fifth-grade (43 questions) students. The test is administered during two class sessions and takes 75–90 minutes to complete. The majority of items are word problems measuring basic, applied, and higher-order thinking skills, and the test also contains a few computational problems, as well as multiple choice and constructed response questions. The authors state that they selected the test because of its validity, reliability, and sensitivity; because it assesses content presented in the latest textbook series available from multiple publishers; and because it reflects NCTM standards. The test is scored by CTB/McGraw-Hill, which provides a normal curve equivalent (NCE) score and scale score. Scorers demonstrated inter-rater reliability on the constructed response items of 0.86 to 0.98 in Resendez and Manley (2005) and 0.81 to 0.90 in Resendez and Azin (2006).
TerraNova CTBS Basic Multiple Assessment Plus (Math Computation)	The TerraNova CTBS Basic Multiple Assessment Plus test is a supplemental test that can be administered with the TerraNova CTBS Basic Multiple Assessment. It provides a separate overall score (the Math Computation score). The test contains 20 multiple-choice items measuring basic and advanced computational skills. The test takes 20 minutes to complete. It is scored by CTB/McGraw-Hill, which provides an NCE score and scale score.

Appendix C: Findings included in the rating for the mathematics achievement domain

			Mean (standard deviation)		WWC calculations			
Outcome measure	Study sample	Sample size	Intervention group	Comparison group	Mean difference	Effect size	Improvement index	*p*-value
Agodini et al., 2010[a]								
ECLS-K	Grade 1 (vs. *Investigations in Number, Data, and Space®*)	57 schools/ 2,396 students	44.54 (8.15)	44.51 (8.04)	0.03	0	0	0.93
ECLS-K	Grade 1 (vs. *Math Expressions*)	55 schools/ 2,481 students	43.85 (8.15)	44.74 (8.52)	–0.89	–0.11	–4	0.02
ECLS-K	Grade 1 (vs. *Saxon Math*)	55 schools/ 2,377 students	44.72 (8.15)	45.23 (7.32)	–0.51	–0.07	–3	0.16
ECLS-K	Grade 2 (vs. *Investigations in Number, Data, and Space®*)	36 schools/ 1,623 students	68.50 (15.74)	69.85 (15.75)	–1.35	–0.09	–3	0.09
ECLS-K	Grade 2 (vs. *Math Expressions*)	35 schools/ 1,633 students	69.49 (15.74)	71.38 (16.70)	–1.89	–0.12	–5	0.02
ECLS-K	Grade 2 (vs. *Saxon Math*)	36 schools/ 1,706 students	69.78 (15.74)	72.53 (16.16)	–2.75	–0.17	–7	0.00
Domain average for mathematics achievement (Agodini et al., 2010)						**–0.09**	**–4**	**Statistically significant**
Resendez & Azin, 2006[b]								
TerraNova Comprehensive Tests of Basics Skills (CTBS) Basic Multiple Assessment (Math Total)	Grades 3 and 5	39 classrooms/ 863 students	654.71 (42.40)	656.00 (47.81)	–1.29	–0.03	–1	nr
TerraNova CTBS Basic Multiple Assessment Plus (Math Computation)	Grades 3 and 5	39 classrooms/ 838 students	633.28 (52.03)	624.83 (52.58)	8.45	0.16	+6	nr
Domain average for mathematics achievement (Resendez & Azin, 2006)						**0.07**	**+3**	**Not statistically significant**
Resendez & Manley, 2005[c]								
TerraNova CTBS Basic Multiple Assessment (Math Total)	Grades 2 and 4	35 classrooms/ 624 students	55.59 (18.49)	54.14 (19.78)	1.45	0.08	+3	0.62
TerraNova CTBS Basic Multiple Assessment Plus (Math Computation)	Grades 2 and 4	35 classrooms/ 491 students	53.89 (21.35)	57.49 (20.46)	–3.60	–0.17	–7	0.19
Domain average for mathematics achievement (Resendez & Manley, 2005)						**–0.05**	**–2**	**Not statistically significant**
Domain average for mathematics achievement across all studies						**–0.02**	**–1**	**na**

Table Notes: For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the average change expected for all students who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average student's percentile rank that can be expected if the student is given the intervention. The WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of each study's domain average was determined by the WWC. nr = not reported. na = not applicable. ECLS-K = Early Childhood Longitudinal Study–Kindergarten.

[a] For Agodini et al. (2010), the unit of assignment is the school. The *p*-values presented here were reported in the original study. The intervention group mean is the unadjusted comparison mean plus the program coefficients from the hierarchical linear modeling (HLM) analysis. The comparison group mean is the unadjusted comparison group mean. A correction for multiple comparisons was needed such that the impact of *Scott Foresman–Addison Wesley Elementary Mathematics* compared to *Math Expressions* is no longer statistically significant. This study is characterized as having statistically significant negative effects because the effect for at least one measure within the domain is negative and statistically significant, and no effects are positive and statistically significant, accounting for clustering and multiple comparisons. For more information, please refer to the WWC Standards and Procedures Handbook, version 2.1, p. 96.

[b] For Resendez & Azin (2006), the unit of assignment is the teacher. The number of students refers to the number of students with posttests. The exact number of students taking both the pretest and posttest is not available. The outcome means are classroom/teacher-level means provided to the WWC by the study authors. The comparison group mean is the unadjusted comparison group classroom-level posttest mean. The intervention group mean is the comparison group classroom-level mean plus the difference in mean classroom-level gains between the intervention and comparison groups. The reported standard deviation is the student-level unadjusted posttest standard deviation obtained from the study's technical supplement. The effect size reported here differs from the effect size reported in the study. The effect size was calculated by the WWC using classroom-level means and student-level standard deviations. The *p*-values are not provided in the study for the specific contrasts of interest to the WWC. This study is characterized as having indeterminate effects because, based on WWC calculations, no effects are statistically significant or substantively important. For more information, please refer to the WWC Standards and Procedures Handbook, version 2.1, p. 96.

[c] For Resendez & Manley (2005), the unit of assignment is the teacher. The number of students indicates the number with posttests. The comparison group mean is the unadjusted comparison group mean reported in the study's technical supplement. The intervention group mean is the unadjusted comparison group mean plus the program coefficients from the HLM analysis as reported in the study's technical supplement. A correction for multiple comparisons was needed but did not affect significance levels. This study is characterized as having indeterminate effects because no effects are statistically significant or substantively important. For more information, please refer to the WWC Standards and Procedures Handbook, version 2.1, p. 96.

Endnotes

[1] The descriptive information for this program was obtained from a publicly available source: the program's website (http://www.pearsonschool.com, downloaded June 2010). The WWC requests developers review the program description sections for accuracy from their perspective. The program description was provided to the developer in January 2012, and we incorporated feedback from the developer. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review. The literature search reflects documents publicly available by December 2012.

[2] The previous report was released in July 2010. This report has been updated to include the review of one study released since that report. This study was within the scope of the protocol and meets evidence standards. A complete list and disposition of all studies reviewed are provided in the references. The studies in this report were reviewed using the Evidence Standards from the WWC Procedures and Standards Handbook (version 2.1), along with those described in the Elementary School Mathematics review protocol (version 2.0). When intervention reports are updated, all studies are re-reviewed under the current WWC standards. One study that met standards with reservations (Resendez & Manley, 2005) in the July 2010 report was re-reviewed for this report and, based on additional information provided to the WWC by the study authors, one portion of the analysis meets evidence standards without reservations, while the other analyses meet evidence standards with reservations. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.

[3] Absence of conflict of interest: One of the studies summarized in this intervention report, Agodini et al. (2010), was prepared by staff of one of the WWC contractors. Because the principal investigator for the WWC review of Elementary School Mathematics is also a staff member of that contractor and
a lead author of this study, the study was rated by staff members from a different organization. The report was then reviewed by the principal investigator, a WWC Quality Assurance reviewer, and an external peer reviewer.

[4] For criteria used in the determination of the rating of effectiveness and extent of evidence, see the WWC Rating Criteria on page 17. These improvement index numbers show the average and range of student-level improvement indices for all findings across the studies.

[5] Grade, delivery method, and program type refer to the studies that meet WWC evidence standards without or with reservations.

[6] The study compared intervention and comparison group outcomes on the TerraNova CTBS Basic Multiple Assessment Plus (Math Computation). The study also examined outcomes on the TerraNova CTBS Basic Multiple Assessment (Math Total), but this analysis has a lower evidence rating. The analysis of the TerraNova CTBS Basic Multiple Assessment (Math Total) outcome suffers from high attrition but shows equivalence between the intervention and comparison groups. Therefore, this portion of the study meets WWC evidence standards with reservations.

[7] Number of students indicates the number posttested.

[8] The study presented results based on student-level analysis. However, the analysis included some students who did not take both the pre- and posttests. To make results comparable with other studies in this review, an author query was conducted to obtain results based on classroom-level means. The results in this review are based on the class means.

[9] The exact number of students taking both the pretest and posttest is not available.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013, May). *Elementary School Mathematics intervention report: Scott Foresman–Addison Wesley Elementary Mathematics*. Retrieved from http://whatworks.ed.gov

WWC Rating Criteria

Criteria used to determine the rating of a study

Study rating	Criteria
Meets WWC evidence standards without reservations	A study that provides strong evidence for an intervention's effectiveness, such as a well-implemented RCT.
Meets WWC evidence standards with reservations	A study that provides weaker evidence for an intervention's effectiveness, such as a QED or an RCT with high attrition that has established equivalence of the analytic samples.

Criteria used to determine the rating of effectiveness for an intervention

Rating of effectiveness	Criteria
Positive effects	Two or more studies show statistically significant positive effects, at least one of which met WWC evidence standards for a strong design, AND No studies show statistically significant or substantively important negative effects.
Potentially positive effects	At least one study shows a statistically significant or substantively important positive effect, AND No studies show a statistically significant or substantively important negative effect AND fewer or the same number of studies show indeterminate effects than show statistically significant or substantively important positive effects.
Mixed effects	At least one study shows a statistically significant or substantively important positive effect AND at least one study shows a statistically significant or substantively important negative effect, but no more such studies than the number showing a statistically significant or substantively important positive effect, OR At least one study shows a statistically significant or substantively important effect AND more studies show an indeterminate effect than show a statistically significant or substantively important effect.
Potentially negative effects	One study shows a statistically significant or substantively important negative effect and no studies show a statistically significant or substantively important positive effect, OR Two or more studies show statistically significant or substantively important negative effects, at least one study shows a statistically significant or substantively important positive effect, and more studies show statistically significant or substantively important negative effects than show statistically significant or substantively important positive effects.
Negative effects	Two or more studies show statistically significant negative effects, at least one of which met WWC evidence standards for a strong design, AND No studies show statistically significant or substantively important positive effects.
No discernible effects	None of the studies shows a statistically significant or substantively important effect, either positive or negative.

Criteria used to determine the extent of evidence for an intervention

Extent of evidence	Criteria
Medium to large	The domain includes more than one study, AND The domain includes more than one school, AND The domain findings are based on a total sample size of at least 350 students, OR, assuming 25 students in a class, a total of at least 14 classrooms across studies.
Small	The domain includes only one study, OR The domain includes only one school, OR The domain findings are based on a total sample size of fewer than 350 students, AND, assuming 25 students in a class, a total of fewer than 14 classrooms across studies.

Glossary of Terms

Attrition Attrition occurs when an outcome variable is not available for all participants initially assigned to the intervention and comparison groups. The WWC considers the total attrition rate and the difference in attrition rates across groups within a study.

Clustering adjustment If intervention assignment is made at a cluster level and the analysis is conducted at the student level, the WWC will adjust the statistical significance to account for this mismatch, if necessary.

Confounding factor A confounding factor is a component of a study that is completely aligned with one of the study conditions, making it impossible to separate how much of the observed effect was due to the intervention and how much was due to the factor.

Design The design of a study is the method by which intervention and comparison groups were assigned.

Domain A domain is a group of closely related outcomes.

Effect size The effect size is a measure of the magnitude of an effect. The WWC uses a standardized measure to facilitate comparisons across studies and outcomes.

Eligibility A study is eligible for review and inclusion in this report if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design.

Equivalence A demonstration that the analysis sample groups are similar on observed characteristics defined in the review area protocol.

Extent of evidence An indication of how much evidence supports the findings. The criteria for the extent of evidence levels are given in the WWC Rating Criteria on p.17.

Improvement index Along a percentile distribution of students, the improvement index represents the gain or loss of the average student due to the intervention. As the average student starts at the 50th percentile, the measure ranges from –50 to +50.

Multiple comparison adjustment When a study includes multiple outcomes or comparison groups, the WWC will adjust the statistical significance to account for the multiple comparisons, if necessary.

Quasi-experimental design (QED) A quasi-experimental design (QED) is a research design in which subjects are assigned to intervention and comparison groups through a process that is not random.

Randomized controlled trial (RCT) A randomized controlled trial (RCT) is an experiment in which investigators randomly assign eligible participants into intervention and comparison groups.

Rating of effectiveness The WWC rates the effects of an intervention in each domain based on the quality of the research design and the magnitude, statistical significance, and consistency in findings. The criteria for the ratings of effectiveness are given in the WWC Rating Criteria on p. 17.

Single-case design A research approach in which an outcome variable is measured repeatedly within and across different conditions that are defined by the presence or absence of an intervention.

Standard deviation The standard deviation of a measure shows how much variation exists across observations in the sample. A low standard deviation indicates that the observations in the sample tend to be very close to the mean; a high standard deviation indicates that the observations in the sample tend to be spread out over a large range of values.

Statistical significance Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ($p < 0.05$).

Substantively important A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.

Please see the WWC Procedures and Standards Handbook (version 2.1) for additional details.