What Works Clearinghouse

INSTITUTE OF EDUCATION SCIENCES

Elementary School Math Revised April 30, 2007

Saxon Elementary School Math

Program description

Saxon Elementary School Math, published by Harcourt Achieve, is a core curriculum for students in kindergarten through grade 5. A distinguishing feature of *Saxon Elementary School Math* is its use of a distributed approach, as opposed to a chapter-based approach, for instruction and assessment. The program is built on the premise that students learn best when instruction is incremental and explicit, previously learned concepts are continually reviewed, and assessment is frequent and cumulative. At each grade level, math concepts are introduced, reviewed, and practiced over time in order to move students from understanding to mastery to fluency. For grades K–3, the *Saxon Elementary School Math* curriculum emphasizes hands-on activities and teacher-directed math conversations that engage students in learning. The curriculum for grades 4–5 also uses math conversations to introduce new concepts, but shifts the focus to student-directed learning.

Research

One study of the *Saxon Elementary School Math* program met the What Works Clearinghouse (WWC) evidence standards with reservations. The study included students in grades 1–8 from a range of socioeconomic backgrounds and attending 342 schools across the state of Georgia. This report focuses only on findings for grades 1–5.[1]

The WWC considers the extent of evidence for *Saxon Elementary School Math* to be small for math achievement.

Effectiveness

Saxon Elementary School Math was found to have no discernible effects on math achievement.

	Math achievement
Rating of effectiveness	No discernible effects
Improvement index[2]	na

na = not applicable

1. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.
2. The student-level improvement index could not be computed for *Saxon Elementary School Math*.

Additional program information

Developer and contact

Saxon Elementary School Math was developed by Saxon Publishers, an imprint of Harcourt Achieve. Address: 6277 Sea Harbor Drive, Orlando, Florida 32887. Email: info@SaxonPublishers.com. Web: http://saxonpublishers.harcourtachieve.com. Telephone: (800) 284-7019.

Scope of use

The first Saxon textbook, *Saxon Algebra,* was published in 1979 by John Saxon for junior college students. In 1980, a high school version, *Algebra 1,* was published. In 1981, the program was tested by 20 teachers with approximately 1,400 students. By 1993, the company had become Saxon Publishers and developed programs for kindergarten through high school. Information is not available on the number or demographics of students, schools, or districts using this intervention.

Teaching

Daily lessons in grades 1–3 consist of three components: 1) the meeting, 2) the math lesson, and 3) the written practice, which includes guided class practice and homework. A typical lesson begins with the meeting where students engage in various practical activities (for example, understanding calendars), and enter into math conversations and dialogue with the class and teacher to communicate their understanding of math concepts. Following the meeting, the teacher begins the math lesson in which new concepts are introduced. Hands-on activities are incorporated into the math lesson to encourage student involvement and further the learning of new concepts. The math lesson is followed by written practice, which includes teacher-facilitated guided class practice of newly and previously learned concepts. The day's homework is completed by the students independently. Cumulative and written assessments occur every five lessons. In kindergarten, these components may be separated into different sessions, and assessments are conducted as individual interviews between the teacher and individual students.

For grades 4 and 5, a daily lesson consists of four components: 1) warm-up, 2) lesson introducing new math concept, 3) practice on new concept, and 4) mixed practice including new and previously learned concepts. Students are introduced to concepts incrementally, given opportunities for continual review and practice, and assessed cumulatively and frequently. An assessment score of 80% or lower indicates a need for remediation, and provision for remediation is part of the program.

Cost

Saxon Elementary School Math grades 1–3 can be ordered as a 24-student or 32-student kit that includes all of the teacher, lesson, classroom and student materials. The student kits range from over $600 to over $800 depending on the size of the kit. Individual kit components, such as manipulatives, workbooks, student texts, teacher manuals, and materials in Spanish, can be purchased separately. Grades 4 and 5 have a separate student edition ($50–$55) and teacher manual set ($185). Other ancillary materials, such as black line master books, fact practice workbooks, and a test practice generator, can be purchased separately.

Research

Seven studies reviewed by the WWC investigated the effects of *Saxon Elementary School Math.* One study (Resendez & Manley, 2005) was a quasi-experimental design that met WWC evidence standards with reservations. The remaining six studies did not meet WWC evidence screens.

Resendez & Manley (2005) conducted a retrospective study that included 170 intervention schools in Georgia and 172 comparison schools that were matched to the intervention schools on student demographics, geographical location, and baseline math performance on Georgia's Criterion-Referenced Competency Test (CRCT). The intervention schools used the *Saxon Elementary School Math* program recommended for each grade level in grades 1–8 between 2000 (or prior) and 2005. The comparison schools used a variety of non-*Saxon* curricula. The majority of comparison schools used traditional basal math curricula. One third of the schools used a mix of basal, investigative,

Research *(continued)*

and other approaches, and a small percentage used an investigative approach to teaching math. This intervention report presents findings for grades 1–5 on Georgia's Criterion-Referenced Competency Test (CRCT).

Extent of evidence

The WWC categorizes the extent of evidence in each domain as small or moderate to large (see the What Works Clearinghouse Extent of Evidence Categorization Scheme). The extent of evidence takes into account the number of studies and the total sample size across the studies that met WWC evidence standards with or without reservations.[3]

The WWC considers the extent of evidence for *Saxon Elementary School Math* to be small for math achievement.

Effectiveness

Findings

The WWC review of elementary school mathematics curriculum-based interventions addresses student outcomes in overall math achievement.

Resendez & Manley (2005) reported no significant effects of the *Saxon Elementary School Math* program on overall math achievement at grades 1–5. Using school-level data provided by the authors, the WWC confirmed that *Saxon Elementary School Math* did not have a statistically significant effect on math achievement at each grade level from first to fifth grade. Based on this study finding, the WWC categorized *Saxon Elementary School Math* as having no discernible effects on overall math achievement.

The WWC also calculated effect sizes and significance levels for subtests of the Georgia CRCT using school-level data provided by the authors. The WWC found significant effects of *Saxon Elementary School Math* on patterns, relations, and algebra at grade 2. For grade 4 the WWC found statistically significant effects of *Saxon Elementary School Math* on computations and estimation, and problem solving. Subtest results were not included in the WWC's rating of effectiveness.

Rating of effectiveness

The WWC rates the effects of an intervention in a given outcome domain as: positive, potentially positive, mixed, no discernible effects, potentially negative, or negative. The rating of effectiveness takes into account four factors: the quality of the research design, the statistical significance of the findings (as calculated by the WWC[4]), the size of the difference between participants in the intervention and the comparison conditions, and the consistency in findings across studies (see the WWC Intervention Rating Scheme).

The WWC found *Saxon Elementary School Math* to have no discernible effects on math achievement

Improvement index

The WWC computes an improvement index for each individual finding. In addition, within each outcome domain, the WWC computes an average improvement index for each study as well as an average improvement index across studies (see the Technical Details of WWC-Conducted Computations). The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. Unlike the rating of effectiveness, the improvement index is based entirely on the size of the effect, regardless of the statistical significance of the effect, the study design, or the

3. The Extent of Evidence categorization was developed to tell readers how much evidence was used to determine the intervention rating, focusing on the number and size of studies. Additional factors associated with a related concept, external validity, such as students' demographics and the types of settings in which studies took place, are not taken into account for the categorization.
4. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation, see the WWC Tutorial on Mismatch. See the Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate the statistical significance. In the case of Saxon Elementary School Math, no corrections for clustering or multiple comparisons were needed.

The WWC found *Saxon Elementary School Math* to have no discernible effects on math achievement *(continued)*

analysis. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results. The average and range of the student-level improvement index could not be computed because student-level standard deviations were not available for the single study reviewed.

Summary

The WWC reviewed seven studies on *Saxon Elementary School Math.* One of these studies met WWC standards with reservations and the remaining studies did not meet WWC evidence screens. This study found no discernible effects on math achievement. The evidence presented in this report may change as new research emerges.

References

Met WWC standards with reservations

Resendez, M., & Manley, M. A. (2005). *The relationship between using Saxon Elementary and Middle School Math and student performance on Georgia statewide assessments.* Orlando, FL: Harcourt Achieve.

Did not meet WWC evidence screens

Calvery, R., Bell, D., & Wheeler, G. (1993, November). *A comparison of selected second and third graders' math achievement: Saxon vs. Holt.* Paper presented at the meeting of the Mid-South Educational Research Association, New Orleans, LA.[5]

Fahsl, A. J. (2001). An investigation of the effects of exposure to Saxon math textbooks, socioeconomic status and gender on math achievement scores. *Dissertation Abstracts International 62*(08), 2671A. (UMI No. 3021615)[6]

Hansen, E., & Greene, K. (2000). *A recipe for math: What's cooking in the classroom: Saxon or Traditional?* Retrieved May 4, 2006 from http://www.secondaryenglish.com/recipeformath.html[5]

Klein, D. (2000). *High achievement in mathematics: Lessons from three Los Angeles elementary schools.* Washington, DC: Brookings Institution Press.[7]

Nguyen, K., Elam, P., & Weeter, R. (1993). *The 1992–93 Saxon mathematics program evaluation report.* Oklahoma City: Oklahoma City Public Schools.[8]

Plato, J. (1998). *An evaluation of Saxon math at Blessed Sacrament School.* Retrieved May 4, 2006 from the University of Illinois, College of Education Web site: http://lrs.ed.uiuc.edu/students/plato1/Final.html[7]

For more information about specific studies and WWC calculations, please see the WWC *Saxon Elementary School Math* Technical Appendices.

5. Does not use a strong causal design: the study, which used a quasi-experimental design, did not establish that the comparison group was equivalent to the treatment group at the baseline in a pretest measure of math achievement.
6. Does not use a strong causal design: the study did not use a comparison group.
7. Does not use a strong causal design: this is a qualitative study.
8. Disruption: the study, which used a quasi-experimental design, demonstrated problems with disruption or contamination.

Appendix

Appendix A1 Study characteristics: Resendez & Manley, 2005 (quasi-experimental design)

Characteristic	Description
Study citation	Resendez, M., & Manley, M. A. (2005). *The relationship between using Saxon Elementary and Middle School Math and student performance on Georgia Statewide Assessments.* Orlando, Fla.: Harcourt Achieve.
Participants	The participants in this study were students in grades 1–8 in 170 intervention schools and 172 comparison schools that were matched on student demographics, geographical location, and baseline math performance on Georgia's CRCT. This intervention report focuses only on findings for grades 1–5, because grades 6–8 are outside of the scope of this review.[1] The authors selected Georgia schools that used the *Saxon Elementary School Math* curriculum between 2000 and 2005. The sample was obtained from the Georgia Department of Education (GDE). The authors note that per state policy, only school-level data could be released. Data for the intervention group came from 85 schools for first grade, 85 schools for second grade, 83 schools for third grade, 79 schools for fourth grade and 79 schools for fifth grade. Data for the comparison group came from 144 schools for first grade, 144 schools for second grade, 135 schools for third grade, 131 schools for fourth grade, and 129 schools for fifth grade. The number of schools per grade is not mutually exclusive. Some of the schools contained multiple grades so the numbers presented do not represent distinct clusters of schools.
Setting	The sample schools were distributed across the state of Georgia and represented a mixture of rural, urban, and suburban communities. The gender and racial compositions of the schools were similar in the intervention schools and comparison schools, with roughly equal gender distribution and more than half of the students white. More than half of the students were educationally disadvantaged, 16% had disabilities, 3% had limited English proficiency, and 6% were gifted in both study conditions.
Intervention	The *Saxon Elementary School Math* curriculum was used as a core curriculum in the intervention schools. The elementary schools in the sample used the version of the *Saxon Elementary School Math* program that was appropriate for each grade level, and participating schools had used the program for an average of three years (range 1–15 years).
Comparison	The schools in the comparison group used a mixture of non-*Saxon* curricula. More than half of the schools in the comparison group used basal math curricula with chapter-based approaches to teaching math. Five percent of the schools used curricula with an investigative approach. The remaining third of the schools used curricula that were a mix of basal, investigative, computer-based approaches. The authors report no significant differences in baseline math performance between the *Saxon* and non-*Saxon* schools.
Primary outcomes and measurement	The outcome measure was the Georgia's Criterion-Referenced Competency Test (CRCT) which assesses competency in number sense and numeration, geometry and measurement, patterns and relations/algebra, statistics and probability, computation and estimation, and problem solving. Fourth-grade students were tested in each school year from 1999–2000 to 2004–05. First-grade, second-grade, third-grade, and fifth-grade students were tested in the spring of school years 2001–02, 2003–04, and 2004–05. All posttest scores are from spring 2005 (see Appendix A2 for more detailed descriptions for outcome measures).
Teacher training	No information was provided regarding the teacher training for the intervention.

1. Results from grades 6–8 are being reviewed as part of the WWC Middle School Mathematics review.

Appendix A2 Outcome measures in the math achievement domain

Outcome measure	Description
Georgia's Criterion-Referenced Competency Test (CRCT)[1]	As cited in Resendez & Manley (2005), the CRCT is a criterion-referenced test which is referenced to Georgia's Quality Core Curriculum Goals. According to the Georgia Department of Education, the CRCT is a multiple-choice test that is valid and reliable for Georgia's public school students.[2] The CRCT math scores range from 150–450, with scores below 300 not meeting standards and scores above 350 exceeding standards. The criteria for meeting the standards vary by objective and grade level. Five objectives are covered by the test: 1) number and number sense, 2) geometry and measurement, 3) patterns, relationships, and algebra, 4) computation and estimation, and 5) problem solving. The cut points are set by the state and take into account the difficulty of the specific objective.

1. The original CRCT scores shown in the report are by objective. Upon request from the WWC, the author calculated the mean overall score across all objectives, controlling for pretest, for each grade.
2. Georgia Department of Education, (n.d.). *Criterion-Referenced Competency Tests.* Retrieved on September 13, 2006 from http://doe.k12.ga.us/curriculum/testing/crct.asp.

Appendix A3 Summary of study findings included in the rating for math achievement[1]

Outcome measure	Study sample	Sample size (schools)	Authors' findings from the study: Mean outcome (standard deviation[2]) *Saxon* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) Comparison group[3]	WWC calculations: Mean difference[4] (*Saxon* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Resendez & Manley 2005 (quasi-experimental design)[8]								
GCRCT: Overall % students meeting objectives	Grade 1	229	86.26 (nr)	85.2 (nr)	1.06	na[10]	ns	na[10]
Average[9] for math achievement (Resendez & Manley, 2005): Grade 1						na[10]	ns	na[10]
GCRCT: Overall % students meeting objectives	Grade 2	229	88.31 (nr)	86.86 (nr)	1.45	na[10]	ns	na[10]
Average[9] for math achievement (Resendez & Manley, 2005): Grade 2						na[10]	ns	na[10]
GCRCT: Overall % students meeting objectives	Grade 3	218	86.94 (nr)	85.93 (nr)	1.01	na[10]	ns	na[10]
Average[9] for math achievement (Resendez & Manley, 2005): Grade 3						na[10]	ns	na[10]
GCRCT: Overall % students meeting objectives	Grade 4	210	73.92 (nr)	71.39 (nr)	2.53	na[10]	ns	na[10]
Average[9] for math achievement (Resendez & Manley, 2005): Grade 4						na[10]	ns	na[10]
GCRCT: Overall % students meeting objectives	Grade 5	208	82.46 (nr)	81.66 (nr)	0.80	na[10]	ns	na[10]
Average[9] for math achievement (Resendez & Manley, 2005): Grade 5						na[10]	ns	na[10]
Domain average[9] for math achievement across all grades						na[10]	na	na[10]

ns = not statistically significant
na = not applicable
nr = not reported

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. Subtest findings from the same studies are not included in these ratings, but are reported in Appendix A4.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are; a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. The intervention group and control group means are pretest-adjusted means provided by the authors and differ from what is in the original study.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. The effect sizes were computed based on school-level data, which were likely to be larger than effect sizes based on student-level data. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Resendez & Manley (2005), no corrections for clustering or multiple comparisons were needed.
9. This row provides the study average, which, in this instance, is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.
10. Student-level standard deviations were not available for this study. School-level standard deviations for the intervention group were 6.60 for grade 1, 6.39 for grade 2, 6.50 for grade 3, 8.51 for grade 4, and 6.94 for grade 5. School-level standard deviations for the comparison group were 6.80 for grade 1, 7.35 for grade 2, 7.15 for grade 3, 11.83 for grade 4, and 8.93 for grade 5. Because student-level effect sizes and improvement indices could not be computed, the magnitude of the effect size was not considered for rating purposes. However, the statistical significance for this study is comparable to other studies and is included in the intervention rating. For further details, please see Technical Details of WWC-Conducted Computations.

Appendix A4 Summary of subtest findings for math achievement[1]

Outcome measure	Study sample	Sample size (schools)	Authors' findings from the study: Mean outcome (standard deviation[2]) *Saxon* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) Comparison group[3]	WWC calculations: Mean difference[4] (*Saxon* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Resendez and Manley 2005 (quasi-experimental design)[8]								
GCRCT: Number and number sense	Grade 1	229	89.53 (nr)	88.52 (nr)	1.01	na[9]	ns	na[9]
GCRCT: Geometry and measurement	Grade 1	229	90.34 (nr)	90.29 (nr)	0.05	na[9]	ns	na[9]
GCRCT: Patterns, relations, and algebra	Grade 1	229	87.88 (nr)	86.28 (nr)	1.60	na[9]	ns	na[9]
GCRCT: Computation and estimation	Grade 1	229	78.93 (nr)	77.43 (nr)	1.50	na[9]	ns	na[9]
GCRCT: Problem solving	Grade 1	229	84.64 (nr)	83.49 (nr)	1.15	na[9]	ns	na[9]
GCRCT: Number and number sense	Grade 2	229	88.57 (nr)	86.62 (nr)	1.95	na[9]	ns	na[9]
GCRCT: Geometry and measurement	Grade 2	229	91.46 (nr)	92.36 (nr)	–0.90	na[9]	ns	na[9]
GCRCT: Patterns, relations, and algebra	Grade 2	229	87.05 (nr)	83.58 (nr)	3.47	na[9]	Statistically significant	na[9]
GCRCT: Computation and estimation	Grade 2	229	86.93 (nr)	85.83 (nr)	1.10	na[9]	ns	na[9]
GCRCT: Problem solving	Grade 2	229	87.54 (nr)	85.93 (nr)	1.61	na[9]	ns	na[9]
GCRCT: Number and number sense	Grade 3	218	89.74 (nr)	88.24 (nr)	1.50	na[9]	ns	na[9]
GCRCT: Geometry and measurement	Grade 3	218	93.6 (nr)	92.24 (nr)	1.36	na[9]	ns	na[9]
GCRCT: Patterns, relations, and algebra	Grade 3	218	86.26 (nr)	85.9 (nr)	0.36	na[9]	ns	na[9]

(continued)

Appendix A4 Summary of subtest findings for math achievement[1] *(continued)*

Outcome measure	Study sample	Sample size (schools)	Authors' findings from the study: Mean outcome (standard deviation[2]) *Saxon* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) Comparison group[3]	WWC calculations: Mean difference[4] (*Saxon* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
GCRCT: Statistics and probability	Grade 3	218	87.13 (nr)	85.83 (nr)	1.30	na[9]	ns	na[9]
GCRCT: Computation and estimation	Grade 3	218	86.81 (nr)	85.71 (nr)	1.10	na[9]	ns	na[9]
GCRCT: Problem solving	Grade 3	218	78.11 (nr)	77.64 (nr)	0.47	na[9]	ns	na[9]
GCRCT: Number and number sense	Grade 4	210	71.47 (nr)	70.85 (nr)	0.62	na[9]	ns	na[9]
GCRCT: Geometry and measurement	Grade 4	210	79.22 (nr)	78.16 (nr)	1.06	na[9]	ns	na[9]
GCRCT: Patterns, relations, and algebra	Grade 4	210	69.76 (nr)	67.7 (nr)	2.06	na[9]	ns	na[9]
GCRCT: Statistics and probability	Grade 4	210	82.15 (nr)	80.17 (nr)	1.98	na[9]	ns	na[9]
GCRCT: Computation and estimation	Grade 4	210	73.12 (nr)	67.65 (nr)	5.47	na[9]	Statistically significant	na[9]
GCRCT: Problem solving	Grade 4	210	67.81 (nr)	63.83 (nr)	3.98	na[9]	Statistically significant	na[9]
GCRCT: Number and number sense	Grade 5	208	79.74 (nr)	77.31 (nr)	2.43	na[9]	ns	na[9]
GCRCT: Geometry and measurement	Grade 5	208	80.77 (nr)	81.54 (nr)	–0.77	na[9]	ns	na[9]
GCRCT: Patterns, relations, and algebra	Grade 5	208	76.16 (nr)	74.56 (nr)	1.60	na[9]	ns	na[9]
GCRCT: Statistics and probability	Grade 5	208	79.82 (nr)	81.52 (nr)	–1.70	na[9]	ns	na[9]
GCRCT: Computation and estimation	Grade 5	208	88.74 (nr)	86.62 (nr)	2.12	na[9]	ns	na[9]

(continued)

Appendix A4 Summary of subtest findings for math achievement[1] *(continued)*

			Authors' findings from the study					
			Mean outcome (standard deviation[2])		WWC calculations			
Outcome measure	Study sample	Sample size (schools)	*Saxon* group[3]	Comparison group[3]	Mean difference[4] (*Saxon* – comparison)	Effect size[5]	Statistical significance[6] (at $\alpha = 0.05$)	Improvement index[7]
GCRCT: Problem solving	Grade 5	208	89.55 (nr)	88.43 (nr)	1.12	na[9]	ns	na[9]

na = not applicable
nr = not reported
ns = not statistically significant

1. This appendix presents subscale findings for measures that fall in math achievement. Total scale scores were used for rating purposes and are presented in Appendix A3.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are; a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. The intervention group and control group means are pretest-adjusted means provided by the authors and differ from what is in the original study.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. The effect sizes were computed based on school-level data, which were likely to be larger than effect sizes based on student-level data. The effect sizes may differ from those presented by the authors because the authors calculated Cohen's *d* using the proportion of variance, while the WWC calculated a Hedges' *g* using the adjusted means, unadjusted standard deviations, and sample sizes for each group using additional data provided by the authors.
6. 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 used additional data provided by the authors to calculate significance.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Resendez & Manley (2005), no corrections for clustering or multiple comparisons were needed.
9. Student-level standard deviations and improvement indices were not available for this study. School-level standard deviations, which were requested by the WWC and provided by the first study author, ranged from 4.50 to 10.32 across grade levels and subtests in the intervention group and from 5.41 to 14.75 across grade levels and subtests in the comparison group. Because student-level standard deviations were not available, student-level effect sizes and improvement indices could not be computed. However, the statistical significance of the findings in Resendez & Manley (2005) is comparable to other studies and is reported in this appendix. For further details, please see Technical Details of WWC-Conducted Computations.

Appendix A5 *Saxon Elementary School Math* rating for the math achievement domain

The WWC rates the effects of an intervention in a given outcome domain as: positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of math achievement, the WWC rated the *Saxon Elementary School Math* program as having no discernible effects. It did not meet the criteria for other ratings (positive effects, potentially positive effects, mixed effects, potentially negative effects, or negative effects) because the single study that met WWC evidence standards with reservations did not show statistically significant or substantively important effects.

Rating received

No discernible effects: No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative.*

  Met. The one study of Saxon Elementary School Math that met WWC standards with reservations showed an indeterminate effect.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. No studies of *Saxon Elementary School Math* met WWC evidence standards for a strong design. Further, no studies showed a statistically significant positive effect.

AND

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

  Met. No studies showed a statistically significant or substantively important negative effect.

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

  Not met. No studies showed a statistically significant or substantively important positive effect.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

  Not met. No studies showed a statistically significant important effect, either negative or positive. The one study of *Saxon Elementary School Math* that met WWC evidence standards with reservations showed an indeterminate effect.

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing 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.

  Not met. No studies of *Saxon Elementary School Math* showed a statistically significant or substantively important effect, either positive or negative.

OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

  Not met. No studies of *Saxon Elementary School Math* showed a statistically significant or substantively important effect.

(continued)

Appendix A5 *Saxon Elementary School Math* rating for the math achievement domain *(continued)*

Potentially negative effects: Evidence of a negative effect with no overriding contrary evidence

- Criterion 1: At least one study showing a statistically significant or substantively important *negative* effect.

  Not met. No studies of *Saxon Elementary School Math* showed a statistically significant or substantively important negative effect.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

  Met. No studies of *Saxon Elementary School Math* showed a statistically significant or substantively important positive effect.

Negative effects: Strong evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *negative* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. No studies showed a statistically significant negative effect or met WWC evidence standards for a strong design.

AND

- Criterion 2: No studies showing statistically significant or substantively important *positive* effects.

  Met. No studies of *Saxon Elementary School Math* showed a statistically significant or substantively important positive effect.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain level effect. The WWC also considers the size of the domain level effect for ratings of potentially positive or potentially negative effects. See the WWC Intervention Rating Scheme for a complete description.

Appendix A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[1]
		Schools	Students	
Math achievement	1	299	nr	Small

nr = not reported

1. A rating of "moderate to large" requires at least two studies and two schools across studies in one domain, and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is "small."