What Works Clearinghouse™

INSTITUTE OF EDUCATION SCIENCES

Middle School Math Updated February 2013

Saxon Math

Program Description[1]

Saxon Math, published by Houghton Mifflin Harcourt, is a core curriculum for students in grades K–12. This report includes studies that investigate the potential impact of *Saxon Math* for students in grades 6–8. A distinguishing feature of the curriculum is its use of an incremental approach for instruction and assessment. This approach limits the amount of new math content delivered to students each day and allows time for daily practice. New concepts are introduced gradually and integrated with previously introduced content so that concepts are developed, reviewed, and practiced over time rather than being taught during discrete periods of time, such as in chapters or units.

Research[2]

The What Works Clearinghouse (WWC) identified five studies of *Saxon Math* that both fall within the scope of the Middle School Math topic area and meet WWC evidence standards. One study meets standards without reservations, and four studies meet standards with reservations, and together, they included 6,601 students in grades 6–8 from 52 schools in four states.

The WWC considers the extent of evidence for *Saxon Math* on the math performance of middle school students to be medium to large for the mathematics achievement domain, the only domain examined for studies reviewed under the Middle School Math topic area.

Report Contents

Effectiveness

Saxon Math was found to have mixed effects on mathematics achievement for middle school students.

Table 1. Summary of findings[3]

Outcome domain	Rating of effectiveness	Improvement index *(percentile points)*		Number of studies	Number of students	Extent of evidence
		Average	Range			
Mathematics achievement	Mixed effects	+9	+5 to +16	5	6,601	Medium to large

Program Information

Background

Originally developed by John Saxon, *Saxon Math* is distributed by Houghton Mifflin Harcourt Supplemental Publishers. Address: Specialized Curriculum Group, 9205 Southpark Center Loop, Orlando, FL, 32819. Email: greatservice@hmhpub.com. Website: www.saxonpublishing.com. Telephone: (800) 289-4490. Fax: (800) 289-3994.

Program details

The *Saxon Math* curriculum consists of 120 daily lessons and 12 activity-based investigations for each grade level. Each lesson makes use of three strategies:

- The first strategy involves offering three types of activities: (a) fact-fluency practice that promotes recall when working with math operations and fractions, (b) mental math exercises intended to build number sense and problem-solving strategies, and (c) practice solving challenging, non-routine story problems in which problem solving strategies are emphasized.
- The second strategy is to limit the amount of new math content delivered to students each day. This strategy introduces a relatively small set of new math ideas daily using examples, mathematical conversations, and practice, and integrates the new concepts with ones that were previously introduced.
- The third strategy involves written practice that aims to help students both master new skills and maintain their mastery of concepts previously instructed.

Students complete written, cumulative assessments after every five lessons. The results of these assessments provide teachers with data for instructional decision making and provide feedback for students and parents. In addition to these written assessments, students may demonstrate mastery of math content through alternate interactive opportunities, such as investigations, test-day activities, and performance tasks.

Currently *Saxon Math* for middle school offers three textbooks (*Saxon Math* Course 1 for grade 6, *Saxon Math* Course 2 for grade 7, and *Saxon Math* Course 3 for grade 8). Earlier versions of the curriculum offered different textbooks for middles school grades (*Saxon 7/6*, *Saxon 8/7*, *Saxon Algebra ½*, and *Saxon Algebra 1*). The study descriptions provided in this report indicate the version of *Saxon Math* that was evaluated by the study authors in cases in which that information is available.

Cost

For the 2012 publication of *Saxon Math*, the student edition for each course costs $70.15 per student for a hard copy, $56.15 for an eBook, $17.35 for a one year subscription to an online edition, or $53.70 for a six year subscription to an online edition. The teacher's manual costs $113.65 for a hard copy, $27.35 for a one year subscription to an online edition, or $85.25 for a six year subscription to an online edition. A teacher technology package is available for $225.75 that includes the Teacher's Manual eBook and various electronic teaching and planning resources. Other materials, such as student workbooks, instructional presentations, and manipulative kits, are available and range in price from $3.10 to $352.00.

Research Summary

The WWC identified 21 studies that investigated the effects of *Saxon Math* on the mathematics achievement of middle school students.

Table 2. Scope of reviewed research[4]

Grade	6, 7, 8
Delivery method	Whole class
Program type	Curriculum

The WWC reviewed 16 of those studies against group design evidence standards. One study (Resendez & Azin, 2006) is a randomized controlled trial that meets WWC evidence standards without reservations. Four studies (Crawford & Raia, 1986; Peters, 1992; Resendez, Fahmy, & Manley, 2005, Cohort A; and Resendez, Fahmy, & Manley, 2005, Cohort F) are randomized controlled trials or quasi-experimental designs that meet WWC evidence standards with reservations.[5] Those five studies are summarized in this report. Eleven studies do not meet WWC evidence standards. The remaining five studies do not meet WWC eligibility screens for review in this topic area. Citations for all 21 studies are in the References section, which begins on p. 6.

Summary of study meeting WWC evidence standards without reservations

Resendez and Azin (2006) conducted a randomized controlled trial to investigate the effect of *Saxon Math* on math achievement in one northeastern Ohio middle school and one southwestern Ohio junior high school.[6] The schools served sixth-, seventh-, and eighth-grade students living in urban and suburban locations. Classes were randomly assigned to use *Saxon Math* or comparison curricula during the 2005–06 school year. The analysis sample included 281 students in 14 *Saxon Math* classrooms and 219 students in 11 comparison group classrooms. Classes in the intervention group used one of four *Saxon Math* curricula: (a) *Saxon 7/6*, 2004–4th Ed., (b) *Saxon 8/7*, 2004–4th Ed., (c) *Saxon Algebra ½*, 2004–3rd Ed., or (d) *Saxon Algebra 1*, 2003–3rd Ed. Classes in the comparison group used either a traditional basal program or a mixed curriculum consisting primarily of teacher-created materials.

Summary of studies meeting WWC evidence standards with reservations

Crawford and Raia (1986) conducted a matched-comparison quasi-experiment to investigate the effect of *Saxon Math* on the math achievement of eighth-grade students in Oklahoma City Public Schools during the 1984–85 school year. The analysis sample included 78 eighth-grade students (39 *Saxon* and 39 comparison) taught by four teachers from four middle schools. Students in the intervention group used the *Saxon Algebra ½* (1983) textbook, and students in the comparison group used the *Scott Foresman Mathematics* (1980) textbook.

Peters (1992) conducted a randomized controlled trial in which the integrity of random assignment was compromised because some students did not remain in the study group to which they were randomly assigned—students were reallocated between the intervention and comparison groups to accommodate scheduling difficulties and student requests for other course offerings. The study investigated the effect of *Saxon Math* on the math achievement of 36 "math-talented" eighth-grade students (19 *Saxon Math* and 17 comparison) from one junior high school in Nebraska during the 1991–92 school year.[7] The district borders two large cities (Lincoln and Omaha) and its students lived in rural and suburban areas. Students in the intervention group used the *Saxon Algebra 1* (1981) textbook, while students in the comparison group used the *University of Chicago School Mathematics Project (UCSMP) Algebra* 1st-edition textbook.

Resendez, Fahmy, and Manley (2005, Cohort A) conducted a matched-comparison quasi-experiment to investigate the effect of *Saxon Math* on the math achievement of sixth-, seventh-, and eighth-grade students in 25 middle schools located in rural, suburban, and urban districts in Texas. The intervention and comparison schools were matched based on average demographic characteristics such student ethnicity, poverty, English language proficiency, and mobility. The analysis sample included 1,472 students from 12 intervention schools who received three years of *Saxon Math* exposure in grades 6, 7, and 8 and 1,582 students from 13 comparison schools during the 1998–99 through 2000–01 school years. Schools in the intervention group used three *Saxon Math* curricula (*Saxon 7/6*, *Saxon 8/7*, and *Saxon Algebra ½*).[8] The majority of schools in the comparison group used core basal math curricula implemented with a chapter-based approach to math instruction.

Resendez, Fahmy, and Manley (2005, Cohort F) conducted a matched-comparison quasi-experiment to investigate the effect of *Saxon Math* on the math achievement of sixth-, seventh-, and eighth-grade students in 20 middle schools located in rural, suburban, and urban districts in Texas. The intervention and comparison schools were matched based on average demographic characteristics such student ethnicity, poverty, English language proficiency, and mobility. The analysis sample included 1,526 sixth-grade students from 10 intervention schools who received one year of *Saxon Math* exposure and 1,407 students from 10 comparison schools during the 2003–04 school year. Schools in the intervention group used two *Saxon Math* curricula (*Saxon 7/6* or *Saxon 8/7*).[9] The majority of schools in the comparison group used core basal math curricula implemented with a chapter-based approach to math instruction.

Effectiveness Summary

The WWC review of *Saxon Math* for the Middle School Math topic area 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 *Saxon Math* on the mathematics achievement of middle school students. For a more detailed description of the rating of effectiveness and extent of evidence criteria, see the WWC Rating Criteria on p. 21.

Summary of effectiveness for the mathematics achievement domain

Five studies that meet standards with or without reservations reported findings in the mathematics achievement domain.

Resendez and Azin (2006) found, and the WWC confirmed, no statistically significant differences between the *Saxon Math* and comparison groups in the mathematics achievement domain. The WWC characterizes these study findings as an indeterminate effect.[10]

Crawford and Raia (1986) reported one positive and statistically significant difference between the *Saxon Math* group and the comparison group in the mathematics achievement domain. However, according to WWC calculations (correcting for clustering), this difference was not statistically significant.[11] The average effect size is considered substantively important according to WWC criteria. Therefore, the WWC characterizes these study findings as a substantively important positive effect.

Peters (1992) found, and the WWC confirmed, no statistically significant differences between the *Saxon Math* and comparison groups in the mathematics achievement domain. The WWC characterizes these study findings as an indeterminate effect.

Resendez, Fahmy, and Manley (2005, Cohort A) reported one positive and statistically significant difference between the *Saxon Math* group and the comparison group in the mathematics achievement domain. However, according to WWC calculations (correcting for clustering), this difference was not statistically significant. The WWC characterizes these study findings as an indeterminate effect.

Resendez, Fahmy, and Manley (2005, Cohort F) reported one positive and statistically significant difference between the *Saxon Math* group and the comparison group in the mathematics achievement domain. However, according to WWC calculations (correcting for clustering), this difference was not statistically significant. The average effect size is considered substantively important according to WWC criteria. Therefore, the WWC characterizes these study findings as a substantively important positive effect.

Thus, for the mathematics achievement domain, there were two studies showing substantively important positive effects and three studies showing indeterminate effects, with no studies showing a statistically significant or substantively important negative effect. 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 five studies that reported findings, the estimated impact of the intervention on outcomes in the *mathematics achievement* domain was two studies showing substantively important positive effects and three studies showing indeterminate effects.
Extent of evidence	**Criteria met**
Medium to large	Five studies that included 6,601 students in 52 schools reported evidence of effectiveness in the *mathematics achievement* domain.

References

Study that meets WWC evidence standards without reservations

Resendez, M., & Azin, M. (2006). *Saxon Math randomized control trial: Final report*. Jackson, WY: PRES Associates, Inc.

Studies that meet WWC evidence standards with reservations

Crawford, J., & Raia, F. (1986). *Analyses of eighth grade math texts and achievement*. Oklahoma City, OK: Oklahoma City Public Schools, Planning, Research, and Evaluation Department.

Peters, K. G. (1992). Skill performance comparability of two algebra programs on an eighth-grade population. *Dissertation Abstracts International, 54*(01), 77A. (UMI No. 9314428)

Resendez, M., Fahmy, A., & Manley, M. A. (2005). Cohort A. *The relationship between using Saxon Middle School Math and student performance on Texas statewide assessments*. Retrieved from http://saxonpublishers.hmhco.com/HA/correlations/pdf/s/SXMath_Middle_TX_research_web.pdf

Resendez, M., Fahmy, A., & Manley, M. A. (2005). Cohort F. *The relationship between using Saxon Middle School Math and student performance on Texas statewide assessments*. Retrieved from http://saxonpublishers.hmhco.com/HA/correlations/pdf/s/SXMath_Middle_TX_research_web.pdf

Studies that do not meet WWC evidence standards

Baldree, C. L. P. (2003). *The effectiveness of two mathematical instructional programs on the mathematics growth of eighth grade students* (Unpublished doctoral dissertation). University of Georgia, Athens. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Clay, D. W. (1998). *A study to determine the effects of a non-traditional approach to Algebra instruction on student achievement* (Unpublished master's thesis). Salem-Teikyo University, Salem, WV. (ERIC Document Reproduction Service No. ED428963). 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.

Imrisek, J. P. (1989). *Incremental development: A more effective means of mathematics instruction?* (Unpublished master's thesis). Bloomsburg University, PA. 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.

Lafferty, J. F. (1996). The links among mathematics text, students' achievement, and students' mathematics anxiety: A comparison of the incremental development and traditional texts. *Dissertation Abstracts International, 56*(08), 3014A. (UMI No. 9537085) 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.

Rentschler, R. V. (1994). The effects of Saxon's incremental review on computational skills and problem-solving achievement of sixth-grade students. *Dissertation Abstracts International, 56*(02), 484A. (UMI No. 9518017) The study does not meet 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.

Resendez, M., & Azin, M. (2007). *The relationship between using Saxon Elementary and Middle-School Math and student performance on California statewide assessments*. Jackson, WY: Saxon. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Additional source:

Resendez, M., & Azin, M. (2007). *Saxon Math and California English learner's math performance: Research brief*. Jackson, WY: Saxon.

Resendez, M., & Azin, M. (2008). *The relationship between using Saxon Math at the elementary and middle school levels and student performance on the North Carolina statewide assessment*. Jackson, WY: PRES Associates, Inc. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

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. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Roberts, F. H. (1994). The impact of the Saxon Mathematics program on group achievement test scores. *Dissertation Abstracts International, 55*(06), 1498A. (UMI No. 9430198) The study does not meet 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.

Saxon, J. (1982). Incremental development: A breakthrough in mathematics. *Phi Delta Kappan, 63*(4), 482–484. The study does not meet 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.

Walsh, T. J. (2009). The effect of Saxon Math on student achievement of sixth-grade students. *Dissertation Abstracts International, 70*(06A). (AAI3362003) The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Studies that are ineligible for review using the Middle School Math Evidence Review Protocol

Andrus, H. A. (2005). *Metacognitive instruction in the realm of sixth grade Saxon Math* (Unpublished doctoral dissertation). Mount Mary College, Milwaukee, WI. The study is ineligible for review because it does not use a comparison group design or a single case design.

Fitzpatrick, S. B. (2001). An exploratory study of the implementation of an educational technology in two eighth grade mathematics classes. *Dissertation Abstracts International, 62*(06), 2082A. (UMI No. 3016656) The study is ineligible for review because it does not examine the effectiveness of an intervention.

Harris, K. L. (2008). *Saxon Math: An analysis for middle school students at-risk of low performance* (Unpublished doctoral dissertation). Capella University, Minneapolis. The study is ineligible for review because it does not use a comparison group design or a single-case design.

McNeil, N., Grandau, L., Knuth, E., Alibali, M., Stephens, A., Hattikudur, S., & Krill, D. (2006). Middle-school students' understanding of the equals sign: The books they read can't help. *Cognition and Instruction, 24*(3), 367–385. The study is ineligible for review because it does not include an outcome within a domain specified in the protocol.

Williams, D. D. (1986). *The incremental method of teaching Algebra I*. Kansas City: University of Missouri. The study is ineligible for review because it does not use a sample within the age or grade range specified in the protocol.

Appendix A.1: Research details for Resendez & Azin, 2006

Resendez, M., & Azin, M. (2006). *Saxon Math randomized control trial: Final report*. Jackson, WY: PRES Associates, Inc.

Table A1. Summary of findings

Meets WWC evidence standards without reservations

Outcome domain	Sample size	Study findings	
		Average improvement index *(percentile points)*	Statistically significant
Mathematics achievement	500 students	+8	No

Setting

The study took place in one junior high school located in a suburban area of southwestern Ohio and one middle school located in a large city in northeastern Ohio. The junior high school served students in grades 7–9. The middle school served students in grades 5–8.

Study sample

The study sample included 543 sixth-, seventh-, and eighth-grade students in 25 classes (14 intervention classes with 303 students and 11 comparison classes with 240 students) from two Ohio schools during the 2005–06 school year.[12] Of the total study sample, about 49% were male (49% intervention and 50% comparison), 8% were special education students (7% intervention and 10% comparison), and 25% received free or reduced-price lunch (19% intervention and 33% comparison). Approximately 81% were Caucasian (91% intervention and 70% comparison), 17% were African American (8% intervention and 28% comparison), and 2% were of other racial/ethnic classifications (1% intervention and 3% comparison). None of the study sample was limited English proficient. Based on pretest percentile rankings, 18% were in the lowest quartile (10% intervention and 29% comparison), 29% were in the highest quartile (43% intervention and 11% comparison), and the remaining 53% were in the two middle quartiles (48% intervention and 60% comparison).[13] The intervention and comparison groups were formed through random assignment at the classroom level. There were seven teachers across the two schools. Six of the seven participating teachers taught at least one intervention and one comparison class. This provides an opportunity for spillover (contamination) of the intervention to the comparison group, but the report concludes that contamination was not a problem. The seventh teacher was randomly assigned to teach one *Saxon Math* class. The analysis sample included 500 students for the TerraNova Math Total test (281 intervention and 219 comparison) and 492 students for the TerraNova Math Computation Total test (280 intervention and 212 comparison).[14]

Intervention group

Students in the intervention group were taught using one of four *Saxon Math* curricula during the 2005–06 school year: (a) *Saxon Math 7/6*, (b) *Saxon Math 8/7*, (c) *Saxon Algebra ½*, or (d) *Saxon Algebra 1*. Teachers were expected to implement key program components that included: warm-up activities; teaching a new lesson concept; lesson practice; "mixed practice" that reviewed and built upon previous concepts and prepared students for upcoming lessons; teacher-directed, whole-class investigations; and test day activities.[15] According to the study authors, two of the seven teachers typically did not follow the implementation guidelines. However, the majority of intervention classrooms (10 out of 14) covered at least 83% of the 120 or more *Saxon Math* lessons.[16]

Comparison group

The math curriculum used in comparison classrooms varied by site. In the junior high school, comparison classrooms were taught using an unspecified traditional basal program for math instruction that used a modular approach, emphasized real-world application, and incorporated a variety of activities including exploration, modeling, and using tools such as technology to communicate math. In the middle school, comparison classrooms were taught using a variety of resources consisting primarily of teacher-created materials based on district and state guidelines. The curriculum also included an Internet-based math program, traditional chapter-based textbooks, and other supplemental math resources.

Outcomes and measurement

The primary outcome measures were the TerraNova Math Total and the TerraNova Math Computation Total of the CTB/McGraw-Hill TerraNova Basic Multiple Assessment with Plus Test. The pretest administration occurred between September and October 2005. The posttest administration occurred between May and June 2006. For a more detailed description of these outcome measures, see Appendix B.

Support for implementation

Teachers received about three hours of training before implementing the *Saxon Math* curricula. The *Saxon Math* trainer covered the program's philosophy, key components, and curriculum support materials. A follow-up training session conducted in October/November provided teachers with one-on-one suggestions for using the program.

Appendix A.2: Research details for Crawford & Raia, 1986

Crawford, J., & Raia, F. (1986). *Analyses of eighth grade math texts and achievement*. Oklahoma City, OK: Oklahoma City Public Schools, Planning, Research, and Evaluation Department.

Table A2. Summary of findings

Meets WWC evidence standards with reservations

Outcome domain	Sample size	Study findings	
		Average improvement index *(percentile points)*	Statistically significant
Mathematics achievement	78 students	+16	No

Setting

The study took place in four middle schools in Oklahoma City Public Schools.

Study sample

The study sample included 78 eighth-grade students (39 intervention and 39 comparison) taught by four teachers in four Oklahoma middle schools during the 1984–85 school year.[17] Each teacher taught an intervention class and a comparison class. The authors did not report demographic information. To create similar intervention and comparison groups, the researchers conducted a stratified matching procedure based on pretest total math score on the California Achievement Test (CAT) at the student level, within teachers, to match a comparison student to each student in the intervention group. When more than one student from the comparison group matched a student in the intervention group, the student match was selected at random. When no student from the comparison group matched a student in the intervention group, the student in the intervention group was excluded from the sample.

Intervention group

Students in the intervention group were taught using the *Saxon Algebra ½* (1983) textbook during the 1984–85 school year. Information about the level of implementation was not provided. The intervention was implemented by four teachers, one from each of four schools. Each of these teachers taught intervention classes and comparison classes.

Comparison group

Students in the comparison group were taught using the math textbook in place prior to the pilot study, *Scott Foresman Mathematics* (1980).

Outcomes and measurement

The primary outcome measure was the total math score on the CAT. Pretest data were from the year-end administration of the CAT in 1984, and posttest data came from the end-of-year test administration in 1985. For a more detailed description of this outcome measure, see Appendix B.

Support for implementation

Information on teacher training was not provided.

Appendix A.3: Research details for Peters, 1992

Peters, K. G. (1992). Skill performance comparability of two algebra programs on an eighth-grade population. *Dissertation Abstracts International, 54*(01), 77A. (UMI No. 9314428)

Table A3. Summary of findings

Meets WWC evidence standards with reservations

Outcome domain	Sample size	Study findings	
		Average improvement index *(percentile points)*	Statistically significant
Mathematics achievement	36 students	+6	No

Setting

The study took place in one junior high school in Nebraska. The district borders two large cities (Lincoln and Omaha) and has a mix of students living in rural and suburban locations.

Study sample

The study sample included two classrooms of the same eighth-grade teacher (for a total of 36 students) from one junior high school during the 1991–92 school year. All of the students were "math talented" based on teacher recommendations and prior academic achievement. No information is provided on the specific thresholds that were used in delineating the math-talented criteria; however, all students scored at or above the 87th percentile on the CAT total math battery. Of the total sample, 56% were female (58% intervention and 53% comparison) and 44% were male (42% intervention and 47% comparison). Students were randomly assigned to the teacher's two classrooms, and the teacher used the intervention in one classroom and the comparison curriculum in the other classroom.[18] However, the assignment of students was altered after random assignment to accommodate scheduling difficulties and student requests for other course offerings. The analysis sample included 19 students in the *Saxon Math* group and 17 students in the *UCSMP Algebra* group.

Intervention group

Students in the intervention group were taught using *Saxon Algebra 1* (1981) during the 1991–92 school year. Students in this group participated in daily math sessions for one academic year. In each session, the teacher introduced a new concept, and students had opportunities to practice the new concept and past concepts. Students were assessed every fifth lesson with study-specific unit tests of the material covered in the past few sessions.

Comparison group

Students in the comparison group were taught using the *UCSMP Algebra* curriculum. The *UCSMP Algebra* program was developed based on National Council of the Teachers of Mathematics (NCTM) objectives that emphasized problem-solving skills, reading comprehension, use of technology, and relevant lessons with real-world applications. Each lesson is organized into an introduction of the concept, a reading section that explains the process, and real-life problem situations.

Outcomes and measurement

The primary outcome measure was the Orleans-Hanna Algebra Prognosis Test.[19] This measure was administered as a pretest in August 1991 and as a posttest in May 1992. For a more detailed description of this outcome measure, see Appendix B.

Support for implementation

The teacher who taught both study groups did not have prior experience with the intervention or comparison curricula but had read extensively about both teaching formats. The teacher participated in a one-week summer workshop on *UCSMP Algebra*, and in two one-day workshops given by local consultants on both of the curricula used in this study.

Appendix A.4: Research details for Resendez, Fahmy, & Manley, 2005, Cohort A

Resendez, M., Fahmy, A., & Manley, M. A. (2005). Cohort A. *The relationship between using Saxon Middle School Math and student performance on Texas statewide assessments*. Retrieved from http://saxonpublishers.hmhco.com/HA/correlations/pdf/s/SXMath_Middle_TX_research_web.pdf

Table A4. Summary of findings **Meets WWC evidence standards with reservations**

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
Mathematics achievement	3,054 students	+5	No

Setting

The study took place in 30 Texas middle schools serving sixth-, seventh-, and eighth-grade students between 1998–99 and 2000–01.

Study sample

The study sample included three cohorts with a total of more than 16,000 sixth-, seventh-, and eighth-grade students from 30 schools (15 intervention and 15 comparison). Of the total sample, 46% were Caucasian (46% intervention and 45% comparison), 43% were Hispanic (42% intervention and 43% comparison), 11% were African American (11% intervention and 10% comparison), 6% were limited English proficient (5% intervention and 7% comparison), 15% were special education (15% intervention and 14% comparison), 49% were female (49% intervention and 48% comparison), and 46% were economically disadvantaged (43% intervention and 48% comparison). To create the matched comparison group of schools, the Texas Education Agency (TEA) identified 40 matched comparison schools from which 15 were randomly selected. The intervention and comparison schools were matched on demographic characteristics including race, ethnicity, poverty, English language proficiency, and percentage of mobile students. This review focuses on an analytic subset (referred to by the study authors as Cohort A of Sample 1) of the full study sample. The analysis sample for the portion of the study reviewed here included 3,054 students (1,472 intervention and 1,582 comparison) from 25 schools (12 intervention and 13 comparison) in grades 6–8 during the 1998–99 through 2000–01 school years.[20]

Intervention group

Intervention students were taught using *Saxon Math* curricula (*Saxon 7/6*, *Saxon 8/7*, or *Saxon Algebra ½*) during the 1998–99 (grade 6), 1999–2000 (grade 7), and 2000–01 (grade 8) school years.

Comparison group

Most comparison schools used core basal math curricula, which generally consist of a chapter-based approach to math instruction. Two schools used an investigative approach with an emphasis on making connections among various mathematical topics and between math and problems in other disciplines.

Outcomes and measurement

The primary outcome was the Texas Assessment of Academic Skills (TAAS) Texas Learning Index (TLI) math score. The pretest measure was the TLI math score from grade 5 (taken in the spring of 1998). The posttest measure was an average of the TLI math scores from grade 6 (taken in spring of 1999), grade 7 (taken in spring of 2000), and grade 8 (taken in the spring of 2001). For a more detailed description of these outcome measures, see Appendix B.

Support for implementation

Information on teacher training was not provided.

Appendix A.5: Research details for Resendez, Fahmy, & Manley, 2005, Cohort F

Resendez, M., Fahmy, A., & Manley, M. A. (2005). Cohort F. *The relationship between using Saxon Middle School Math and student performance on Texas statewide assessments*. Retrieved from http://saxonpublishers.hmhco.com/HA/correlations/pdf/s/SXMath_Middle_TX_research_web.pdf

Table A5. Summary of findings **Meets WWC evidence standards with reservations**

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
Mathematics achievement	2,933 students	+10	No

Setting

The study took place in 30 Texas middle schools serving sixth-, seventh-, and eighth-grade students during 2003–04.

Study sample

The study sample included three cohorts with a total of more than 18,000 sixth-, seventh-, and eighth-grade students from 30 schools (15 intervention and 15 comparison). Of the total sample, 42% were Caucasian (38% intervention and 45% comparison), 44% were Hispanic (47% intervention and 41% comparison), 13% were African American (14% intervention and 12% comparison), 6% were limited English proficient (5% intervention and 6% comparison), 15% were special education (14% intervention and 15% comparison), 49% were female (49% intervention and 49% comparison), and 52% were economically disadvantaged (48% intervention and 55% comparison). To create the matched comparison group, the TEA identified 40 matched comparison schools from which 15 were randomly selected. The intervention and comparison schools were matched on demographic characteristics including race, ethnicity, poverty, English language proficiency, and percentage of mobile students. This review focuses on an analytic subset (referred to by the study authors as Cohort F of Sample 3) of the full study sample. The analysis sample for the portion of the study reviewed here included 2,933 students (1,526 intervention and 1,407 comparison) from 20 schools (10 intervention and 10 comparison) in grade 6 during the 2003–04 school year.[21]

Intervention group

Intervention students were taught using the *Saxon Math* curriculum during the 2003–04 school year. The majority used *Saxon 7/6*, and the remainder used *Saxon 8/7*.

Comparison group

Most comparison schools used core basal math curricula, which generally consist of a chapter-based approach to math instruction. Two schools used an investigative approach with an emphasis on making connections among various mathematical topics and between math and problems in other disciplines.

Outcomes and measurement

The primary outcome was the Texas Assessment of Knowledge and Skills (TAKS). The pretest measure was the TAKS math performance scores from grade 5 (taken in the spring of 2003). The posttest measure was the TAKS math performance scores from grade 6 (taken in the spring of 2004). For a more detailed description of these outcome measures, see Appendix B.

Support for implementation

Information on teacher training was not provided.

Appendix B: Outcome measures for the mathematics achievement domain

Mathematics achievement	
California Achievement Test (CAT) General Mathematics Exam	The CAT is a standardized achievement test. The mathematics section includes subtests on mathematics computation and mathematics concepts and applications. Normal Curve Equivalent (NCE) scores were used in the analysis (outcome measure used in Crawford & Raia, 1986; information obtained from the test publisher website; edition of the test was not reported).
Orleans-Hanna Algebra Prognosis Test	The nationally normed Orleans-Hanna Algebra Prognosis test consists of 60 items and is used to predict student success in future algebra study by comparing the actual test score with the students' most recent math grades. In contrast to an achievement test, students are required to answer questions by following a procedure or set of operations using mathematical or verbal expressions parallel to but different from those contained in the model lessons. This test is often used to predict the ability to succeed in a first-year algebra course of study. Raw scores converted by the teacher to standard scores were used in the analysis (as cited in Peters, 1992).
TerraNova Math Computation Total Scale Score	The TerraNova Math Computation Total scale score is based on a portion of CTB/McGraw-Hill's TerraNova Basic Multiple Assessment (Level 16-sixth grade, Level 17-seventh grade, and Level 18-eighth grade) with Plus Test. This 20-minute portion of the TerraNova standardized test has 20 multiple-choice, computational problems. The objectives tested include: multiplying whole numbers (sixth grade only), dividing whole numbers (sixth grade only), decimals (sixth and seventh grade), fractions, integers (seventh and eighth grade), percents (seventh and eighth grade), and order of operations (seventh and eighth grade) (as cited in Resendez & Azin, 2006).
TerraNova Math Total (MC/CR) Scale Score	The TerraNova Math Total (MC/CR) scale score is based on a portion of CTB/McGraw-Hill's TerraNova Basic Multiple Assessment (Level 16-sixth grade, Level 17-seventh grade, and Level 18-eighth grade) with Plus Test. This 90-minute test contains 41 or 42 multiple-choice and constructed-response items, consisting of a few computational problems but mostly word problems. The objectives tested include: number and number relations; computation and estimation; measurement; geometry and spatial sense; data, statistics, and probability; patterns, function, and algebra (seventh grade and eighth grade only); problem solving and reasoning; and communication (as cited in Resendez & Azin, 2006).
TerraNova Objective Performance Indices (OPI)	CTB/McGraw-Hill provides OPI for each objective measured on the TerraNova Basic Multiple Assessment with Plus Test. The OPI is an estimate of the number of items a student or group of students could be expected to answer correctly if there had been 100 such items on the test for that objective (as cited in Resendez & Azin, 2006).
Texas Assessment of Knowledge and Skills (TAKS) Math Test	The TAKS covers numbers, operations, and quantitative reasoning; patterns, relationships, and algebraic reasoning; geometry and spatial reasoning; concepts and uses of measurement; probability and statistics; and mathematical processes and tools. A scaled score was used in the analysis (as cited in Resendez, Fahmy, & Manley, 2005, Cohort F).
Texas Learning Index (TLI) Math Score (based on the Texas Assessment of Academic Skills [TAAS])	The TAAS is a criterion-referenced state test that measures problem-solving and critical-thinking skills. The TLI is an outcome metric, based on student performance on the TAAS, allowing for comparisons between administrations and between grades. The TAAS was used in Texas from 1990–2002. It was replaced by the Texas Assessment of Knowledge and Skills in 2003 (as cited in Resendez, Fahmy, & Manley, 2005, Cohort A).

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

Outcome measure	Study sample	Sample size	Mean (standard deviation) Intervention group	Mean (standard deviation) Comparison group	WWC calculations Mean difference	WWC calculations Effect size	WWC calculations Improvement index	*p*-value
Resendez & Azin, 2006[a]								
TerraNova Math Computation Total	Grades 6–8	492 students	674.00 (49.37)	662.84 (46.47)	11.16	0.23	+9	0.06
TerraNova Math Total	Grades 6–8	500 students	679.11 (40.47)	673.32 (38.11)	5.79	0.15	+6	0.17
Domain average for mathematics achievement (Resendez & Azin, 2006)						**0.19**	**+8**	**Not statistically significant**
Crawford & Raia, 1986[b]								
California Achievement Test	Grade 8	78 students	55.56 (11.86)	50.72 (11.75)	4.84	0.41	+16	0.01
Domain average for mathematics achievement (Crawford & Raia, 1986)						**0.41**	**+16**	**Not statistically significant**
Peters, 1992[c]								
Orleans-Hanna Prognosis Test	Grade 8 (math talented)	36 students	95.67 (4.53)	95.06 (4.09)	0.61	0.14	+6	> 0.05
Domain average for mathematics achievement (Peters, 1992)						**0.14**	**+6**	**Not statistically significant**
Resendez, Fahmy, & Manley, 2005, Cohort A[d]								
Texas Learning Index Score	Grade 8	3,054 students	83.95 (6.99)	82.98 (7.62)	0.97	0.13	+5	< 0.01
Domain average for mathematics achievement (Resendez, Fahmy, & Manley, 2005, Cohort A)						**0.13**	**+5**	**Not statistically significant**
Resendez, Fahmy, & Manley, 2005, Cohort F[e]								
Texas Assessment of Knowledge and Skills Math Scale Score	Grade 6	2,933 students	2,229.02 (225.89)	2,174.49 (205.10)	54.53	0.25	+10	< 0.01
Domain average for mathematics achievement (Resendez, Fahmy, & Manley, 2005, Cohort F)						**0.25**	**+10**	**Not statistically significant**
Domain average for mathematics achievement across all studies						**0.22**	**+9**	**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 change (measured in standard deviations) in an average student's outcome that can be expected if the student is given the intervention. 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. na = not applicable.

[a] For Resendez & Azin (2006), no corrections for clustering or multiple comparisons were needed. The *p*-values presented here were reported in the original study. The authors' analysis utilized hierarchical linear modeling (HLM), which accounts for the nesting of cases within clusters (in this case, students within classrooms). This approach obviates the need for a clustering correction, which might otherwise be needed given the classroom-level random assignment. The intervention group value is the unadjusted comparison group mean plus the program coefficient from the HLM analysis. The effect sizes reported here differ from those reported in the original study due to differences in the effect-size formulas used.

[b] The *p*-values presented here were reported in the original study. For Crawford & Raia (1986), a correction for clustering was needed and results in significance levels that differ from those in the original study.

[c] For Peters (1992), no corrections for clustering or multiple comparisons were needed. The *p*-values presented here were reported in the original study. The WWC calculated the program group mean using a difference-in-differences approach (see the *WWC Procedures and Standards Handbook*) by adding the impact of the program (i.e., the difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest mean.

[d] The *p*-values presented here were reported in the original study. For Resendez, Fahmy, & Manley (2005), Cohort A, a correction for clustering was needed and results in significance levels that differ from those in the original study. The means for the *Saxon Math* group and comparison group are repeated measures ANCOVA adjusted means during grade 8 for Cohort A (grade 8 of Sample 1). The mean difference between these two scores represents the effect of three years' exposure to *Saxon Math*. The standard deviations are the unadjusted standard deviations for grade 8 provided to the WWC by the study authors in response to a query.

[e] The *p*-values presented here were reported in the original study. For Resendez, Fahmy, & Manley (2005), Cohort F, a correction for clustering was needed and results in significance levels that differ from those in the original study. The study authors provided the WWC with unadjusted standard deviations for the *Saxon Math* and comparison groups in response to a query.

Appendix D: Description of supplemental findings for the mathematics achievement domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Resendez & Azin, 2006[a]								
TerraNova OPI Communication subscale	Grades 6–8	484 students	44.07 (19.63)	43.80 (20.57)	0.27	0.01	+1	0.87
TerraNova OPI Computation and Estimation subscale	Grades 6–8	498 students	59.45 (18.65)	57.00 (19.06)	2.45	0.13	+5	0.26
TerraNova OPI Data, Statistics, & Probability subscale	Grades 6–8	503 students	54.58 (19.81)	52.23 (20.43)	2.35	0.12	+5	0.29
TerraNova OPI Decimals subscale	Grades 6 & 7	212 students	66.00 (19.13)	56.84 (20.09)	9.16	0.47	+18	0.02
TerraNova OPI Division subscale	Grade 6	102 students	59.69 (27.48)	40.84 (25.10)	18.85	0.71	+26	0.02
TerraNova OPI Fractions subscale	Grades 6–8	491 students	46.59 (26.48)	39.75 (25.30)	6.84	0.26	+10	0.06
TerraNova OPI Geometry & Spatial Sense subscale	Grades 6–8	503 students	48.84 (20.63)	46.72 (19.94)	2.12	0.10	+4	0.33
TerraNova OPI Integers subscale	Grades 7 & 8	387 students	60.45 (24.71)	50.84 (27.73)	9.61	0.37	+14	0.01
TerraNova OPI Measurement subscale	Grades 6–8	502 students	43.29 (22.52)	38.27 (22.52)	5.02	0.22	+9	0.01
TerraNova OPI Multiplication subscale	Grade 6	105 students	67.27 (25.10)	50.78 (27.20)	16.49	0.62	+23	0.18
TerraNova OPI Number and Number Relations subscale	Grades 6–8	499 students	66.22 (20.71)	64.53 (22.17)	1.69	0.08	+3	0.36
TerraNova OPI Order of Operations subscale	Grades 7 & 8	387 students	72.60 (17.76)	68.99 (23.53)	3.61	0.18	+7	0.11
TerraNova OPI Patterns, Functions, & Algebra subscale	Grades 7 & 8	396 students	55.30 (18.10)	53.24 (18.53)	2.06	0.11	+4	0.36
TerraNova OPI Percents subscale	Grades 7 & 8	384 students	57.69 (26.42)	46.25 (23.62)	11.44	0.45	+17	0.01
TerraNova OPI Problem Solving and Reasoning subscale	Grades 6–8	503 students	46.08 (20.65)	41.03 (19.57)	5.05	0.25	+10	0.01
Crawford & Raia, 1986[b]								
CAT Math Computation subtest	Grade 8	78 students	57.66 (13.35)	51.44 (14.14)	6.22	0.45	+17	0.01
CAT Math Concepts subtest	Grade 8	78 students	53.18 (12.44)	50.00 (12.40)	3.18	0.25	+10	0.10

Table Notes: The supplemental findings presented in this table are additional findings from the studies in this report that do not factor into the determination of the intervention rating. 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 change (measured in standard deviations) in an average student's outcome that can be expected if the student is given the intervention. 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. OPI = Objective Performance Indices. CAT = California Achievement Test.

[a] For Resendez & Azin (2006), a correction for multiple comparisons was needed and results in significance levels that differ from those in the original study. The *p*-values presented here were reported in the original study. Due to the multiple comparisons adjustment, the *p*-values of 0.01 for *TerraNova OPI Measurement*; 0.01 for *TerraNova OPI Problem Solving and Reasoning*; 0.02 for *TerraNova OPI Division*; 0.02 for *TerraNova OPI Decimals*; 0.01 for *TerraNova OPI Integers*; and 0.01 for *TerraNova OPI Percents* were higher than the critical *p*-value for statistical significance; therefore, the WWC does not find the result to be statistically significant. The authors' analysis utilized hierarchical linear modeling (HLM), which accounts for the nesting of cases within clusters (in this case, students within classrooms). This approach obviates the need for a clustering correction, which might otherwise be needed given the classroom-level random assignment. The intervention group value is the unadjusted comparison group mean plus the program coefficient from the HLM analysis. The effect sizes reported here differ from those reported in the original study due to differences in the effect-size formulas used.

[b] For Crawford & Raia (1986), a correction for clustering was needed and results in significance levels that differ from those in the original study. The *p*-values presented here were reported in the original study.

Endnotes

[1] The descriptive information for this program was obtained from publicly available sources: the program's website (http://saxonpublishers.hmhco.com, downloaded June 2010) and directly from the authors in the case of Resendez and Azin (2006). The WWC requests distributors review the program description sections for accuracy from their perspective. The program description was provided to the distributor in February 2012, and the WWC incorporated feedback from the distributor. 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 2011.

[2] The previous report was released in April 2010. This report has been updated to include reviews of two studies that have been released since 2010. Of the additional studies, one was not within the scope of the review protocol for the Middle School Math topic area, and one was within the scope of the review protocol for the Middle School Math topic area but did not meet evidence standards. A complete list and disposition of all studies reviewed are provided in the references. The studies in this report were reviewed using WWC Evidence Standards, version 2.1, as described in the Middle School Math review protocol, version 2.0. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.

[3] For criteria used in the determination of the rating of effectiveness and extent of evidence, see the WWC Rating Criteria on p. 21. These improvement index numbers show the average and range of student-level improvement indices for all findings across the studies.

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

[5] Two of these studies (Resendez, Fahmy, & Manley, 2005, Cohort A, and Resendez, Fahmy, & Manley, 2005, Cohort F) were included within one research report. The study authors reported on three independent samples containing multiple cohorts of students. This report includes Cohort A (grade 8 of Sample 1) and Cohort F (grade 6 of Sample 3). Data from these two cohorts were treated as separate studies because they examined the effects of *Saxon Math* on different samples of students at two different times. Students in Cohort A were in grade 8 in 2000–01; students in Cohort F were in grade 6 in 2003–04. Sample 2 of the study was excluded from WWC review because it was used for a pre-post analysis of the students in *Saxon Math* schools and did not include a comparison group. Cohorts B and C of Sample 1 were used by the study authors only in an analysis of tenth-grade math performance. Because it is unknown whether the intervention and comparison groups for these cohorts attended similar schools in grades 9 and 10, it is impossible to determine whether the effects can be attributed solely to *Saxon Math*; therefore, the WWC excluded these cohorts from this review. Cohorts G and H of Sample 3 were excluded from WWC review because pre-*Saxon Math* achievement data were not available for these students and, consequently, baseline equivalence could not be established.

[6] In the case of Resendez and Azin (2006), differences between intervention and comparison groups on pretest, and several demographic characteristics, were statistically significant at the student level but not at the classroom level. The study authors statistically controlled for these baseline differences in their analysis. Because random assignment was well executed, the WWC categorizes the study as a randomized controlled trial.

[7] The "math-talented" designation is based on teacher recommendations and prior academic achievement. No information is provided on the specific thresholds that were used in delineating the math-talented criteria; however, all students in the sample scored at or above the 87th percentile on the CAT total math battery.

[8] The authors did not provide information on the edition and publication year of the *Saxon Math* texts used in the study.

[9] The authors did not provide information on the edition and publication year of the *Saxon Math* texts used in the study.

[10] The effect size formula used by Resendez and Azin (2006) differs from the formula used by the WWC and yields a different effect size.

[11] Crawford and Raia (1986) reported, and the WWC confirmed, that there was not a statistically significant difference in posttest means between the *Saxon Math* and comparison groups and that the two groups had identical pretest means. In an additional analysis, the authors reported that controlling for pretest produced a statistically significant effect of *Saxon Math*. The WWC could not confirm the reported significance level for this analysis, and WWC calculations indicated the effect was not statistically significant.

[12] Initial study participants included 549 students (307 intervention and 242 comparison) who were enrolled at the beginning of the fall semester. The study authors included in their final study sample only students who remained in the study throughout the year. Six students left during the school year (four intervention and two comparison), resulting in a study sample of 543 students still enrolled at the end of the spring semester.

[13] Because random assignment was well executed, the WWC categorizes the study as a randomized controlled trial. The study authors statistically controlled for baseline differences between the intervention and comparison groups in their analysis.

[14] The study authors also conducted subgroup analyses. Because the authors did not report sufficient information to calculate effect sizes, the WWC excluded the subgroup analyses from this review.

[15] In the junior high school, approximately 45 minutes of additional math instruction were provided to students in need of remediation. In particular, 38 seventh-grade remedial students received a "double-dose" of instruction: 22 students were in a *Saxon Math* class and a *Saxon Math* lab, 11 students were in a *Saxon Math* class and a comparison math lab, and five students were in a comparison math class and a comparison math lab. According to the study authors, this additional math instruction did not affect the findings for the two primary measures.

[16] In the middle school, three of the four teachers stopped or reduced their use of their respective curricula for at least one to two months in order to focus on preparation for state testing in March. This disruption reduced use of both the *Saxon Math* and the comparison programs. The disruption in the *Saxon Math* group included stopping or reducing use of the *Saxon* program for six to ten weeks.

[17] The study authors reported on three "studies": one that compared *Saxon* students to all other eighth-grade students in Oklahoma City Public Schools, one that compared *Saxon* students in the pilot schools to non-*Saxon* students attending the same schools, and one that compared *Saxon* students taught by teachers who used both the *Saxon* and non-*Saxon* textbooks to non-*Saxon* students taught by the same teachers. The third study included an analysis in which the authors matched students on pretest within strata formed by teachers who used both *Saxon* and non-*Saxon* textbooks. This WWC review focuses only on this third, within-teacher matched comparison analysis because it is the only analysis for which the authors demonstrated baseline equivalence.

[18] Because both the intervention and comparison curricula were monitored on a weekly basis by the researcher to help maintain the integrity of implementation, and because there is no indication in the study that the teacher was biased toward one of the conditions, this design was accepted for review.

[19] The study author described the Orleans-Hanna Algebra Prognosis Test as the primary measure of student math achievement. The study also examined four study-generated criterion unit tests, not from the Orleans-Hanna Algebra Prognosis Test, designed to descriptively measure student understanding of algebraic components. However, the author did not provide information on the reliability or validity of these four tests. Accordingly, analyses based on these four unit tests were not considered in this version of the report.

[20] The study authors excluded from the analysis three intervention schools that were not using *Saxon Math* during the 1998–99 school year; two comparison schools were subsequently dropped. In addition, the WWC excluded Cohorts B and C from this review because they were included only in an analysis of tenth-grade math performance. Because it is unknown whether the intervention and comparison groups for these cohorts attended similar schools in ninth- and tenth-grade, it is impossible to determine whether the effects can be attributed solely to *Saxon Math*; therefore, the WWC excluded these cohorts from this review.

[21] The study authors excluded from the analysis five intervention schools that were not using *Saxon Math* during the 2003–04 school year; five comparison schools were subsequently dropped. In addition, the WWC excluded Cohorts G and H from review because pre-*Saxon Math* achievement data were not available for these students and, consequently, baseline equivalence could not be established.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013, February). *Middle School Math intervention report: Saxon Math*. 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. 21.

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. 21.

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.