WWC Intervention Report

A summary of findings from a systematic review of the evidence

INSTITUTE OF EDUCATION SCIENCES

Secondary Mathematics May 2016*

Saxon Math

Program Description[1]

Saxon Math is a core curriculum for students in grades K–12 that uses an incremental approach to 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. This review focuses on studies of *Saxon Math's* secondary courses, including *Saxon Algebra I, Saxon Geometry, Saxon Algebra II,* and *Saxon Advanced Math*.

Research[2]

The What Works Clearinghouse (WWC) identified two studies of *Saxon Algebra I* that both fall within the scope of the Secondary Mathematics topic area and meet WWC group design standards. One study meets WWC group design standards without reservations and the other study meets WWC group design standards with reservations. Together, these studies included 198 secondary students in grades 8–9 in two locations.

The WWC considers the extent of evidence for *Saxon Algebra I* on the mathematics achievement of secondary students to be small for the algebra domain. There were no studies that meet WWC group design standards in the five other domains, so this intervention report does not report on the effectiveness of *Saxon Math* for those domains. (See the Effectiveness Summary on p. 5 for more details of effectiveness by course and domain.)

The findings in this report pertain to *Saxon Algebra I* only. No studies of *Saxon Geometry, Saxon Algebra II*, or *Saxon Advanced Math* fall within the scope of the Secondary Mathematics review protocol and meet WWC group design standards.

Effectiveness

Saxon Algebra I was found to have no discernible effects on algebra for secondary students.

Report Contents

This intervention report presents findings from a systematic review of *Saxon Math* conducted using the WWC Procedures and Standards Handbook, version 3.0, and the Secondary Mathematics review protocol, version 3.1.

Table 1. Summary of findings[3]

Course and outcome domain	Rating of effectiveness	Improvement index (percentile points)		Number of studies	Number of students	Extent of evidence
		Average	Range			
Saxon Algebra I						
Algebra	No discernible effects	+5	+4 to +6	2	198	Small
Saxon Geometry	No evidence					
Saxon Algebra II	No evidence					
Saxon Advanced Math	No evidence					

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

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

- Each daily lesson consists of learning a new mathematical concept, working on practice problems relating to that concept, and solving a number of problems that include the current and previous material.
- A relatively small set of new math ideas is introduced daily using examples, mathematical conversations, and practice, such that new ideas and concepts are integrated with ones that were previously introduced.
- The lessons include 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.

The secondary curriculum includes *Saxon Algebra I, Saxon Geometry, Saxon Algebra II*, and *Saxon Advanced Math*, which can be purchased separately.

Cost

As of November 2015, the student editions of *Saxon Algebra I, Saxon Algebra II, Saxon Geometry*, and *Saxon Advanced Math* cost $82.90, $86.45, $93.60, and $89.95, respectively. The teacher editions of each subject cost $84.30, $92.20, $131.45, and $91.35, respectively. Other curriculum materials include eBooks, student workbooks, and teaching materials, and range in price from $10.85 for a *Saxon Algebra I* Student Practice Workbook to $2,244.15 for 6-year *Saxon Algebra I* online edition with Destination Math.

Research Summary

This research summary includes information from studies of all available *Saxon Math* courses for secondary students.

Table 2. Scope of reviewed research

Grade	8–12
Delivery method	Whole class
Program type	Curriculum

The WWC identified 11 eligible studies for review:

- Ten eligible studies investigated the effects of *Saxon Algebra I* on the mathematics achievement of secondary students.
- One eligible study investigated the cumulative effect of 3 years of *Saxon Math* courses (including *Saxon Algebra I, Saxon Algebra II*, and *Saxon Advanced Mathematics*) on the mathematics achievement of secondary students.[4]

The WWC reviewed the 11 eligible studies against group design standards. One study (Pierce, 1984) is a randomized controlled trial that meets WWC group design standards without reservations, and one study (Peters, 1992) is a randomized controlled trial that meets WWC group design standards with reservations. Those two studies are summarized in this report. The remaining nine studies do not meet WWC group design standards.

An additional 46 studies were identified but do not meet WWC eligibility criteria for review in this topic area. Citations for all 57 studies are in the References section, which begins on p. 6.

Summary of *Saxon Algebra I* study meeting WWC group design standards without reservations

Pierce (1984) conducted a randomized controlled trial in one high school in the southern midwest United States. The study sample included ninth-grade students enrolled in beginning algebra. Two eligible teachers were randomly selected to participate in the study. For each teacher, half of their classes were randomly assigned to receive *Saxon Algebra I*, while the other half were randomly assigned to receive the business-as-usual algebra curriculum (*Holt Algebra I*). While the study author refers to the study as using a quasi-experimental design, this random assignment of classes led the WWC to review the study as a randomized controlled trial. Students were assigned via the school's normal computerized scheduling procedures. The author notes that students who were assigned to beginning algebra after the third week of school were assigned to one of the 12 other classrooms not in the study. The final sample included 162 students (77 intervention and 85 comparison).

Summary of *Saxon Algebra I* study meeting WWC group design standards with reservations

Peters (1992) conducted a randomized controlled trial that investigated the effect of *Saxon Algebra I* on the mathematics 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.[5] The district borders two large cities (Lincoln and Omaha), and its students lived in rural and suburban areas. Students in the intervention group used *Saxon Algebra I* (1981), while students in the comparison group used the *University of Chicago School Mathematics Project (UCSMP) Algebra I* First-Edition textbook. 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 demonstrated baseline equivalence on the analysis sample and therefore *meets WWC group design standards with reservations*.

Effectiveness Summary

The WWC review of *Saxon Algebra I* for the Secondary Mathematics topic area includes student outcomes in six domains: algebra, geometry, statistics and probability, trigonometry/precalculus, calculus, and general mathematics achievement. The two studies of *Saxon Algebra I* that meet WWC group design standards reported findings in one of the six domains: algebra. The findings below present the authors' estimates and WWC-calculated estimates of the size and statistical significance of the effects of *Saxon Algebra I* on secondary students. Additional comparisons are presented as supplemental findings in the appendix. The supplemental findings do not factor into the intervention's rating of effectiveness. For a more detailed description of the rating of effectiveness and extent of evidence criteria, see the WWC Rating Criteria on p. 19.

Summary of *Saxon Algebra I* effectiveness for the algebra domain

Table 3. Rating of effectiveness of *Saxon Algebra I* and extent of evidence for the algebra domain

Rating of effectiveness	Criteria met
No discernible effects *No evidence of statistically significant or substantively important effects, either positive or negative.*	In the two studies that reported findings, the estimated impact of the intervention on outcomes in the *algebra* domain was neither statistically significant nor large enough to be considered substantively important.
Extent of evidence	**Criteria met**
Small	Two studies that included 198 students in two schools reported evidence of effectiveness in the *algebra* domain.

Two studies of *Saxon Algebra I* that meet WWC group design standards with or without reservations reported findings in the algebra domain.

Pierce (1984) found, and the WWC confirmed, no statistically significant or substantively important differences between the *Saxon Algebra I* and comparison groups in the algebra domain. The WWC characterizes these study findings as an indeterminate effect.

Peters (1992) found, and the WWC confirmed, no statistically significant or substantively important differences between the *Saxon Algebra I* and comparison groups in the algebra domain. The WWC characterizes these study findings as an indeterminate effect.

Thus, for the algebra domain, two studies of *Saxon Algebra I* showed indeterminate effects. This results in a rating of no discernible effects, with a small extent of evidence.

References

Study of *Saxon Algebra I* that meets WWC group design standards without reservations

Pierce, R. D. (1984). *A quasi-experimental study of Saxon's Incremental Development Model and its effects on student achievement in first-year algebra* (Unpublished doctoral dissertation). University of Tulsa, OK.

Study of *Saxon Algebra I* that meet WWC group design standards with reservations

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)

Studies of *Saxon Algebra I* that do not meet WWC group design standards

Abrams, B. J. (1989). *A comparison study of the Saxon Algebra I text* (Unpublished doctoral dissertation). University of Colorado, Boulder. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Aquino, A., & Zoet, C. (1985). Reinforcement in Algebra I: A study in the use of the Saxon Algebra I textbook. *Mathematics in Michigan*, 23–28. The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention.

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. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Denson, P. S. (1989). *A comparison of the effectiveness of the Saxon and Dolciani texts and theories about the teaching of high school algebra* (Unpublished doctoral dissertation). Claremont Graduate University, CA. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Johnson, D. M., & Smith, B. (1987). An evaluation of Saxon's algebra text. *Journal of Educational Research, 81*(2), 97–102. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Lawrence, L. K. (1992). *The long-term effects of an incremental development model of instruction upon student achievement and student attitude toward mathematics* (Unpublished doctoral dissertation). University of Tulsa, OK. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Mayers, K. S. (1995). *The effect of using the Saxon Algebra I textbook on the achievement of ninth-grade Algebra I students from 1989–1993* (Unpublished doctoral dissertation). Delta State University, Cleveland, MS. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Williams, D. D. (1986). *The incremental method of teaching Algebra I*. Kansas City: University of Missouri. The study does not meet WWC group design standards because the eligible outcomes do not meet WWC requirements.

Study of *multiple Saxon courses* that does not meet WWC group design standards

Sanders, B. B. (1997). *The effects of using the Saxon Mathematics method of instruction vs. a traditional method of mathematical instruction on the achievement of high school juniors*. Americus: Georgia Southwestern State University. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Studies of *Saxon Geometry* that meet WWC group design standards

None.

Studies of *Saxon Algebra II* that meet WWC group design standards

None.

Studies of *Saxon Advanced Algebra* that meet WWC group design standards

None.

Studies of *Saxon Math* that are ineligible for review using the Secondary Mathematics Evidence Review Protocol

Agodini, R., Harris, B., Seftor, N., Remillard, J., & Thomas, M. (2013). *After two years, three elementary math curricula outperform a fourth* (NCEE 2013-4019). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. This study is ineligible for review because it does not use a sample aligned with the protocol.

Additional sources:

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

Agodini, R., Harris, B., Atkins-Burnett, S., Heaviside, S., Novak, T., & Murphy, R. (2009). *Achievement effects of four early elementary school math curricula: Findings from first graders in 39 schools* (NCEE 2009–4052). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

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

Agodini, R., Harris, B., Thomas, M., Murphy, R., & Gallagher, L. (2010). *Achievement effects of four early elementary school math curricula: Findings for first and second graders* (NCEE 2011–4001). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Clements, D. H., Agodini, R., & Harris, B. (2013). *Instructional practices and student math achievement: Correlations from a study of math curricula* (NCEE 2013-4020). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Remillard, J. T., Harris, B., & Agodini, R. (2014). The influence of curriculum material design on opportunities for student learning. *ZDM: The International Journal on Mathematics Education, 46*(5), 735–749.

Andrus, H. A. (2005). *Metacognitive instruction in the realm of sixth grade Saxon Math* (Unpublished doctoral dissertation). Mount Mary College, Milwaukee, WI. This study is ineligible for review because it does not use an eligible design.

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. This study is ineligible for review because it does not use a sample aligned with the protocol.

Bell, G. (2012). The effects of Saxon Math instruction on middle school students' mathematics achievement. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 73*(4-A), 1339. This study is ineligible for review because it does not use a sample aligned with the protocol.

Bhatt, R., & Koedel, C. (2012). Large-scale evaluations of curricular effectiveness: The case of elementary mathematics in Indiana. *Educational Evaluation and Policy Analysis, 34*(4), 391–412. This study is ineligible for review because it does not use a sample aligned with the protocol.

Black, A. M. (2012). *Student motivation in math* (Unpublished doctoral dissertation). Alaska Pacific University, Anchorage. This study is ineligible for review because it is out of scope of the protocol.

Bolser, S., & Gilman, D. A. (2003). *Saxon Math, Southeast Fountain Elementary School: Effective or ineffective?* (ERIC Document Reproduction Service No. ED 474537). This study is ineligible for review because it does not use a sample aligned with the protocol.

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. This study is ineligible for review because it does not use a sample aligned with the protocol.

Christofori, P. (2005). *The effect of direct instruction math curriculum on higher-order problem solving* (Unpublished doctoral dissertation). University of South Florida, Sarasota. This study is ineligible for review because it does not use an eligible design.

Clewell, B. C., Cosentino de Cohen, C., Campbell, P. B., & Perlman, L. (2005). *Review of evaluation studies of mathematics and science curricula and professional development models.* Washington, DC: The Urban Institute. Retrieved from http://www.urban.org/ This study is ineligible for review because it does not use an eligible design.

Crawford, J., & Raia, F. (1986). *Analyses of eighth grade math texts and achievement*. Oklahoma City: Oklahoma City Public Schools, Planning, Research, and Evaluation Department. This study is ineligible for review because it does not use a sample aligned with the protocol.

Crowe, E. C., Connor, C. M., & Mazzocco, M. M. M. (2011). *Examining the impact of child x instruction interactions in first grade* [Abstract]. Evanston, IL: Society for Research on Educational Effectiveness. This study is ineligible for review because it is out of scope of the protocol.

Cummins-Colburn, B. J. L. (2007). Differences between state-adopted textbooks and student outcomes on the Texas Assessment of Knowledge and Skills examination. *Dissertation Abstracts International, 68*(06A), 168-2299. This study is ineligible for review because it does not use a sample aligned with the protocol.

Educational Research Institute of America. (2009). *A longitudinal analysis of state mathematics scores for Florida schools using Saxon Math No. 365*. Bloomington, IN: Author. This study is ineligible for review because it does not use an eligible design.

Educational Research Institute of America. (2009). *A longitudinal analysis of state mathematics scores for Indiana schools using Saxon Math No. 362*. Bloomington, IN: Author. This study is ineligible for review because it does not use an eligible design.

Educational Research Institute of America. (2009). *A longitudinal analysis of state mathematics scores for Oklahoma schools using Saxon Math No. 363*. Bloomington, IN: Author. This study is ineligible for review because it does not use an eligible design.

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), 2681A. This study is ineligible for review because it does not use an eligible design.

Good, K., Bickel, R., & Howley, C. (2006). *Saxon Elementary Math program effectiveness study.* Charlestown, WV: Edvantia. This study is ineligible for review because it does not use a sample aligned with the protocol.

Hansen, E., & Greene, K. (2000). *A recipe for math. What's cooking in the classroom: Saxon or traditional?* This study is ineligible for review because it does not use a sample aligned with the protocol.

Harcourt Achieve, Inc. (2005). *Case study research summaries of Saxon Math*. Boston, MA: Author. Retrieved from http://www.hmhco.com/ This study is ineligible for review because it is out of scope of the protocol.

Harris, K. L. (2008). *Saxon Math: An analysis for middle school students at-risk of low performance* (Unpublished doctoral dissertation). Capella University, Minneapolis, MN. This study is ineligible for review because it does not use an eligible design.

Imrisek, J. P. (1989). *Incremental development: A more effective means of mathematics instruction?* (Unpublished master's thesis). Bloomsburg University, PA. This study is ineligible for review because it does not use a sample aligned with the protocol.

Klein, D. (2000). *High achievement in mathematics: Lessons from three Los Angeles elementary schools*. Washington, DC: Brookings Institution Press. This study is ineligible for review because it does not use an eligible design.

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. This study is ineligible for review because it does not use an eligible design.

McBee, M. (1982). *Dolciani versus Saxon: A comparison of two algebra I textbooks with high school students.* Oklahoma City: Oklahoma City Public Schools. This study is ineligible for review because it is out of scope of the protocol.

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. This study is ineligible for review because it does not use a sample aligned with the protocol.

Additional source:

McNeil, N. M., Grandau, L., Stephens, A. C., Krill, D. E., Alibali, M. W., & Knuth, E. J. (2004). Middle-school students' experience with the equal sign: Saxon Math ≠ Connected Mathematics. In D. McDougall (Ed.), *Proceedings of the XXVI Annual Conference of the North American Chapter of the International Group for the Psychology of Mathematics Education* (PME-NA), Toronto, Canada (Vol. 1, pp. 271–276). Columbus, OH: ERIC.

Nguyen, K., & Elam, P. (1993). *The 1992-93 Saxon Mathematics program evaluation report*. Oklahoma City: Oklahoma City Public Schools. This study is ineligible for review because it is out of scope of the protocol.

Plato, J. (1998). *An evaluation of Saxon Math at Blessed Sacrament School*. University of Illinois at Urbana-Champaign. Retrieved from http://lrs.ed.uiuc.edu/. This study is ineligible for review because it does not use an eligible design.

Rau, M. A., Aleven, V., & Rummel, N. (2015). Successful learning with multiple graphical representations and self-explanation prompts. *Journal of Educational Psychology, 107*(1), 30–46. This study is ineligible for review because it is out of scope of the protocol.

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. This study is ineligible for review because it does not use an eligible design.

Resendez, M., & Azin, M. (2005). *The relationship between using Saxon elementary and middle school math and student performance on Georgia statewide assessments*. Jackson, WY: PRES Associates. This study is ineligible for review because it is out of scope of the protocol.

Resendez, M., & Azin, M. (2006). *Saxon Math randomized control trial: Final report*. Jackson, WY: PRES Associates. This study is ineligible for review because it is out of scope of the protocol.

Resendez, M., & Azin, M. (2007). *The relationship between using Saxon elementary and middle school math and student performance on California statewide assessments: Final report*. Jackson, WY: PRES Associates. This study is ineligible for review because it is out of scope of the protocol.

Additional source:

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

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. This study is ineligible for review because it is out of scope of the protocol.

Resendez, M., Fahmy, A., & Azin, M. (2005). *Cohort A. The relationship between using Saxon middle school math and student performance on Texas statewide assessments*. Jackson, WY: PRES Associates. This study is ineligible for review because it is out of scope of the protocol.

Resendez, M., Fahmy, A., & Azin, M. (2005). *Cohort F. The relationship between using Saxon middle school math and student performance on Texas statewide assessments*. Jackson, WY: PRES Associates. This study is ineligible for review because it is out of scope of the protocol.

Roan, C. (2012). *A comparison of elementary mathematics achievement in Everyday Math and Saxon Math schools in Illinois* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3507509) This study is ineligible for review because it does not use a sample aligned with the protocol.

Roberts, F. H. (1994). The impact of Saxon Mathematics program on group achievement test scores. *Dissertation Abstracts International, 55*(06), 1498A. (UMI No. 9430198) This study is ineligible for review because it is out of scope of the protocol.

Saxon, J. (1982). Incremental development: A breakthrough in mathematics. *Phi Delta Kappan, 63*(4), 482–484. This study is ineligible for review because it is out of scope of the protocol.

Severns, L. D. (2014). Saxon Math and student achievement: A multiyear investigation (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 1662809559) This study is ineligible for review because it is out of scope of the protocol.

Silvious, N. B. (2008). *Effects of Saxon Math program of instruction on the mathematics achievement of students with learning disabilities in grades 2 through 8*. Chester, PA: Widener University. This study is ineligible for review because it does not use a sample aligned with the protocol.

Slavin, R. E., & Lake, C. (2007). Effective programs in elementary mathematics: A best-evidence synthesis. *The Best Evidence Encyclopedia, 1*(2). Retrieved from http://www.bestevidence.org/ This study is ineligible for review because it does not use an eligible design.

Additional source:

Slavin, R. E., & Lake, C. (2009). *Effective programs for elementary mathematics: A best evidence synthesis*. Educator's summary. Retrieved from http://www.bestevidence.org/

Slavin, R. E., Lake, C., & Groff, C. (2009). Effective programs in middle and high school mathematics: A best-evidence synthesis. *Review of Educational Research, 79*(2), 839–911. This study is ineligible for review because it does not use an eligible design.

Steed, J. E. (2012). *A comparison of standardized test scores on the open-ended math portion of the Alabama reading and mathematics test between students who received Saxon Math instruction and students that received Alabama Math, Science, and Technology Initiative instructions* (Unpublished doctoral dissertation). Troy University, New York. This study is ineligible for review because it is out of scope of the protocol.

Vinogradova, E., King, C., & Rhoades, T. (2008, July). *Success for all students: What works? Best practices in Maryland public schools.* Paper presented at the annual meeting of the American Sociological Association, Boston, MA. This study is ineligible for review because it is out of scope of the protocol.

Walsh, T. J. (2009). The effect of Saxon Math on student achievement of sixth-grade students. *Dissertation Abstracts International, 70*(06A). (AAI3362003) This study is ineligible for review because it is out of scope of the protocol.

Appendix A.1: Research details for Pierce (1984)

Pierce, R. D. (1984). *A quasi-experimental study of Saxon's Incremental Development Model and its effects on student achievement in first-year algebra* (Unpublished doctoral dissertation). University of Tulsa, OK.

Table A1. Summary of *Saxon Algebra I* findings **Meets WWC group design standards without reservations**

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
Algebra	162 students	+4	No

Setting The study was conducted in a high school in the southern midwest United States.

Study sample The study sample included ninth-grade students enrolled in beginning algebra. Prior to the start of the school year, teachers with at least 3 years of experience and an even number of ninth-grade beginning algebra classes were identified. Two of the teachers who agreed to participate were randomly selected to take part in the study. One teacher taught four classes of beginning algebra, and the other taught two classes. For each teacher, half of their classes were randomly assigned to receive *Saxon Math*, while the other half were randomly assigned to receive the business-as-usual algebra curriculum (*Holt Algebra I*). Participating students were assigned via the schools' normal computerized class scheduling procedures (82 intervention students and 92 comparison). Students who were assigned to beginning algebra after the third week of school were assigned to one of 12 non-study classrooms. Analytic sample sizes for the unadjusted analysis were 77 intervention students and 85 comparison students. The school district was described as primarily White and middle or upper-middle class. Ninth graders enrolled in beginning algebra were on the average mathematics track and constituted the majority of ninth graders at the study school.

Intervention group Students in the intervention group received algebra instruction using *Saxon Algebra I*. Instruction in both conditions occurred over the course of an entire academic year during daily 55-minute math instructional blocks. Teachers organized their classroom instruction in "equivalent" ways for both the intervention and comparison sections, with respect to structure: utilizing 10–15 minutes for homework review, 10 minutes for review of the text, 15 minutes to review new material, and 15–20 minutes of work on problems due during the next class period. Class pacing was similar between the two groups; however, instructional approaches and problem set content were different based on the differential approaches of the intervention and comparison curricula. For example, students in the intervention group received ongoing review in accordance with *Saxon's Incremental Development Model*.

Comparison group Students in the comparison group received instruction using *Holt Algebra I*, the school's business-as-usual algebra curriculum. This curriculum followed standard mathematical pedagogy, including chapter reviews and practice problems, with problem sets and quarterly review tests.

Outcomes and measurement

The primary outcome measure used in the study was the Lankton First-Year Algebra Test, a standardized and normed multiple choice algebra assessment. The assessment was administered at the conclusion of the school year. The author also reported subscales defined by their alignment with instructional objectives and instructional content, which were defined as part of the Lankton assessment. These scores are presented in Appendix D. All scores on the Lankton measures are reported as the raw number of test items answered correctly. The author presented supplemental findings for several subscores on the assessment. The supplemental findings do not factor into the intervention's rating of effectiveness. For a more detailed description of the outcome measure, see Appendix B.

Support for implementation

The two study teachers met with each other weekly and with the researcher monthly during the school year. The meetings were designed as an opportunity to discuss common problems encountered implementing *Saxon Algebra I* and share ideas about the study. Teachers agreed at the outset of the study to provide "equivalent" instruction to each study condition to maintain the integrity of the study.

Appendix A.2: 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 A2. Summary of *Saxon Algebra I* findings **Meets WWC group design standards with reservations**

Outcome domain	Sample size	Study findings	
		Average improvement index *(percentile points)*	Statistically significant
Algebra	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 36 students from two classrooms taught by the same eighth-grade teacher in 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 California Achievement Test (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 curriculum in one classroom and the comparison curriculum in the other. However, the assignment of students was altered after random assignment to accommodate scheduling difficulties and student requests for other course offerings. The analytic sample included 19 students in the *Saxon Math* group and 17 students in the comparison group.

Intervention group Students in the intervention group were taught using *Saxon Algebra I* (1981) during the 1991–92 school year. Students participated in daily math lessons for one academic year. In each lesson, 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 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.[6] This measure was administered as a pretest in August 1991 and as a posttest in May 1992. For the pretest measure, the author reported both a standardized score and a raw score. Only the standardized score is used in this review because, per the author, the standardized score allows comparability between the pretest and posttest. For a more detailed description of the 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 instructional approaches. The teacher participated in a 1-week summer workshop on *UCSMP Algebra*, and in two 1-day workshops given by local consultants on both curricula used in the study. The study author also conducted weekly monitoring to help maintain implementation integrity.

Appendix B: Outcome measures for the algebra domain

Algebra	
Lankton First-Year Algebra Test	The Lankton First-Year Algebra Test is a standardized and normed algebra assessment consisting of 50 multiple-choice questions. The test is designed to assess knowledge of the basic objectives and content of a first-year algebra course. The objectives include an understanding of the concepts of number, set, operation, structure, and relation. Test items are also organized into content categories, including: defining terms, meanings of signs and symbols; algebraic operations, factoring and square roots; equations and inequalities; algebraic expressions and formulas, functions, variation, problem solving; and graphic representation. Reported reliability for this measure was 0.86 (as cited in Pierce, 1984). The author also reports subscales defined by instructional objectives and content. The five content subscales include: 1. Definition of terms, meaning of signs and symbols; 2. Fundamental algebraic operations, factoring and extracting roots; 3. Equations and inequalities; 4. Algebraic expressions and formulas, functions, variation, problem solving; and 5. Graphic representation. The five instructional objectives include: 1. Number; 2. Set; 3. Operation; 4. Structure; and 5. Relation. The subscores are reported as supplemental findings in Appendix D.
Orleans-Hanna Algebra Prognosis Test	The 60-item nationally normed Orleans-Hanna Algebra Prognosis Test is used to place and group students in algebra courses and inform the development of lessons plans. It was developed in 1928 and revised in 1980. This measure was administered as a pretest in August 1991 and as a posttest in May 1992. For the pretest measure, the author reports both a standardized score and a raw score. Only the standardized score is used in this review because, per the author, the standardized score allows comparability between the pretest and posttest. The reported sample Kuder-Richardson reliability estimate was .96 (as cited in Peters, 1992).

Appendix C: Findings included in the rating for studies of Saxon Algebra I for the algebra domain

			Mean (standard deviation)		WWC calculations			
Outcome measure	**Study sample**	**Sample size**	**Intervention group**	**Comparison group**	**Mean difference**	**Effect size**	**Improvement index**	***p*-value**
Pierce (1984)[a]								
Lankton First-Year Algebra Test	Grade 9	162 students	22.88 (6.34)	22.34 (5.48)	0.54	0.09	+4	.44
Domain average for algebra (Pierce, 1984)						**0.09**	**+4**	**Not statistically significant**
Peters (1992)[b]								
Orleans-Hanna Algebra Prognosis Test	Grade 8 (math-talented)	36 Students	95.63 (4.53)	95.06 (4.09)	0.61	0.14	+6	> .05
Domain average for algebra (Peters, 1992)						**0.14**	**+6**	**Not statistically significant**
Domain average for algebra across all studies						**0.11**	**+5**	**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 outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual 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. Some statistics may not sum as expected due to rounding. na = not applicable.

[a] For Pierce (1984), a correction for clustering was needed but did not affect whether any of the contrasts were found to be statistically significant. The *p*-value presented here was reported in the original study. The author also conducted an ANCOVA analysis that adjusted for student pretest scores on a measure of mathematical knowledge. However, due to missing data in the ANCOVA analysis, the WWC focuses on the unadjusted analysis reported in the study. Reported *p*-values are similar across the two analyses. Because pretest standard deviations were not reported, calculation of the program group mean using a difference-in-differences approach was not feasible. This study is characterized as having an indeterminate effect because the estimated effect is neither statistically significant nor substantively important. For more information, please refer to the WWC Standards and Procedures Handbook (version 3.0), p. 26.

[b] For Peters (1992), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The WWC calculated the program group mean using a difference-in-differences approach (see WWC Procedures and Standards Handbook) by adding the impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. Please see the WWC Procedures and Standards Handbook (version 3.0) for more information. The *p*-value presented here was reported in the original study. This study is characterized as having an indeterminate effect because the estimated effect is neither statistically significant nor substantively important. For more information, please refer to the WWC Standards and Procedures Handbook (version 3.0), p. 26.

Appendix D: Description of supplemental findings for studies of *Saxon Algebra I* for the algebra domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	*p*-value
Pierce (1984)[a]								
Lankton First-Year Algebra Test: Content Subscale 1–Defining terms, meanings of signs and symbols	Grade 9	162 students	4.23 (1.64)	3.84 (1.58)	0.39	0.24	+10	.21
Lankton First-Year Algebra Test: Content Subscale 2–Algebraic operations, factoring and square roots	Grade 9	162 students	5.47 (2.08)	5.97 (1.79)	–0.50	–0.26	–10	.15
Lankton First-Year Algebra Test: Content Subscale 3–Equations and inequalities	Grade 9	162 students	6.29 (2.19)	6.04 (2.08)	0.25	0.12	+5	.37
Lankton First-Year Algebra Test: Content Subscale 4–Algebraic expressions and formulas, functions, variation, problem solving	Grade 9	162 students	3.53 (1.73)	3.53 (1.59)	0.00	0.00	0	.98
Lankton First-Year Algebra Test: Content Subscale 5–Graphic representation	Grade 9	162 students	3.38 (1.62)	2.93 (1.62)	0.45	0.28	+11	.04
Lankton First-Year Algebra Test: Instructional Subscale 1–Number	Grade 9	162 students	6.27 (1.80)	6.13 (2.01)	0.14	0.07	+3	.44
Lankton First-Year Algebra Test: Instructional Subscale 2–Set	Grade 9	162 students	3.78 (1.61)	3.58 (1.64)	0.20	0.12	+5	.30
Lankton First-Year Algebra Test: Instructional Subscale 3–Operation	Grade 9	162 students	4.84 (1.92)	5.02 (1.55)	–0.18	–0.10	–4	.50
Lankton First-Year Algebra Test: Instructional Subscale 4–Structure	Grade 9	162 students	5.26 (2.05)	4.87 (1.84)	0.39	0.20	+8	.17
Lankton First-Year Algebra Test: Instructional Subscale 5–Relation	Grade 9	162 students	2.75 (1.34)	2.74 (1.22)	0.01	0.01	0	.76

Table Notes: The supplemental findings presented in this table are additional findings from studies in this report that meet WWC design standards with or without reservations, but 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 outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. Some statistics may not sum as expected due to rounding.

[a] For Pierce (1984), the *p*-values presented here were reported in the original study. A correction for clustering was needed and resulted in a WWC-computed *p*-value of .50 for *Lankton First-Year Algebra Test: Content Subscale 5–Graphic representation*; therefore, the WWC does not find the result to be statistically significant. The author also conducted an ANCOVA analysis that adjusted for student pretest scores on a measure of mathematical knowledge. However, due to missing data in the ANCOVA analysis, the WWC focuses on the unadjusted analysis reported in the study. Reported *p*-values are similar across the two analyses. Because pretest means and standard deviations were not reported, calculation of the program group mean using a difference-in-differences approach was not feasible.

Endnotes

[*] Due to the 2015 restructuring of the Mathematics topic area from three areas (Elementary, Middle, and High School) to two areas (Primary and Secondary Mathematics), this report is considered a new report rather than an updated report. The information in this report combines the research examined in the prior reports and presents the conclusions differently.

[1] The descriptive information for this program was obtained from a publicly available source: the publisher's website (www.hmhco.com). The WWC requests publishers review the program description sections for accuracy from their perspective. The program description was provided to the publisher in August 2014, and the WWC incorporated feedback from the publisher. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review.

[2] The literature search reflects documents publicly available by November 2015. Two previous intervention reports were released in February 2011 and February 2013 under the High School Math and Middle School Math topic areas, respectively. This report includes reviews of 33 studies that were not included in the prior reports. Of the additional studies, 32 were not within the scope of the review protocol for the Secondary Mathematics topic area, and one was within the scope of the review protocol but did not meet WWC group design standards. A complete list and disposition of all studies reviewed are provided in the references.

The report includes reviews of all studies that previously met WWC group design standards with or without reservations and resulted in a revised disposition for five studies.

Resendez and Azin (2006) and Resendez, Fahmy, and Manley (2005), in the Middle School Mathematics intervention report on *Saxon Math* issued in February 2013, are excluded from this report because they fall outside of the scope of the Secondary Mathematics review protocol, since the majority of the samples in each study are using a primary mathematics intervention. An author query was conducted to request results for the subsample of students who are eligible for review under the Secondary Mathematics review protocol, but no data to support such an analysis were received.

Crawford and Raia (1986), also included in the Middle School Mathematics intervention report on *Saxon Math* issued in February 2013, is excluded from this review because it is not eligible for review under the Secondary Mathematics review protocol. The sample uses a primary mathematics intervention (*Saxon Algebra ½*) and is eligible for review under the Primary Mathematics topic area.

Pierce (1984) was previously listed as not eligible for review, given the study's publication date based on an earlier evidence review protocol; however, the current Secondary Mathematics review protocol specifies a publication timeframe of 1983 or later, meaning this study is now eligible for review.

In the previous report, Abrams (1989) was reported to meet WWC group design standards with reservations. However, an updated review determined the study does not establish equivalence of the analytic sample used to estimate impacts, and no additional information is available, so it does not meet WWC group design standards in this report.

The studies in this report were reviewed using the Standards from the WWC Procedures and Standards Handbook (version 3.0), along with those described in the Secondary Mathematics topic area review protocol (version 3.1). 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. 19. These improvement index numbers show the average and range of student-level improvement indices for all findings across the studies.

[4] No studies examining the effectiveness of S*axon Geometry, Saxon Algebra II,* and *Saxon Advanced Math* fall within the scope of the Secondary Mathematics review protocol and meet WWC group design standards. Because no studies meet WWC group design standards at this time, the WWC is unable to draw any conclusions based on research about the effectiveness or ineffectiveness of *Saxon Geometry, Saxon Algebra II*, and *Saxon Advanced Math* on secondary students. Additional research that meets WWC standards is needed to determine the effectiveness or ineffectiveness of these courses.

[5] 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 California Achievement Test (CAT) Total Math Battery.

[6] 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.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2016, May). *Secondary Mathematics 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 group design standards without reservations	A study that provides strong evidence for an intervention's effectiveness, such as a well-implemented RCT.
Meets WWC group design 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 group design 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 group design 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. 19.

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

Intervention An educational program, product, practice, or policy aimed at improving student outcomes.

Intervention report A summary of the findings of the highest-quality research on a given program, product, practice, or policy in education. The WWC searches for all research studies on an intervention, reviews each against design standards, and summarizes the findings of those that meet WWC design standards.

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 study participants 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 eligible study participants are randomly assigned to 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. 19.

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.

Glossary of Terms

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% ($p < .05$).

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

Systematic review A review of existing literature on a topic that is identified and reviewed using explicit methods. A WWC systematic review has five steps: 1) developing a review protocol; 2) searching the literature; 3) reviewing studies, including screening studies for eligibility, reviewing the methodological quality of each study, and reporting on high quality studies and their findings; 4) combining findings within and across studies; and, 5) summarizing the review.

Please see the WWC Procedures and Standards Handbook (version 3.0) for additional details.

Intervention Report | Practice Guide | Quick Review | Single Study Review

An **intervention report** summarizes the findings of high-quality research on a given program, practice, or policy in education. The WWC searches for all research studies on an intervention, reviews each against evidence standards, and summarizes the findings of those that meet standards.

This intervention report was prepared for the WWC by Mathematica Policy Research under contract ED-IES-13-C-0010.