What Works Clearinghouse

INSTITUTE OF EDUCATION SCIENCES

High School Math February 2011

Saxon Math

Program Description[1]

Saxon Math is a textbook series covering grades K–12 based on incremental development and continual review of mathematical concepts to give students time to learn and practice concepts throughout the year. A distinguishing feature of *Saxon Math* is its use of a distributed approach—spreading practice and instruction for any single math content strand across the course of the entire instructional year—as opposed to a chapter-based approach for instruction and assessment. The program is built on the premise that students learn best when instruction is incremental and explicit, previously learned concepts are continually reviewed, and assessment is frequent and cumulative. At each grade level, math concepts are introduced, reviewed, and practiced over time in order to move students from understanding to fluency.

Research[2]

One study of *Saxon Math* that falls within the scope of the High School Math review protocol meets What Works Clearinghouse (WWC) evidence standards with reservations. The one study included 278 high school students in two districts in Colorado.[3]

Based on the one study, the WWC considers the extent of evidence for *Saxon Math* on high school students to be small for math achievement.

Effectiveness

Saxon Math was found to have no discernible effects on math achievement for high school students.

	Math achievement
Rating of effectiveness	No discernible effects
Improvement index[4]	Average: –15 percentile points Range: –17 to –14 percentile points

1. The descriptive information for this program was obtained from a publicly available source: the program's website (http://saxonpublishers.hmhco.com/en/sxnm_home.htm, downloaded October 2010). The WWC requests developers to review the program description sections for accuracy from their perspective. 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 October 2009.
2. The studies in this report were reviewed using WWC Evidence Standards, Version 2.0 (see the WWC Procedures and Standards Handbook, Chapter III), as described in protocol Version 2.0.
3. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.
4. These numbers show the average and range of student-level improvement indices for all findings across the study.

Additional program information

Developer and contact

Originally developed by John Saxon, *Saxon Math* is distributed by Saxon Publishers, an imprint of Houghton Mifflin Harcourt Specialized Curriculum. Address: Specialized Curriculum Group, 181 Ballardvale Street, Wilmington, MA 01887. Email: suppesker-orders@hmhpub.com. Web: http://saxonpublishers.hmhco.com/en/sxnm_home.htm. Telephone: (800) 289-4490.

Scope of use

The first textbook, *Saxon Algebra*, was published in 1979 by John Saxon for junior college students. In 1980, a high school version, *Saxon Algebra 1*, was published. By 1993, Saxon Publishers had developed programs for kindergarten through high school; Saxon joined Harcourt in 2004. School districts in all 50 states use Saxon products.

Teaching

The *Saxon Math* curriculum for each grade level or course consists of at least 120 daily lessons and 12 activity-based investigations. A daily lesson consists of learning a new mathematical concept, working on practice problems relating to that lesson, and solving a number of problems that include the current and previous material. This daily cycle is interrupted for tests and additional topics. Some versions of the curriculum include a teacher's edition with support and options for differentiated instruction.

Cost

Individual copies of the student and teacher editions of the *Saxon Algebra 1* textbook cost $70.95 and $103.75, respectively. Other available products include practice guides, manipulatives, and teaching materials, ranging from $8.75 for a student edition practice workbook to $534.40 for a manipulatives kit.

Research

Eight studies reviewed by the WWC High School Math topic area investigated the effects of *Saxon Math*. One study is a quasi-experimental design that meets WWC evidence standards with reservations. The remaining seven studies do not meet either WWC evidence standards or eligibility screens.

Meets evidence standards with reservations

Abrams (1989) conducted a quasi-experiment using treatment and comparison classrooms in two Colorado school districts during the 1988–89 school year. Within these districts, algebra teachers in four schools volunteered to participate in the study. Classrooms had already been formed (students assigned) at the beginning of the study, and subsequently, school officials determined which classrooms received the treatment and which were in the comparison group. The final sample included eight teachers and 278 students (126 intervention and 152 comparison).

Extent of evidence

The WWC categorizes the extent of evidence in each domain as small or medium to large (see the WWC Procedures and Standards Handbook, Appendix G). The extent of evidence takes into account the number of studies and the total sample size across the studies that meet WWC evidence standards with or without reservations.[5]

The WWC considers the extent of evidence for *Saxon Math* to be small for math achievement for high school students.

5. The extent of evidence categorization was developed to tell readers how much evidence was used to determine the intervention rating, focusing on the number and size of studies. Additional factors associated with a related concept—external validity, such as the students' demographics and the types of settings in which studies took place—are not taken into account for the categorization. Information about how the extent of evidence rating was determined for *Saxon Math* is in Appendix A6.

Effectiveness

Findings

The WWC review of interventions for High School Math addresses student outcomes in one domain: math achievement. The findings below present the authors' estimates and WWC-calculated estimates of the size and the statistical significance of the effects of *Saxon Math* on high school students.[6]

Math achievement. Abrams (1989) reported negative and statistically significant effects of *Saxon Math* on math achievement based on the Cooperative Mathematics Test, Algebra I subtest, and Mathematics Problem Solving Part II, Problem Solving Strategies subtest. After accounting for clustering and multiple comparisons, the WWC determined these findings were neither statistically significant nor large enough to be considered substantively important according to WWC criteria (i.e., an effect size of at least 0.25). As the findings for both measures within the domain were neither statistically significant nor substantively important, this study is characterized by the WWC as having an indeterminate effect on math achievement.

Rating of effectiveness

The WWC rates the effects of an intervention in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative. The rating of effectiveness takes into account four factors: the quality of the research design, the statistical significance of the findings, the size of the difference between participants in the intervention and the comparison conditions, and the consistency in findings across studies (see the WWC Procedures and Standards Handbook, Appendix E).

The WWC found *Saxon Math* to have no discernible effects on math achievement for high school students

Improvement index

The WWC computes an improvement index for each individual finding. In addition, within each outcome domain, the WWC computes an average improvement index for each study and an average improvement index across studies (see WWC Procedures and Standards Handbook, Appendix F). The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the percentile rank of the average student in the comparison condition. Unlike the rating of effectiveness, the improvement index is entirely based on the size of the effect, regardless of the statistical significance of the effect, the study design, or the analysis. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results for the intervention group.

The average improvement index for math achievement is –15 percentile points in the one study, with a range of –17 to –14 percentile points across findings.

Summary

The WWC High School Math topic area reviewed eight studies on *Saxon Math*. One of these studies meets WWC evidence standards with reservations; the remaining seven studies do not meet either WWC evidence standards or eligibility screens. Based on the one study, the WWC found no discernible effects on math achievement for high school students. The conclusions presented in this report may change as new research emerges.

6. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Abrams (1989), corrections for clustering and multiple comparisons were needed, so the significance levels may differ from those reported in the original study.

References

Meets WWC evidence standards with reservations

Abrams, B. J. (1989). *A comparison study of the Saxon Algebra I text* (Unpublished doctoral dissertation). University of Colorado, Boulder.

Studies that fall outside the High School Math review protocol or do not meet WWC evidence standards

Aquino, A., & Zoet, C. (1985, Spring). Reinforcement in algebra I: A study in the use of the Saxon Algebra I textbook. *Mathematics in Michigan*, 23–28. The study is ineligible for review because it does not occur within the time frame specified in the protocol.

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. This study fails to meet evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

Johnson, D. M., & Smith, B. (1987). An evaluation of Saxon's algebra text. *Journal of Educational Research, 81*(2), 97–102. The study is ineligible for review because it does not occur within the time frame specified in the protocol.

Mayers, K. S. (1995). *The effects 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. This study fails to meet evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

McBee, M. (1982). *Dolciani versus Saxon: A comparison of two algebra I textbooks with high school students.* Oklahoma City, OK: Oklahoma City Public Schools. The study is ineligible for review because it does not occur within the time frame specified in the protocol.

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). The University of Tulsa, OK. The study is ineligible for review because it does not occur within the time frame specified in the protocol.

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 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.

Appendix

Appendix A1 Study characteristics: Abrams, 1989

Characteristic	Description
Study citation	Abrams, B. J. (1989). *A comparison study of the Saxon Algebra I text* (Unpublished doctoral dissertation). University of Colorado, Boulder.
Participants	The researchers targeted schools in Colorado that were reported by the developer to have purchased the *Saxon* textbooks. Many of the contacted districts were no longer using the materials, but two districts that were agreed to participate in the study. Within these districts, nine teachers in four schools initially agreed to participate. Their 18 classrooms had already been formed when the study was started, so school officials determined the treatment status of each classroom. The initial sample of students was 468 (228 treatment and 240 comparison), although some classrooms were excluded from the analysis (due to switching texts or the teacher dropping out) and some students did not complete posttests. The resulting sample had seven teachers and 13 classes with 278 students (four teachers and 7 classes with 126 treatment group students and three teachers and 6 classes with 152 comparison group students), for which equivalence was demonstrated on pretests. Both groups were about 48% male. The schools served a mix of populations, from rural to urban and lower-middle-class to upper-middle-class families.
Setting	The study took place in two Colorado school districts. Within these districts, students of eight teachers in four schools comprised the analysis sample.
Intervention	Students in the treatment group were taught using *Saxon Algebra I* in the 1988–89 academic year. Teachers were free to supplement the texts with additional materials.
Comparison	Students in the comparison group used one of two traditional texts: Gobran's *Beginning Algebra* and Coxford and Payne's *Algebra 1*. Teachers were free to supplement the texts with additional materials.
Primary outcomes and measurement	The two study outcomes are Algebra (Cooperative Mathematics Tests: Algebra I) and Problem Solving (Mathematics Problem Solving Part II: Problem Solving Strategies). The tests were administered to all students one to two weeks before the end of the term in spring 1989, during the regular algebra class time. For a more detailed description of these outcome measures, see Appendix A2.
Staff/teacher training	All teachers in the study were regular classroom teachers and had used both the *Saxon* and traditional curricula in the past.

Appendix A2 Outcome measures for the math achievement domain

Outcome measure	Description
Cooperative Mathematics Test: Algebra I	A 40-item assessment of algebra knowledge from the Educational Testing Service, 1962 (as cited in Abrams, 1989). The test was broken into two subscales based on the categorization of each question: knowledge and skill or understanding and application. This characterization was done by the classroom teachers and two mathematics education graduate students.
Mathematics Problem Solving Part II: Problem Solving Strategies	A 38-item assessment of problem-solving skills from the Wisconsin Pupil Assessment Program, 1985 (as cited in Abrams, 1989).

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

			Author's findings from the study		WWC calculations			
			Mean outcome (standard deviation)[2]					
Outcome measure	Study sample	Sample size (classes/ students)	*Saxon Math* group	Comparison group	Mean difference[3] (*Saxon Math* – comparison)	Effect size[4]	Statistical significance[5] (at $\alpha = 0.05$)	Improvement index[6]
Abrams, 1989[7]								
Cooperative Mathematics Test: Algebra I	Algebra I students	13/278	21.10 (5.68)	23.30 (5.68)	–2.20	–0.36	ns	–14
Mathematics Problem Solving Part II: Problem Solving Strategies	Algebra I students	13/278	19.60 (5.01)	21.90 (5.01)	–2.30	–0.43	ns	–17
Domain average for math achievement[8]						**–0.39**	**na**	**–15**

ns = not statistically significant
na = not applicable

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the math achievement domain. Subtest findings from the same study are not included in these ratings but are reported in Appendix A4.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. In the case of Abrams (1989), the standard deviation was computed from the standard error of measurement for each instrument and assigned to both groups.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results for the intervention group.
7. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Abrams (1989), corrections for clustering and multiple comparisons were needed, so the significance levels may differ from those reported in the original study.
8. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

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

			Author's findings from the study		WWC calculations			
			Mean outcome (standard deviation)[2]					
Outcome measure	Study sample	Sample size (classes/ students)	*Saxon Math* group	Comparison group	Mean difference[3] (*Saxon Math* – comparison)	Effect size[4]	Statistical significance[5] (at α = 0.05)	Improvement index[6]
Abrams, 1989[7]								
Cooperative Mathematics Test: Algebra I, Knowledge and Skill	Algebra I students	13/278	14.00 (3.82)	15.80 (3.82)	–1.80	–0.44	ns	–17
Cooperative Mathematics Test: Algebra I, Understanding and Application	Algebra I students	15/278	7.10 (2.61)	7.50 (2.61)	–0.40	–0.15	ns	–6

ns = not statistically significant

1. This appendix presents subtest findings for measures that fall wwin the math achievement domain. Total test scores were used for rating purposes and are presented in Appendix A3.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. In the case of Abrams (1989), the standard deviation was computed from the standard error of measurement for each instrument and assigned to both groups.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Abrams (1989), corrections for clustering and multiple comparisons were needed, so the significance levels may differ from those reported in the original study.

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

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of math achievement, the WWC rated *Saxon Math* as having no discernible effects for high school students.

Rating received

No discernible effects: No affirmative evidence of effects.

- Criterion 1: No studies showing a statistically significant or substantively important effect, either *positive* or *negative*.

  Met. The sole study did not show a statistically significant or substantively important effect.

Other ratings considered

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

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

  Not met. No studies showed a statistically significant positive effect.

AND

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

  Met. The sole study did not show a statistically significant or substantively important negative effect.

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

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

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

AND

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

  Not met. The sole study showed an indeterminate effect.

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

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

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

OR

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

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

(continued)

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

- Criterion 1: One study showing a statistically significant or substantively important *negative* effect and no studies showing a statistically significant or substantively important positive effect.

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

OR

- Criterion 2: Two or more studies showing statistically significant or substantively important *negative* effects, at least one study showing a statistically significant or substantively important *positive* effect, and more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

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

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

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

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

OR

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

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

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

Appendix A6 Extent of evidence by domain

		Sample size		
Outcome domain	Number of studies	Schools	Students	Extent of evidence[1]
Math achievement	1	4	278	Small

1. A rating of “medium to large” requires at least two studies and two schools across studies in one domain and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is “small.” For more details on the extent of evidence categorization, see the WWC Procedures and Standards Handbook, Appendix G.