

Appendix

Appendix A1.1 Study characteristics: Williams, 1986 (randomized controlled trial)

Characteristic	Description
Study citation	Williams, D. D. (1986). <i>The incremental method of teaching algebra I</i> . Kansas City, MO: University of Missouri.
Participants	Forty-six ninth-grade students from one high school were randomly assigned to intervention or comparison using a computer program. ¹ The proportion of female students was 56% in the intervention group and 44% in the comparison group. The majority of participants came from middle-class homes.
Setting	The study took place in one high school in Missouri serving a rural-suburban community. ²
Intervention	The intervention group was taught Algebra I using the <i>Saxon Math</i> textbook. Each lesson lasted approximately 57 minutes, which consisted of 20 minutes of lecture and about 37 minutes of directed practice. The teacher held a Master's degree in education and had 15 years of teaching experience. The same teacher taught both the intervention and comparison groups. ³
Comparison	The comparison group used the <i>Algebra I</i> book by Dolciani. The comparison group lessons also lasted approximately 57 minutes, which consisted of 20 minutes of lecture and about 37 minutes of directed practice. The teacher used the same number of instructional aides and supplemental materials in the intervention and comparison groups. According to the study author, the only difference between conditions was in the textbooks.
Primary outcomes and measurement	The primary outcome measure was a teacher-developed mathematics test administered at the beginning of the school year and again in middle of April. (See Appendix A2 for more detailed descriptions of outcome measures.)
Teacher training	Teacher training was not reported in this study.

1. The study did not provide information on the nature of the computer program used (e.g., if this was a classroom scheduler) or if scheduling considerations were used at the time of assignment of students into groups.
2. This study was accepted for review because the focus of this topic review is on grades 6 through 9 regardless of setting (i.e., middle school, junior high school, or high school). For further details see the [Middle School Math Protocol](#).
3. There was no indication that the teacher might have been biased towards the *Saxon Math* curriculum or the comparison curriculum. Therefore, this design met WWC evidence screens.

Appendix A1.2 Study characteristics: Peters, 1992 (randomized controlled trial with randomization problems)

Characteristic	Description
Study citation	Peters, K. G. (1992). Skill performance comparability of two algebra programs on an eighth-grade population. <i>Dissertation Abstracts International</i> , 54(01), 77A. (UMI No. 9314428).
Participants	Participants of this study included 36 eighth-grade students from one school. All the students were “math-talented” based on teacher recommendations, prior academic achievement, and personal maturity. Students were randomly assigned to one of two classrooms (one intervention classroom and one comparison classroom). However, following assignment to groups, one student was moved to intervention; so 19 students were assigned to intervention, and 17 students were assigned to comparison.
Setting	The study took place in a junior high school in a rural suburban district abutting Lincoln, Nebraska.
Intervention	Participants in the intervention group were taught using the <i>Saxon Math</i> curriculum for eighth-grade students (<i>Algebra ½</i>). Students in this group participated in daily sessions for one year. In each session, the teacher introduced a new concept incrementally, and students had opportunities to practice the new concept and past concepts during each session. Students were assessed every fifth lesson. The intervention is designed to cover 120 lessons in one school year.
Comparison	Participants in the comparison group were taught using the University of Chicago School Mathematics Project textbook. This curriculum, based on NCTM standards, was designed to build independent learners and thinkers, build understanding of math vocabulary (such as mathematical signs), emphasize reviewing concepts within existing lessons, and increase student comprehension. The same teacher taught both the intervention and comparison groups.
Primary outcomes and measurement	The primary outcome measures were the Orleans-Hanna Algebra Prognosis Test and Understanding of Algebraic Components Test. (See Appendix A2 for more detailed descriptions of outcome measures.)
Teacher training	The teacher did not have prior experience with the intervention or comparison curriculum, but read extensively about both teaching formats. Agreed-upon components of both the intervention and comparison curricula were monitored on a weekly basis by the researcher to help maintain the integrity of implementation.

Appendix A1.3 Study characteristics: Crawford & Raia, 1986 (quasi-experimental design)

Characteristic	Description
Study citation	Crawford, J., & Raia, F. (1986). <i>Analyses of eighth grade math texts and achievement</i> . Oklahoma City, OK: Oklahoma City Public Schools, Planning, Research, and Evaluation Department.
Participants	The reviewed study ¹ included 72 students in the intervention group and 259 students in the comparison group. All students were in grade 8 and came from 17 classes taught by four teachers. The groups were matched on pretest California Achievement Test (CAT) scores. All students in the study were designated “high performing” students. No demographic information for the study sample was reported by the study authors.
Setting	The study took place in four middle schools in the Oklahoma City Public Schools. Four teachers taught both the intervention and the comparison groups.
Intervention	Participants in the intervention group were taught using the <i>Saxon Math</i> curriculum for eighth-grade students (<i>Algebra ½</i>) during the 1984–85 academic year. Specific information about the level of implementation was not provided. The intervention is designed to cover 120 lessons in one year with students participating in daily lessons, approximately 60 minutes a lesson.
Comparison	Participants in the comparison group were taught using the curricula already in place prior to the start of the study. At least three of the schools used the <i>Scott-Foresman Mathematics</i> curriculum in the comparison classrooms. Additional information about the curriculum used in comparison classrooms, including implementation, was not provided.

(continued)

Appendix A1.3 Study characteristics: Crawford & Raia, 1986 (quasi-experimental design) (continued)

Characteristic	Description
Primary outcomes and measurement	The primary outcome measure is the California Achievement Test (CAT). (See Appendix A2 for more detailed descriptions of outcome measures.)
Teacher training	Teacher training was not reported in this study.

1. Crawford & Raia (1986) reported on three types of analyses: those that had different schools assigned to each study condition, those used in the *Saxon Math* pilot schools only, and those that had teachers assigned to teach both the intervention and the comparison curriculum. This WWC review focuses only on the third type of analysis report, as it controls for school and teacher characteristics that might be associated with the choice of one curriculum over the other.

Appendix A1.4 Study characteristics: Resendez, Fahmy, & Manley, 2005 (quasi-experimental design)

Characteristic	Description
Study citation	Resendez, M., Fahmy, A., & Manley, M. A. (2005). <i>The relationship between using Saxon Middle School Math and student performance of Texas statewide assessments</i> . Retrieved from http://saxonpublishers.harcourtachieve.com/HA/correlations/pdf/s/SXMath_Middle_TX_research_web.pdf
Participants	This WWC review focused on the one-, two-, and three-year analyses of Sample 1 and the one-year analysis of Sample 3 using archival data. The study also included an analysis with part of the schools in Sample 1 (referred to as Sample 2 in the original report), but this analysis did not examine impact relative to comparison group, and therefore was not reviewed. For the examination of TAAS scores, more than 5,000 students in one cohort were tracked from grades 6 to 8. For the examination of TAKS scores, outcomes for more than 5,000 students were examined only at the end of sixth grade. The intervention and comparison schools were matched on demographic characteristics including ethnicity, poverty, English language proficiency, and percentage of mobile students.
Setting	Data was collected from 15 intervention schools and 15 comparison schools located in rural, suburban, and urban districts in Texas.
Intervention	The intervention group used the <i>Saxon 76</i> , <i>Saxon 87</i> , and <i>Saxon Algebra 1/2</i> textbooks. The school districts made decisions about which textbook series to use independent of the researcher.
Comparison	The 15 comparison schools were randomly selected from a pool of identified 40 matched comparison schools. The majority of the comparison schools used core basal math curricula, which typically consist of a chapter-based approach to math instruction.
Primary outcomes and measurement	Two measures were used in this study: the Texas Assessment of Academic Skills (TAAS) and the Texas Assessment of Knowledge and Skills (TAKS). (See Appendix A2 for more detailed descriptions of outcome measures.)
Teacher training	Teacher training was not reported in this study.

Appendix A1.5 Study characteristics: Resendez & Manley, 2005 (quasi-experimental design)

Characteristic	Description
Study citation	Resendez, M., & Manley, M. A. (2005). <i>The relationship between using Saxon elementary and middle school math and student performance on Georgia statewide assessments</i> . Orlando, FL: Harcourt Achieve.
Participants	This study investigated the impact of <i>Saxon Math</i> in elementary and middle schools using an existing database obtained from the state. This WWC intervention report focuses on the analyses that included cohorts of students followed up from grade 6 to grade 8. Participants came from a variety of ethnic and socio-economic backgrounds. Intervention and comparison schools were matched on demographic characteristics, including gender, ethnicity, poverty, and percentage of migrant students.
Setting	The data collection included 17 intervention schools and 15 comparison schools for grade 6 data, and 16 intervention schools and 12 comparison schools for grade 7 and 8 data. ¹ These schools were located in rural, suburban and urban districts across Georgia.
Intervention	The intervention group was taught using <i>Saxon 76</i> , <i>Saxon 87</i> , <i>Algebra ½</i> , or <i>Algebra 1</i> textbooks. Participating schools implemented <i>Saxon Math</i> in at least 75% of their classrooms.
Comparison	Comparison students used core basal math curricula. The study reported that across the elementary and middle school samples, most comparison schools (62%) used traditionally organized math texts, which take a topic-based approach to math instruction. A few schools (5%) used an investigative approach stressing making connections among mathematics topics. The remaining schools (33%) mixed basal, investigative, computer-based and non-textbook based math instruction. The study did not provide information on the specific number of middle schools using each type of comparison program.
Primary outcomes and measurement	The primary outcome measure was the Criterion Referenced Competency Test (CRCT). (See Appendix A2 for more detailed descriptions of outcome measures.)
Teacher training	Teacher training was not reported in this study.

1. Information on the number of schools was requested by the WWC and obtained from the first study author.

Appendix A1.6 Study characteristics: Roberts, 1994 (quasi-experimental design)

Characteristic	Description
Study citation	Roberts, F. H. (1994). The impact of the <i>Saxon Mathematics</i> program on group achievement test scores. <i>Dissertation Abstracts International</i> , 55(06), 1498A. (UMI No. 9430198).
Participants	The sample consisted of 185 eighth-grade students. About one-third of the students in the sample were African American and the remaining students were Caucasian. ¹
Setting	All students were from six schools in two rural counties in Mississippi. The two school districts were similarly economically deprived and contained little industry.
Intervention	The intervention used the <i>Saxon 76</i> , <i>Saxon 87</i> , and <i>Saxon Algebra ½</i> textbooks. Saxon texts are arranged incrementally rather than by topic.
Comparison	Comparison schools use the Houghton-Mifflin text <i>Mathematics</i> in the seventh grade and the Holt, Rinehart, & Winston text <i>Mathematics Unlimited</i> in the eighth grade.
Primary outcomes and measurement	The primary outcomes measure was the Stanford Achievement Test, Eighth Edition (SAT-8) total math subtest score. (See Appendix A2 for more detailed descriptions of outcome measures.)
Teacher training	Teacher training was not reported in this study.

1. In Roberts (1994), additional analyses were done after removing schools scoring at the extremes (one treatment, one comparison). No statistically significant differences were reported for this comparison.

Appendix A2 Outcome measures in the math achievement domain

Outcome measure	Description
End-of-course math test	The 50-item teacher-developed test includes 37 computation and 13 word problems asking students to solve one- and two-variable equations, inequalities, or systems using a variety of mathematical formats including whole numbers, fractions, decimals, exponents, square roots, set notation, and graphing (as cited in Williams, 1986).
Orleans-Hanna Prognostic Test	This nationally normed test consists of 60 multiple-choice items based on nine model lessons and five questionnaire items that require students to report their course grades and predict their final grade if they were to take algebra. 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 (as cited in Peters, 1992).
Understanding of algebraic components	Four unit criterion-referenced tests designed to examine understanding of 12 algebraic components. The four units focus on algebraic terms and expressions, linear equations, exponents and polynomials, and systems, parabolas, and quadratic equations. Across units, this measure includes a total of 120 items (as cited in Peters, 1992). A performance average of the percentage of skills mastered by each of the students was used.
The California Achievement Test (CAT)	The California Achievement Test is a standardized achievement test. The mathematics section includes subtests on mathematics computation and mathematics concepts and applications, which assess the ability to perform fundamental mathematics operations, apply mathematical concepts, and use a variety of problem-solving strategies (as cited in Crawford & Raia, 1986; edition of the test was not reported).
The Stanford Achievement Test, Eighth Edition (SAT-8)	The math part of this nationally normed standardized achievement test includes three subtests: concepts of numbers (34 items), mathematics computation (44 items), and mathematics application (40 items). These three subtests combine to form a total mathematics score. In grade six the intermediate level Form K was used. In grade eight the Advanced level 2 form was used (as cited in Roberts, 1994).
The Texas Assessment of Academic Skills (TAAS)—TLI score	The Texas Assessment of Academic Skills (TAAS) is the state standardized test that reports the Texas Learning Index (TLI) or math scaled score, which was used in the analysis. The TAAS looks at 13 different objectives a student could have mastered (as cited in Resendez, Fahmy, & Manley, 2005).
The Texas Assessment of Knowledge and Skills (TAKS)	The Texas Assessment of Knowledge and Skills (TAKS) was used. This 48-item test covers 6 objectives. The math subtest of the TAKS included 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).
Criterion Referenced Competency Test (CRCT)	This Georgia standardized state test measures content outlined in the state's core curriculum. The math subtest includes number sense and numeration; geometry and measurement; patterns, relationships, and algebraic reasoning; statistics and probability; computation and estimation; and problem solving (as cited in Resendez & Manley, 2005).

Appendix A3 Summary of study findings included in the rating for the math achievement domain¹

Outcome measure	Study sample	Sample size (schools/students)	Author's findings from the study			WWC calculations		
			Mean outcome (standard deviation ²)		Mean difference ³ (Saxon Math – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			Saxon Math group	Comparison group				
Williams, 1986 (randomized controlled trial)⁷								
End-of-course math test	Grade 9	1/46	24.1 (10.50)	18.1 (7.23)	6.00	0.65	Statistically significant	+24
Average⁸ for math achievement (Williams, 1986)						0.65	Statistically significant	+24
Peters, 1992 (randomized controlled trial with randomization problems)⁷								
Orleans-Hanna Prognostic Test	Grade 8 (math-talented)	1/36	95.63 (4.53)	95.06 (4.09)	0.57	0.13	ns	+5
Understanding of algebraic components	Grade 8 (math-talented)	1/36	16.09 (5.23)	17.44 (4.16)	-1.35	-0.28	ns	-11
Average⁸ for math achievement (Peters, 1992)						-0.08	ns	-3
Crawford & Raia, 1986 (quasi-experimental design)⁷								
The California Achievement Test (CAT)	Grade 8	4/78	55.56 (11.86)	50.72 (11.75)	4.84	0.41	Statistically significant	+16
Average⁸ for math achievement (Crawford & Raia, 1986)						0.41	Statistically significant	+16
Resendez, Fahmy, & Manley, 2005 (quasi-experimental design)⁷								
The Texas Assessment of Academic Skills (TAAS)—TLI score	Grade 8	30/3,054	83.95 (7.00)	82.98 (7.68)	0.97	0.13	Statistically significant	+5
The Texas Assessment of Knowledge and Skills (TAKS)	Grade 6	30/2,933	2,229.02 (225.89)	2,174.49 (205.10)	54.53	0.25	Statistically significant	+10
Average⁸ for math achievement (Resendez, Fahmy, & Manley, 2005)						0.19	Statistically significant	+8

(continued)

Appendix A3 Summary of study findings included in the rating for the math achievement domain *(continued)*

Outcome measure	Study sample	Sample size (schools/students)	Author's findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ³ (Saxon Math – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			Saxon Math group	Comparison group				
Resendez & Manley, 2005 (quasi-experimental design)⁷								
Criterion Referenced Competency Test (CRCT)	Grade 8	28/nr	66.63 (nr)	64.61 (nr)	2.02	na ⁹	ns	na ⁹
Average⁸ for math achievement (Resendez & Manley, 2005)						na ⁹	ns	na ⁹
Roberts, 1994¹⁰ (quasi-experimental design)⁷								
The Stanford Achievement Test, Eighth Edition (SAT-8)	Grade 8	6/185	49.97 (21.06)	52.68 (21.06)	-2.71	-0.13	ns	-5
Average⁸ for math achievement (Roberts, 1994)						-0.13	ns	-5
Domain average⁸ for math achievement across all studies						0.21	na	+8

na = not applicable

nr = not reported

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the improvement index. Subscale and subgroup findings from the same studies are not included in these ratings but are reported in Appendices A4.1 and A4.2.
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. Standard deviations for the analyses reported in Resendez, Fahmy, & Manley (2005) and Resendez & Manley (2005) were requested by the WWC and received from the first study author. The means in Roberts (1994) are gains between pretest and posttest in NCE mean scores. The standard deviations in this study were not available and therefore estimated based on the distribution of NCE scores, which has a mean of 50 and standard deviation of 21.06.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. In Williams (1986), the study groups were equivalent on pretest, so unadjusted posttest means were used for this review. In Roberts (1994), the intervention group mean is the comparison group mean plus the mean difference.
4. For an explanation of the effect size calculation, please see the [Technical Details of WWC-Conducted Computations](#).
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.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of Resendez, Fahmy, & Manley (2005), no correction for clustering was needed, as the study author provided the WWC with an aligned analysis that takes into account clustering and the results of that analysis demonstrated statistically significant differences between the groups; however, corrections for multiple comparisons were needed for this study. In the case of Crawford & Raia (1986) and Roberts (1994), corrections for clustering but not for multiple comparisons were needed. In the case of Williams (1986), Peters (1992), and Resendez & Manley (2005), no corrections for clustering or multiple comparisons were needed.
8. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.
9. Student-level standard deviations were not available for this study. School-level standard deviations were 7.99 for the intervention group and 7.96 for the comparison group. Because the student-level effect size and improvement index could not be computed, the magnitude of the effect size was not considered for rating purposes. However, the statistical significance for this study is comparable to other studies and is included in the intervention rating.
10. Roberts (1994) reported on sub-group analyses by gender and ethnicity. However, those analyses were not reviewed by the WWC because complete statistical information was not available. For further details please see [Technical Details of WWC-Conducted Computations](#).

Appendix A4.1 Summary of subtest findings for the math achievement domain¹

Outcome measure	Study sample	Sample size (schools/students)	Author's findings from the study			WWC calculations		
			Mean outcome (standard deviation ²)		Mean difference ³ (Saxon Math – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			Saxon Math group	Comparison group				
Crawford & Raia, 1986 (quasi-experimental design)⁷								
The California Achievement Test (CAT)								
Math computations	Grade 8	4/78	57.66 (13.35)	51.44 (14.14)	6.22	0.45	Statistically significant	+17
Math concepts	Grade 8	4/78	53.18 (12.44)	50.00 (12.40)	3.18	0.25	ns	+10
Resendez, Fahmy, & Manley, 2005 (quasi-experimental design)⁷								
The Texas Assessment of Knowledge and Skills (TAKS)								
Numbers, operations, and quantitative reasoning	Grade 6	30/2,933	7.21 (2.15)	6.53 (2.21)	0.68	0.31	Statistically significant	+12
Patterns, relationships, and algebraic reasoning	Grade 6	30/2,933	6.11 (2.18)	5.52 (2.16)	0.59	0.27	Statistically significant	+11
Geometry and spatial reasoning	Grade 6	30/2,933	5.36 (1.49)	5.01 (1.50)	0.35	0.23	Statistically significant	+9
Concepts and uses of measurement	Grade 6	30/2,933	3.09 (1.26)	2.99 (1.32)	0.10	0.08	Statistically significant	+3
Probability and statistics	Grade 6	30/2,933	4.54 (1.29)	4.45 (1.41)	0.09	0.07	Statistically significant	+3
Mathematical processes and tools	Grade 6	30/2,933	6.6 (1.74)	6.29 (1.89)	0.31	0.17	Statistically significant	+7
Resendez & Manley, 2005 (quasi-experimental design)⁷								
Criterion Referenced Competency Test (CRCT)								
Number sense and numeration	Grade 8	29/nr	72.22 (nr)	69.90 (nr)	2.32	na ⁸	ns	na ⁸
Problem solving	Grade 8	29/nr	68.3 (nr)	64.00 (nr)	4.30	na ⁸	ns	na ⁸

(continued)

Appendix A4.1 Summary of subtest findings for the math achievement domain *(continued)*

Outcome measure	Study sample	Sample size (schools/students)	Author's findings from the study			WWC calculations		
			Mean outcome (standard deviation ²)		Mean difference ³ (Saxon Math – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			Saxon Math group	Comparison group				
Resendez & Manley, 2005 (quasi-experimental design)⁷ <i>(continued)</i>								
Criterion Referenced Competency Test (CRCT) <i>(continued)</i>								
Geometry and measurement	Grade 8	29/nr	64.48 (nr)	62.84 (nr)	1.64	na ⁸	ns	na ⁸
Patterns, relationships, and algebraic reasoning	Grade 8	29/nr	64.06 (nr)	65.13 (nr)	-1.07	na ⁸	ns	na ⁸
Statistics and probability	Grade 8	29/nr	68.44 (nr)	63.53 (nr)	4.91	na ⁸	ns	na ⁸
Computation and estimation	Grade 8	29/nr	62.30 (nr)	62.27 (nr)	0.03	na ⁸	ns	na ⁸

na = not applicable

nr = not reported

ns = not statistically significant

1. This appendix presents subtest findings for measures that fall in the math achievement domain. Total scores and scale scores were used for rating purposes and are presented in Appendix A3.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. Standard deviations for the analyses reported in Resendez, Fahmy, & Manley (2005) and Resendez & Manley (2005) were requested by the WWC and received from the first study author.
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, please see the [Technical Details of WWC-Conducted Computations](#).
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.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of Crawford & Raia (1986), corrections for clustering were needed.
8. Student-level standard deviations were not available for this study. School-level standard deviations for the intervention and comparison groups were, respectively, 8.42 and 5.69 for number sense and numeration; 7.85 and 11.12 for problem solving; 9.51 and 11.57 for geometry and measurement; 7.39 and 11.58 for patterns, relationships, and algebraic reasoning; 8.83 and 7.54 for statistics and probability; and 9.58 and 8.30 for computation and estimation. Student-level effect sizes and improvement indices could not be computed for this study, but the statistical significance is comparable to other studies. For further details please see [Technical Details of WWC-Conducted Computations](#).

Appendix A4.2 Summary of subgroup findings for the math achievement domain¹

Outcome measure	Study sample	Sample size (schools/students)	Author's findings from the study			WWC calculations		
			Mean outcome (standard deviation ²)		Mean difference ³ (Saxon Math – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			Saxon Math group	Comparison group				
Resendez, Fahmy, & Manley, 2005 (quasi-experimental design)⁷								
The Texas Assessment of Academic Skills (TAAS)—TLI score	Grade 6	30/3,403	83.66 (7.72)	82.50 (9.42)	1.16	0.13	Statistically significant	+5
The Texas Assessment of Academic Skills (TAAS)—TLI score	Grade 7	30/3,054	83.78 (8.19)	82.27 (9.47)	1.52	0.17	Statistically significant	+7
Resendez & Manley, 2005 (quasi-experimental design)⁷								
CRCT total score	Grade 6	32/nr	73.35 (nr)	71.84 (nr)	1.51	na ⁸	ns	na ⁸
CRCT total score	Grade 7	33/nr	76.94 (nr)	76.57 (nr)	0.37	na ⁸	ns	na ⁸

na = not applicable

nr = not reported

ns = not statistically significant

1. This appendix presents subgroup findings for measures that fall in the math achievement domain. These findings were not used for rating purposes because the samples overlap with the most recent follow-ups 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. Standard deviations for the analyses reported in Resendez, Fahmy, & Manley (2005) and Resendez & Manley (2005) were requested by the WWC and received from the first study author.
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, please see the [Technical Details of WWC-Conducted Computations](#).
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.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of Resendez, Fahmy, & Manley (2005), no correction for clustering was needed.
8. Student-level standard deviations were not available for this study. School-level standard deviations for the intervention and comparison groups were, respectively, 6.70 and 9.29 for grade 6 and 4.64 and 5.54 for grade 7. Student-level effect sizes and improvement indices could not be computed for this study, but the statistical significance is comparable to other studies. For further details please see [Technical Details of WWC-Conducted Computations](#).

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

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

For the outcome domain of math achievement, the WWC rated *Saxon Middle School Math* as having positive effects. The remaining ratings (potentially positive effects, mixed effects, no discernible effects, potentially negative effects, negative effects) were not considered because *Saxon Middle School Math* was assigned the highest applicable rating.

Rating received

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.

Met. Of the six studies reviewed, three studies showed statistically significant positive effects; one of those studies met WWC evidence standards for a strong design. A fourth study showed substantively important positive effects. The remaining two studies showed indeterminate effects.

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

Met. No studies showed statistically significant or substantively important negative effects.

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