Appendix

Appendix A1.1 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 *Skill performance comparability of two algebra programs on an eighth-grade population.* Unpublished doctoral dissertation, University of Nebraska, Lincoln, NE.
Participants	The study included 36 eighth-grade students. All of the students were "math talented" based on teacher recommendations, prior academic achievement, and personal maturity. The students scored at the 87th percentile or above on the California Achievement Test total math battery.
Setting	The study took place in one junior high school in a rural suburban district abutting Lincoln, Nebraska. Students were randomly[1] assigned to one of two classrooms (one intervention classroom and one comparison). The same teacher taught both the intervention and comparison groups.
Intervention[2]	Participants in the intervention group were taught the *UCSMP Algebra* curriculum. No information was provided about the fidelity of implementation.
Comparison	Participants in the comparison group were taught using the *Saxon Middle School Math* curriculum for eighth-grade students (Algebra 1/2). Students in this group participated in daily sessions for one academic 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 *Saxon Math* curriculum is designed to cover 120 lessons in one year.
Primary outcomes and measurement	The primary outcome measures are the Orleans-Hanna Algebra Prognosis Test and the Understanding of Algebraic Components test.[3] (See Appendix A2 for more detailed descriptions of outcome measures.)
Teacher training	The study noted that the teacher who taught both study groups did not have prior experience with the intervention or comparison curricula, but read extensively about both teaching formats. The teacher participated in a one-week summer workshop on *UCSMP Algebra,* and in two additional one-day workshops given by local consultants on the curricula used in this study. Further, 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.

1. The study indicates that a random selection of numbers was used to divide the 36 participants between the intervention and comparison groups. But due to scheduling problems for other course offerings, the number of students in each group was changed to 17 in the *UCSMP Algebra* group and 19 in the *Saxon Math* group. The study meets standards with reservations because baseline differences were controlled for in the statistical analysis.
2. The same teacher taught both the intervention and comparison groups. Because there is no indication in the study to assume that the teacher was biased toward one of the conditions, this design was accepted for review.
3. The Orleans-Hanna prognosis test is typically administered as a measure of students' basic Algebra skills and does not cover the full scope of the knowledge and skills taught in a *UCSMP Algebra* course. It was accepted for review because it assesses important mathematics knowledge and skills relevant to middle school math.

Appendix A1.2 Study characteristics: Thompson, Senk, Witonsky, Usiskin, & Kaeley, 2006 (quasi-experimental design)

Characteristic	Description
Study citation	Thompson, D. R., Senk, S. L., Witonsky, D., Usiskin, Z., & Kaeley, G. (2006). *An evaluation of the second edition of UCSMP Algebra.* Chicago: University of Chicago School Mathematics Project.
Participants	The study included 189 students (98 in the intervention group and 91 in the comparison group) in six matched pairs of classrooms. In some of the participating schools, classrooms were randomly assigned to conditions.[1] About 85% of students were in ninth grade; the remaining students were enrolled in upper grade levels. A prerequisite for participation was a strong pre-algebra course or UCSMP *Transition Mathematics* in the preceding year. The sample consisted of a diverse student population, including ethnic minority students and students from low socioeconomic status. About half of the participants were female (53.1% in the intervention group, 50.5% in the comparison group).
Setting	The study participants attended three high schools in the West, Northeast, and South.[2] Two of the high schools were located in suburban districts, and the third school served inner-city and suburban populations.
Intervention	The intervention group used *UCSMP Algebra* 2nd edition, which was tested in this study and revised afterwards. Class time spent on instruction ranged 43–58 minutes, with a mean of 51.3 minutes and a standard deviation of 7.6 minutes. The intervention group teachers were expected to cover 12 chapters; one of the teachers was expected to cover only the first sections of those chapters. Based on teachers' reports, the study reported major differences in students' opportunity to learn different mathematic concepts. For example, *UCSMP Algebra* students studied application of the concepts in the algebra text, while comparison students in two schools had limited exposure to applications. Intervention group students had more access to calculators than comparison group students. Neither intervention nor comparison group students had access to computers.
Comparison	The comparison group used the mathematics textbooks that were in place in the schools prior to the study—*Algebra I: An Incremental Development* (Saxon), *Algebra: Structure and Method Book I* (Houghton Mifflin), and *Algebra 1* (Prentice Hall). Class time spent on instruction was similar to that reported for the intervention group. Comparison group teachers were expected to cover 10–11 of the chapters in the textbook; one teacher was expected to cover 12 of the 13 chapters.
Primary outcomes and measurement	This study used three measures–High School Subject Tests: Algebra, an algebra readiness test, and a problem-solving and understanding test.[3] Achievement on the general mathematics test and the algebra readiness test was analyzed in three ways: overall achievement using all test items; using only the test items for which all study participants in the same school had the opportunity to learn the content needed to answer the items, so the test was different in each school (referred to as "fair test"); and using only the test items for which all students in all participating schools had the opportunity to learn the content needed to answer the items, so the test was the same in all schools (referred to as "conservative test"). For the purposes of this WWC review, only the analyses focusing on overall achievement were used for rating purposes. (See Appendix A2 for more detailed descriptions of outcome measures.)
Teacher training	The teachers in the intervention group did receive training before or during the school year. These teachers received the textbook chapters, lesson notes, and answers to questions in stages throughout the school year. According to the study, none of the teachers had previously taught using *UCSMP Algebra* textbooks. No information was provided for training of teachers in the comparison group.

1. The study also reported on a comparison between *UCSMP Algebra* first edition and second edition and found no statistically significant differences. Because this intervention report reviews the evidence for both versions of the curriculum, this comparison was not reviewed for rating purposes. The study did not include a comparison of first edition with a non–*UCSMP Algebra* comparison group.
2. Although the study participants attended high schools, this sample is relevant to the scope of the Middle School Math review, which focuses on students in grades 6–9 regardless of the setting (for example, middle school, junior high school, or high school).
3. The algebra readiness test and the problem-solving and understanding test were developed by UCSMP.

Appendix A2 Outcome measures in the math achievement domain

Outcome measure	Description
Orleans-Hanna Algebra Prognosis 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 (as cited in Peters, 1992).
Understanding of Algebraic Components	Four unit 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.
High School Subject Tests: Algebra	This test consists of 40 items (20 arithmetic skills items and 20 arithmetic uses items). The test was developed by Scott, Foresman, and Company (as cited in Thompson et al., 2006).
Algebra Readiness	This test consists of 11 multiple-choice items and 10 short constructed-response items. The test, constructed by UCSMP, was accepted for review based on description of its content (as cited in Thompson et al., 2006).
Problem-solving and Understanding	This open-ended problem-solving test was developed by UCSMP (as cited in Thompson et al., 2006). This test was administered in two different forms—"even form" and "odd form"—and each form examines different mathematics skills and knowledge (for example, probability, mean and median, decimals and fractions, and area and perimeter). The test was accepted for review based on description of its content. Half of the students in each class were randomly assigned each form type.

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

Outcome measure	Study sample	Sample size (classrooms/ students)	Author's findings from the study: Mean outcome (standard deviation[2]) — *UCSMP Algebra* group	Author's findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*UCSMP Algebra* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Peters, 1992 (randomized controlled trial with randomization problems)[7]								
Orleans-Hanna Algebra Prognosis Test	Grade 8 (math talented)	2/36	95.06 (4.09)	95.63 (4.53)	–0.57	–0.13	ns	–5
Understanding of Algebraic Components[8]	Grade 8 (math talented)	2/36	17.44 (4.16)	16.09 (5.23)	1.35	0.28	ns	+11
Average[9] for math achievement (Peters, 1992)						0.08	ns	+3
Thompson et al., 2006 (quasi-experimental design)[7]								
High School Subject Tests: Algebra	Grades 9–12	12/189	47.90 (16.30)	46.00 (14.90)	1.90	0.12	ns	+5
Algebra Readiness	Grades 9–12	12/189	49.50 (16.30)	37.30 (14.90)	12.20	0.78	Statistically significant	+28
Problem-solving and Understanding	Grades 9–12	12/189	6.23 (3.69)[10]	3.39 (2.54)	2.84	0.89	Statistically significant	+31
Average[9] for math achievement (Thompson et al., 2006)						0.60	Statistically significant	+22
Domain average[9] for math achievement across all studies						0.34	na	+13

ns = not statistically significant
na = not applicable

1. This appendix reports findings considered for the effectiveness rating and the improvement index.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. 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 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 results favorable to the intervention group.
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 Peters (1992), no corrections for clustering or multiple comparisons were needed. In the case of Thompson et al. (2006), corrections for clustering and multiple comparisons were needed, so the significance levels may differ from those reported in the original study.
8. Means and standard deviations for this student outcome were calculated by the WWC based on raw data presented in the appendices of the original study report.
9. 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 size.
10. Means and standard deviations for the problem-solving and understanding test were calculated by the WWC based on the means and standard deviations presented for the even and odd forms of this test (see Appendix A4), taking into account the number of students who completed each type of form.

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

Outcome measure	Study sample	Sample size (schools/ students)	Author's findings from the study: Mean outcome (standard deviation[2]) *UCSMP Algebra* group	Author's findings from the study: Mean outcome (standard deviation[2]) Comparison group	WWC calculations: Mean difference[3] (*UCSMP Algebra* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Thompson et al., 2006 (quasi-experimental design)[7]								
Problem-solving and Understanding—odd form	Grades 9–12	12/101	5.20 (3.30)	2.70 (2.20)	2.50	0.88	Statistically significant	+31
Problem-solving and Understanding—even form	Grades 9–12	12/88	7.40 (3.80)	4.20 (2.70)	3.20	0.96	Statistically significant	+33

ns = not statistically significant
nr = not reported

1. This appendix presents findings for the even form and odd form versions of the problem-solving and understanding test reported in Thompson et al. (2006). Aggregated scores across forms were used for rating purposes and are presented in Appendix A3.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. 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 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 results favorable to the intervention group.
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 Thompson et al. (2005), a correction for clustering was needed.

Appendix A5 *UCSMP Algebra* rating for the math achievement domain

The WWC rates an intervention's effects 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 *UCSMP Algebra* as having potentially positive effects. It did not meet the criteria for positive effects because no studies met WWC evidence standards for a strong design. The other ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *UCSMP Algebra* was assigned the highest applicable rating.

Rating received

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.

  Met. One study showed a statistically significant positive effect.

- 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.

  Met. One study showed a statistically significant positive effect, and one study showed an indeterminate effect. No studies showed a statistically significant or substantively important negative 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 met WWC evidence standards for a strong design, and only one study showed statistically significant positive 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 effect. The WWC also considers the size of the domain level effect for ratings of potentially positive or potentially negative effects. See the WWC Intervention Rating Scheme for a complete description.