Appendix

Appendix A1 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. *Dissertation Abstracts International, 54*(01), 77A. (UMI No. 9314428)
Participants	The study included 36 students in grade 8 in two classrooms in one 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 in the sample scored at or above the 87th percentile on the California Achievement Test total math battery. The sample consisted of 20 girls (9 treatment, 11 comparison) and 16 boys (8 treatment, 8 comparison).
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. Students were randomly assigned to one of two classrooms (one intervention classroom and one comparison classroom).[1] The same teacher taught both the intervention and comparison groups.
Intervention[2]	Participants in the intervention group were taught the *UCSMP Algebra* curriculum during the 1991/92 school year. The intervention curriculum was monitored weekly by the researcher to help maintain 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 measure is the Orleans-Hanna Algebra Prognosis Test.[3] The pretest administration occurred in August 1991, and the posttest administration occurred in May 1992. For a more detailed description of this outcome measure, see Appendix A2.
Staff/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 had 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 author indicates that a random selection of numbers was used to divide participants between the intervention and comparison groups. However, the assignment of students was altered to accommodate scheduling difficulties and student requests for other course offerings. The analysis sample includes 17 students in the *UCSMP Algebra* group and 19 in the *Saxon Math* group. The study author demonstrated the baseline equivalence of the *UCSMP Algebra* and *Saxon Math* groups at pretest.
2. The same teacher taught both the intervention and comparison groups. Because both the intervention and comparison curricula were monitored on a weekly basis by the researcher to help maintain the integrity of implementation, and because there is no indication in the study to assume that the teacher was biased toward one of the conditions, this design was accepted for review.
3. The author described only the Orleans-Hanna Algebra Prognosis Test as the 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.

Appendix A2 Outcome measure for 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). For Peters (1992), pretest scores on the Orleans-Hanna Prognosis Test were from an August 1991 administration and posttest scores were from a May 1992 administration.

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

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *UCSMP Algebra* group	Authors' 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	95.02[8] (4.09)	95.63[9] (4.53)	–0.61	–0.14	ns	–6
Average for math achievement (Peters, 1992)[10]						**–0.14**	**ns**	**–6**
Domain average for math achievement across all studies[10]						**–0.14**	**ns**	**–6**

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the math achievement domain.
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 author 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. For the formulas the WWC used to calculate statistical significance, see Technical Details of WWC-Conducted Computations. In the case of Peters (1992), no correction for clustering or multiple comparisons was needed.
8. The intervention group value from Peters (1992) is the comparison group mean plus the difference in mean gains between the intervention (*UCSMP Algebra*) and control groups.
9. The comparison group mean from Peters (1992) is unadjusted.
10. 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.

Appendix A4 *UCSMP Algebra* 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 *UCSMP Algebra* as having no discernible effects.

Rating received

No discernible effects: No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative.*

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

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 statistically significant positive effects.

AND

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

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

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 study that evaluated math achievement and met WWC standards with reservations showed indeterminate effects.

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 effect, either positive or negative.

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 statistically significant or substantively important effects, either positive or negative.

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 Intervention Rating Scheme.

Appendix A5 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[1]
		Schools	Students	
Math achievement	1	1	36	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."