Appendix

Appendix A1 Study characteristics: Schoen & Hirsch, 2002

Characteristic	Description
Study citation	Schoen, H. L., & Hirsch, C. R. (2002). The *Core-Plus Mathematics* project: Perspectives and student achievement. In S. Senk & D. Thompson (Eds.), *Standards-based school mathematics curricula: What are they? What do students learn?* (pp. 311–343). Hillsdale, NJ: Lawrence Erlbaum Associates. ***Additional source:*** Schoen, H. L., Hirsch, C. R., & Ziebarth, S. W. (1998). *An emerging profile of the mathematical achievement of students in the* Core-Plus Mathematics *project*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.
Participants	Among an initial sample of 36 high schools that were field testing the *Core-Plus Mathematics* curriculum, 11 schools volunteered to administer pretests and posttests to students in both *Core-Plus Mathematics* and traditional classrooms. The authors state that schools were encouraged to create heterogeneous classroom groupings, although this was not always possible. The authors utilized a stratified matched-pairs design to select the intervention and comparison samples. Students in comparison classrooms were grouped by their most recently completed math course, and then matched to students in the intervention group using pretest scores, school, and gender, in that order. This process was conducted separately during each of the two years of the study (only five of the 11 schools from year one agreed to posttest students in the comparison group in year two). The main analysis included 1,050 students (525 intervention and 525 comparison) in year one and 390 students (195 intervention and 195 control) in year two. Additional analyses (reported in Appendices A3 and A4) varied in sample size, with baseline equivalence information presented separately for each of these samples.
Setting	The full set of 36 field-test schools were located in Alaska, California, Colorado, Georgia, Idaho, Iowa, Kentucky, Michigan, Ohio, South Carolina, and Texas. The 11 schools in the year one analysis included six from the Midwest (one urban, one rural, and four suburban), three from the West (one urban and two rural), one urban school from the East, and one rural school from the South. At each site, there were from two to five *Core-Plus Mathematics* teachers and from one to three comparison teachers. Five of the 11 schools continued into the year two analysis: two suburban, Midwestern schools and three urban schools, one from the South and two from the West.
Intervention	The intervention as implemented in the study included Course 1 and Course 2 of the *Core-Plus Mathematics* curriculum. The *Core-Plus Mathematics* Course 1 curriculum was used with ninth-grade students in year one, and *Core-Plus Mathematics* Course 2 was for tenth-grade students in year two. The authors note that the field-test versions of the *Core-Plus Mathematics* curriculum used in the study underwent revisions prior to the curriculum's formal publication.
Comparison	According to the authors, the nature of the instruction in the comparison classrooms was not specified in advance; a variety of traditional textbooks were used. Comparison classrooms during year one included 20 Algebra, five Pre-algebra, three General Mathematics, and two ninth-grade accelerated Geometry classes. Students in the year two comparison group were enrolled in either Algebra, Geometry, or Accelerated Advanced Algebra.
Primary outcomes and measurement[1]	Student math achievement was assessed using several measures. The full analysis sample for years one and two completed the Iowa Tests of Educational Development mathematics subtest. Slightly smaller numbers of students completed two author-created outcome measures: the Course 1 CPMP Posttest and Course 2 CPMP Posttest. The SAT Mathematics subtest also served as an outcome measure for a subsample of students. For a more detailed description of these outcome measures, see Appendix A2.
Staff/teacher training	From each school, a minimum of one *Core-Plus Mathematics* teacher attended a two-week workshop prior to teaching a *Core-Plus Mathematics* course. In this workshop, teachers worked through the course materials by using a small-group investigative approach similar to the one that they would be using with their own students. The comparison teachers had no special in-service program.

1. The study presented analyses on additional outcomes that are not included in this report. The samples for the ITED-Q two year trend analysis, NAEP, and college placement exam analyses were not equivalent at baseline; there was too much time between the establishment of equivalence (6th grade) and the start of the treatment (9th grade) for the ACT analysis; and the college performance outcomes are out of scope for this review.

Appendix A2 Outcome measures for the mathematics achievement domain

Outcome measure	Description
Iowa Tests of Educational Development Mathematics subtest (ITED-Q)	This nationally standardized achievement test is designed to measure skills on quantitative thinking processes that are important for anyone with at least a high school education. It is divided into three subtests (Understanding Mathematical Concepts and Procedures, Interpreting Information, and Solving Problems), and includes questions on whole numbers, exponents, fractions, decimals, percents, ratios, geometry, measurement, estimation, rounding, statistics, probability, tables, and graphs (as cited in Schoen & Hirsch, 2002).
Course 1 CPMP Posttest (Part 1)	This author-designed open-ended achievement test was designed to be a test of content that both *Core-Plus Mathematics* and comparison students would have had an opportunity to learn. It is divided into three subtests: two contextual subtests requiring algebraic methods, and a third subtest of procedural algebra. Specifically, the first two subtests require students to demonstrate their comprehension of algebraic concepts, such as linear equations, tables and graphs, and inequalities, by applying and interpreting them to specific examples, and the third subtest requires students to solve linear equations and simplify linear expressions (as cited in Schoen & Hirsch, 2002).
Course 2 CPMP Posttest (Part 1)	This author-designed open-ended achievement test was designed to be a test of content that both *Core-Plus Mathematics* and comparison students would have had an opportunity to learn. It is divided into three subtests: two contextual subtests (one algebraic and one geometric), and a third subtest of procedural algebra. The algebra subtests are similar in design to those in the Course 1 CPMP Posttest but include some work with exponents and quadratic expressions (as cited in Schoen & Hirsch, 2002).
SAT I Mathematics subtest (SAT)	One of the components of the SAT college entrance examination (SAT I), the Mathematics subtest, measures mathematical reasoning and symbol sense, drawing on content from arithmetic, algebra, and geometry (as cited in Schoen & Hirsch, 2002).

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

			Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation)[2]					
Outcome measure	Study sample	Sample size (students)	*Core-Plus Mathematics* group	Comparison group	Mean difference[3] (*Core-Plus Mathematics* – comparison)	Effect size[4]	Statistical significance[5] (at $\alpha = 0.05$)	Improvement index[6]
			Schoen & Hirsch (2002)[7]					
ITED-Q	Grade 9	1,050	266.0 (39.5)	257.1 (46.2)	8.9	0.21	Statistically significant	+8
ITED-Q	Grade 10	390	281.4 (32.0)	280.0 (32.3)	1.4	0.04	ns	+2
CPMP1 – Contextual Algebra I subtest	Grade 9	947	10.11 (4.20)	6.42 (4.22)	3.69	0.88	Statistically significant	+31
CPMP1 – Contextual Algebra II subtest	Grade 9	947	4.34 (2.64)	3.09 (2.26)	1.25	0.51	Statistically significant	+19
CPMP1 – Procedural Algebra subtest	Grade 9	947	8.92 (5.05)	10.87 (5.32)	–1.95	–0.38	Statistically significant	–15
CPMP2 – Contextual Algebra subtest	Grade 10	237	7.14 (3.97)	3.94 (2.74)	3.20	0.94	Statistically significant	+33
CPMP2 – Procedural Algebra subtest	Grade 10	237	7.54 (4.05)	8.30 (3.94)	–0.76	–0.19	ns	–8
CPMP2 – Coordinate Geometry subtest	Grade 10	237	16.10 (4.70)	11.13 (4.53)	4.97	1.07	Statistically significant	+36
SAT	Grades 11 and 12	98	484.6 (53.8)	467.0 (67.5)	17.6	0.29	ns	+11
Domain average for mathematics achievement[8]						**0.37**	**na**	**+15**

ns = not statistically significant
na = not applicable
CPMP1 = Course 1 CPMP Posttest (Part 1)
CPMP2 = Course 2 CPMP Posttest (Part 1)
ITED-Q = Iowa Tests of Educational Development Mathematics subtest
SAT = SAT I Mathematics subtest

(continued)

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the mathematics achievement domain. Subgroup 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.
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 Schoen and Hirsch (2002), 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 subgroup findings for the mathematics achievement domain[1]

			Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation)[2]					
Outcome measure	Study sample	Sample size (students)	*Core-Plus Mathematics* group	Comparison group	Mean difference[3] (*Core-Plus Mathematics* – comparison)	Effect size[4]	Statistical significance[5] (at $\alpha = 0.05$)	Improvement index[6]
			Schoen & Hirsch (2002)[7]					
ITED-Q	Grade 9 Pre-algebra	218	240.4 (35.5)	218.8 (40.2)	21.6	0.57	Statistically significant	+21
ITED-Q	Grade 9 Algebra	734	269.2 (37.9)	262.2 (42.0)	7.0	0.17	Statistically significant	+7
ITED-Q	Grade 9 Accelerated Geometry	98	299.1 (23.7)	304.0 (21.6)	–4.9	–0.21	ns	–8
ITED-Q	Grade 10 Algebra	62	252.0 (33.2)	248.6 (22.0)	3.4	0.12	ns	+5
ITED-Q	Grade 10 Geometry	278	283.7 (28.3)	281.4 (30.1)	2.3	0.08	ns	+3
ITED-Q	Grade 10 Accelerated Advanced Algebra	50	305.0 (23.8)	311.4 (17.5)	–6.4	–0.30	ns	–12
CPMP1 – Contextual Algebra I subtest	Grade 9 Algebra	655	10.60 (3.99)	6.98 (4.07)	3.62	0.90	Statistically significant	+32
CPMP1 – Contextual Algebra II subtest	Grade 9 Algebra	655	4.56 (2.49)	3.43 (2.25)	1.13	0.48	Statistically significant	+18
CPMP1 – Procedural Algebra subtest	Grade 9 Algebra	655	9.41 (5.04)	11.90 (4.74)	–2.49	–0.51	Statistically significant	–19
CPMP1 – Contextual Algebra I subtest	Grade 9 Accelerated Geometry	91	12.48 (3.33)	9.60 (4.14)	2.88	0.76	Statistically significant	+28
CPMP1 – Contextual Algebra II subtest	Grade 9 Accelerated Geometry	91	5.91 (3.05)	4.23 (2.29)	1.68	0.62	Statistically significant	+23

(continued)

Appendix A4 Summary of subgroup findings for the mathematics achievement domain[1] *(continued)*

			Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation)[2]					
Outcome measure	Study sample	Sample size (students)	*Core-Plus Mathematics* group	Comparison group	Mean difference[3] (*Core-Plus Mathematics* – comparison)	Effect size[4]	Statistical significance[5] (at $\alpha = 0.05$)	Improvement index[6]
CPMP1 – Procedural Algebra subtest	Grade 9 Accelerated Geometry	91	11.75 (4.80)	14.89 (3.77)	–3.14	–0.72	Statistically significant	–27
CPMP2 – Contextual Algebra subtest	Grade 10 Algebra	58	7.04 (4.05)	2.54 (2.05)	4.50	1.41	Statistically significant	+42
CPMP2 – Procedural Algebra subtest	Grade 10 Algebra	58	7.23 (4.41)	6.26 (2.86)	0.97	0.26	ns	+10
CPMP2 – Coordinate Geometry subtest	Grade 10 Algebra	58	13.15 (3.75)	8.26 (3.41)	4.89	1.35	Statistically significant	+41
CPMP2 – Contextual Algebra subtest	Grade 10 Geometry	136	6.70 (3.99)	3.99 (2.83)	2.71	0.78	Statistically significant	+28
CPMP2 – Procedural Algebra subtest	Grade 10 Geometry	136	7.23 (4.41)	7.51 (3.19)	–0.28	–0.07	ns	–3
CPMP2 – Coordinate Geometry subtest	Grade 10 Geometry	136	16.84 (4.66)	10.97 (4.19)	5.87	1.32	Statistically significant	+41

ns = not statistically significant
CPMP1 = Course 1 CPMP Posttest (Part 1)
CPMP2 = Course 2 CPMP Posttest (Part 1)
ITED-Q = Iowa Tests of Educational Development Mathematics subtest

1. This appendix presents subgroup findings for measures that fall in the mathematics achievement domain. Aggregated scores were used for rating purposes and are presented in Appendix A3. Results on the Course 1 CPMP Posttest for the Pre-algebra subgroup and Course 2 CPMP Posttest for the Accelerated Advanced Algebra subgroup are not included because the groups were not equivalent at baseline and the analyses did not control for pretest differences.
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 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 Schoen and Hirsch (2002), corrections for clustering and multiple comparisons were needed, so the significance levels may differ from those reported in the original study.

Appendix A5 *Core-Plus Mathematics* rating for the mathematics 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 mathematics achievement, the WWC rated *Core-Plus Mathematics* as having potentially positive effects for high school students. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered, as *Core-Plus Mathematics* 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. The sole study showed a statistically significant 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.

  Met. The sole study showed a statistically significant positive 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. Only one study showed a statistically significant positive effect.

AND

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

  Met. The sole study showed a statistically significant 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]
Mathematics achievement	1	11	1,050	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.