Appendix

Appendix A1 Study characteristics: Waite, 2000

Characteristic	Description
Study citation	Waite, R. D. (2000). A study of the effects of *Everyday Mathematics®* on student achievement of third-, fourth-, and fifth-grade students in a large north Texas urban school district. *Dissertation Abstracts International, 61*(10), 3933A. (UMI No. 9992659)
Participants	The participants were third-, fourth-, and fifth-grade students.[1] Six schools volunteered to implement the first edition of *Everyday Mathematics®*, and a comparison group of 12 schools in the same school district was selected and matched on previous mathematics scores, socioeconomic status, and ethnicity. The final sample consisted of 732 students in the intervention group and 2,704 students in the comparison group.
Setting	All the schools in this study were located in a large urban school district in north Texas.
Intervention	The intervention group consisted of six schools that volunteered to be part of a pilot program implementing the first edition of *Everyday Mathematics®*. The study examined implementation that occurred during the 1998–99 school year.
Comparison	Based on a profile of the intervention group, a comparison group of 12 schools in the same district that were similar in socioeconomic status, grade level, ethnic diversity, and previous year's Iowa Test of Basic Skills mathematics score were selected. The comparison group used a more traditional mathematics curriculum approved by the school district.
Primary outcomes and measurement	1999 Texas Assessment of Academic Skills (TAAS) mathematics scores. For a more detailed description of this outcome measure, see Appendix A2.
Staff/teacher training	Teachers in the intervention schools received 40 hours of training for the use of the *Everyday Mathematics®* curriculum and also received the "Teacher's Resource Package."

1. The distribution of grades across treatment and comparison students differed somewhat. However, the effect size would be similar if an effect size aggregating the separate effects for each grade was recalculated, weighting by the sample size in each grade.

Appendix A2 Outcome measure for the math achievement domain

Outcome measure	Description
1999 Texas Assessment of Academic Skills (TAAS)	As cited in Waite (2000), the 1999 TAAS was a criterion-referenced assessment, developed by the Texas Education Agency (TEA) from the state-mandated curriculum to assess higher order thinking and problem-solving skills across all public schools in Texas. TEA reports an internal consistency reliability range of 0.88 to 0.92 for the assessment. Only the mathematics scores from this assessment were used in this study.

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

Outcome measure	Study sample	Sample size (students/ schools)	Author's findings from the study: Mean outcome (standard deviation)[2]: *Everyday Mathematics®* group	Author's findings from the study: Mean outcome (standard deviation)[2]: Comparison group	WWC calculations: Mean difference[3] (*Everyday Mathematics®* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Waite, 2000[7]								
TAAS Mathematics test	Grades 3, 4, and 5	3,436/18	78.82 (11.5)	74.93 (14.8)	3.89	0.27	ns	+11
Domain average for math achievement[8]						**0.27**	**na**	**+11**

ns = not statistically significant
na = not applicable
TAAS = Texas Assessment of Academic Skills

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.
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. In the case of Waite (2000), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
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 (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. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C. In the case of Waite (2000), a correction for clustering was 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]

Outcome measure	Study sample	Sample size (students/ schools)	Author's findings from the study: Mean outcome (standard deviation)[2] — *Everyday Mathematics®* group	Author's findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[3] (*Everyday Mathematics®* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Waite, 2000[7]								
TAAS Math: Concepts	Grades 3, 4, and 5	3,436/18	17.51 (2.6)	16.75 (3.1)	0.76	0.25	ns	+10
TAAS Math: Operations	Grades 3, 4, and 5	3,436/18	13.08 (2.9)	12.20 (3.5)	0.88	0.26	ns	+10
TAAS Math: Problem solving	Grades 3, 4, and 5	3,436/18	9.73 (3.6)	8.63 (3.6)	1.10	0.31	ns	+12

ns = not statistically significant

TAAS = Texas Assessment of Academic Skills

1. This appendix presents subtest findings for measures that fall in the math achievement domain. Total 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. Pooled standard deviations across grades within the domain were calculated by the WWC and confirmed with the study authors.
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. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C. In the case of Waite (2000), corrections for clustering and multiple comparisons were needed, so the significance levels may differ from those reported in the original study.

Appendix A5 *Everyday Mathematics®* 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 *Everyday Mathematics®* as having potentially positive effects for elementary students. It could not achieve a rating of positive effects because there was only one study. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, or negative effects) were not considered, as *Everyday 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. *Everyday Mathematics®* had one study showing a 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.

  Met. *Everyday Mathematics®* had no studies showing 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. *Everyday Mathematics®* had no studies with a strong design.

AND

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

  Met. *Everyday Mathematics®* had no studies showing 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. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

Appendix A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[1]
		Schools	Students	
Math achievement	1	18	3,436	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.