

# Appendix

## Appendix A1 Study characteristics: Resendez & Manley, 2005 (randomized controlled trial)

Characteristic	Description
<b>Study citation</b>	Resendez, M., and Manley, M. A. (2005). <i>Final report: A study on the effectiveness of the 2004 Scott Foresman-Addison Wesley Elementary Math program</i> . Jackson, WY: PRES Associates, Inc.
<b>Participants</b>	The participants in this study were second- and fourth-grade students. Ten classes of 200 second graders and eight classes of 175 fourth graders in six schools were randomly assigned to the intervention condition using <i>Scott Foresman-Addison Wesley Elementary Mathematics</i> . In the same six schools, nine classes of 188 second graders and eight classes of 156 fourth graders were assigned to the comparison condition using five distinct elementary math programs.
<b>Setting</b>	The six elementary schools were located in urban, suburban, and rural communities in Washington (one urban school), Wyoming (one rural and one suburban school), Virginia (one urban school), and Kentucky (two suburban schools).
<b>Intervention</b>	Students used the 2004 <i>Scott Foresman-Addison Wesley</i> (SFAW) text during the school year. This was the first year this program had been implemented in these schools. Teachers covered 70% (SD=15.3%) of the curriculum.
<b>Comparison</b>	Students used five distinct comprehensive math programs that took a basal or investigative approach and covered the same content as <i>Scott Foresman-Addison Wesley Elementary Mathematics</i> . Teachers covered 75% (SD=18.2%) of their math program.
<b>Primary outcomes and measurement</b>	The primary outcome measure was the TerraNova CTBS, Basic Multiple Assessment (Level 12, 2nd grade and level 14, 4th grade) with Plus Test. As cited in Resendez and Manley (2005), the TerraNova CTBS is a reliable and valid standardized test consisting of multiple choice, constructed response, and computational problems. According to the authors, it offers broad coverage of the mathematics content in most textbooks and reflects the National Council of Teachers of Mathematics (NCTM) standards. The TerraNova CTBS provides two overall scores from two separate tests: the TerraNova Math Total and TerraNova Math Computation Total. (See Appendix A2 for more detailed descriptions of outcome measures.)
<b>Teacher training</b>	Teachers in the intervention classrooms met with a SFAW professional trainer for a half-day session prior to implementing the curriculum in their classes. Two follow-up sessions of approximately two hours were conducted: one session occurred 4–8 weeks after the teachers began implementation, and a second session occurred 10–18 weeks after implementation. The training did not focus on professional development in the form of effective teaching strategies but instead focused on the vision of the program and how teachers could use the SFAW math program to help students make sense of mathematics.

## Appendix A2 Outcome measures in the mathematics achievement domain

Outcome measure	Description
<b>TerraNova CTBS Math Total score</b>	The TerraNova CTBS Basic Multiple Assessment (Level 12, 2nd grade and Level 14, 4th grade) with Plus Test is a standardized test that, as described by Resendez and Manley (2005), was chosen for the study because of its validity, reliability, and sensitivity. The authors report that the test assesses content from the latest textbook series that are available from multiple publishers and reflects NCTM standards. Inter-rater reliability for the constructed response items was calculated by CTB McGraw Hill to range from 0.86 to 0.98 for all items on the second- and fourth-grade level tests (as cited in Resendez & Manley, 2005). The test provides two overall scores from two parts: The TerraNova Math Total and TerraNova Math Computation Total. The TerraNova Math Total (TNMT) score is based on multiple choice and constructed response for 43 Level 14 items and 34 Level 12 items. It is administered during two class sessions for a total of 75 minutes. The majority of items are word problems measuring basic, applied, and higher order thinking skills. TNMT Scores are reported as “scaled” and “normal curve equivalent” (as cited in Resendez & Manley, 2005).
<b>TerraNova Math Computation Total score</b>	The TerraNova Math Computation (TNMC) Total score is based on items in the CTBS- <i>Plus</i> test. It is a 20-item, multiple choice supplemental test that is administered for 20 minutes. The TNMC measures basic and advanced math computational skills (as cited in Resendez & Manley, 2005).

## Appendix A3 Summary of study findings included in the rating for the mathematics achievement domain<sup>1</sup>

Outcome measure	Study sample	Sample size (students/ teachers)	Author's findings from the study		Mean difference <sup>4</sup> (Scott Foresman–Addison Wesley – comparison)	WWC calculations		
			Mean outcome (standard deviation <sup>2</sup> )	Comparison group		Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
<b>Resendez &amp; Manley, 2005 (randomized controlled trial)</b>								
TerraNova Math Total score	Second- and fourth-grade students	645/35	55.59 (18.49)	54.14 (19.78)	1.45	0.08	ns	+3
TerraNova Math Computation score	Second- and fourth-grade students	533/35	53.89 (21.35)	57.49 (20.46)	-3.60	-0.17	ns	-7
<b>Domain average<sup>8</sup> for mathematics achievement</b>						-0.05	ns	-2

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices.
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. The authors used multi-level modeling to explore the differences in outcomes. The intervention group means are the control means plus the program coefficient from the HLM analyses.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, please see the [Technical Details of WWC-Conducted Computations](#).
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. 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](#). See the [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of *Scott Foresman–Addison Wesley Elementary Mathematics*, no corrections were needed.
7. 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.
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 Scott Foresman–Addison Wesley Elementary Mathematics rating for the math achievement domain

The WWC rates interventions as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.

For the outcome domain of math achievement, the WWC rated *Scott Foresman–Addison Wesley Elementary Mathematics* as having no discernible effects. It did not meet the criteria for positive effects because it had only one study, which had no statistically significant positive effects. Further, it did not meet the criteria for other ratings (potentially positive, mixed, potentially negative, or negative effects), because the study did not show statistically significant or substantively important effects, either positive or negative.

### 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.** The only study of *Scott Foresman–Addison Wesley Elementary Mathematics* showed an indeterminate 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 of *Scott Foresman–Addison Wesley Elementary Mathematics* examined math achievement outcomes. Although it met WWC standards for a strong design, it did not show statistically significant effects.

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

**Met.** The WWC analysis found no statistically significant or substantively important negative effects in this domain.

**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.** The WWC analysis found no statistically significant or substantively important positive effects in this domain.

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect. Fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

**Not met.** The only study of *Scott Foresman–Addison Wesley Elementary Mathematics* showed an indeterminate effect.

**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. 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.** The WWC analysis found no statistically significant or substantively important effects in this domain.

OR

(continued)

**Appendix A4**     **Scott Foresman–Addison Wesley Elementary Mathematics rating for the math achievement domain** *(continued)*

- 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.** The WWC analysis found no statistically significant or substantively important effects in this domain.

**Potentially negative effects:** Evidence of a negative effect with no overriding contrary evidence

- Criterion 1: At least one study showing a statistically significant or substantively important *negative* effect. No studies showing a statistically significant or substantively important *positive* effect. The number of studies showing statistically significant or substantively important *negative* effects is greater than the number showing statistically significant or substantively important *positive* effects.

**Not met.** The WWC analysis found no statistically significant or substantively important negative effects in this domain.

**Negative effects:** Strong evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *negative* effects, at least one of which is based on a strong design.

**Not met.** The only study of *Scott Foresman–Addison Wesley Elementary Mathematics* showed an indeterminate effect.

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

**Met.** The WWC analysis found no statistically significant or substantively important positive effects in this domain.