Appendix

Appendix A1.1 Study characteristics: Ysseldyke & Bolt, 2007

Characteristic	Description
Study citation	Ysseldyke, J., & Bolt, D. M. (2007). Effect of technology-enhanced continuous progress monitoring on math achievement. *School Psychology Review, 36*(3), 453–467.
Participants	The initial study sample included 3,309 students from 133 classrooms in nine schools, representing eight school districts in eight states. Sample students were in grades 2–8 during the 2003–04 school year. In the initial study sample, 38% of the students were Hispanic, 28% were African-American, 24% were Caucasian, 1% were Asian, 0% were Native American, and 8% were not specified. The baseline sample for grades 2–5 was 896 students (441 treatment and 455 control) in 40 classrooms (20 treatment and 20 control). The results here are drawn from the test-takers in the 40 elementary school classrooms included in the analysis—587 students took the STAR Math test (315 treatment, 272 control) and 700 took the Terra Nova test (355 treatment, 345 control). Postattrition treatment and control groups were equivalent on pretests at baseline.
Setting	The elementary school sample analyzed here comprises five schools in four states: Alabama, Florida, South Carolina, and Texas.
Intervention	Students were taught by teachers using the *Accelerated Math*™ program during the 2003–04 school year. Teachers assigned to the treatment group were asked to use *Accelerated Math*™ with their regular math curriculum. The existing curricula included: *Harcourt Math* or *Silver Burdett Math* in Alabama, *Houghton Mifflin Math Central* in Florida and South Carolina, and *Sharon Wells Math* or *Harcourt Math* in Texas. In practice, the program was not implemented for approximately 40% of grade 2–8 students in the initial treatment group; the authors did not report the percentage of grade 2–5 students in the treatment group of the analysis sample that did not participate in *Accelerated Math*™.
Comparison	Students in the control group were taught using the same set of math curricula as the treatment group, but without the addition of *Accelerated Math*™.
Primary outcomes and measurement	Participating students were pretested in October 2003 and posttested in May 2004 using two nationally normed, standardized tests (STAR Math and Terra Nova) for math achievement. Students in the treatment and control groups were compared using a linear regression analysis in which posttest scores were regressed on pretest scores. In the paper, the authors include school level dummies and interactions between the school and the treatment group in their analysis of outcomes for students in grades 2–8. The results presented in this report related to students in grades 2–5 (provided to the WWC by the author) include only controls for pretests. For a more detailed description of these outcome measures, see Appendix A2.
Staff/teacher training	Teachers in the intervention group were trained to use *Accelerated Math*™. During the school year, teachers using *Accelerated Math*™ received three to five visits from a Renaissance Learning math consultant, who guided teachers on how to improve their use of the program. Teachers also had unlimited access to technical support.

Appendix A1.2 Study characteristics: Nunnery & Ross, 2007

Characteristic	Description
Study citation	Nunnery, J., & Ross, S. M. (2007). The effects of the School Renaissance program on student achievement in reading and mathematics. *Research in the Schools, 14*(1), 40–59.
Participants	The analysis sample included 865 students (416 treatment, 449 comparison) from 18 elementary schools (nine treatment and nine comparison). These students were in grade 5 during the 2001–02 school year. Students in the analysis sample remained in the same school and had matched data available for three consecutive years (the 1999–2000 to 2001–02 school years). Characteristics of the student sample varied across the 18 schools. Between 0% and 59% in each school qualified for free or reduced-price lunch. Between 0% and 35% of students in each school were limited English proficient, and the proportion of students who were Caucasian ranged from 25% to 95%.
Setting	The treatment group schools came from one suburban school district in Texas. Comparison schools came from other school districts with similar populations of students in Texas.
Intervention	In the 2000–01 school year, schools in the treatment group began implementing School Renaissance, a comprehensive school reform model that includes *Accelerated Math*™. Students in the treatment group experienced two years of the *Accelerated Math*™ program as their primary mathematics curriculum. Treatment schools may have supplemented with other materials.
Comparison	Schools in the comparison condition were from Texas school districts that had not implemented the full School Renaissance package. It is possible that some elements of School Renaissance (e.g., *Accelerated Math*™) were present in the comparison schools; however, the comparison group curriculum is unknown.
Primary outcomes and measurement	The study used the Texas Learning Index math scores (based on the Texas Assessment of Academic Skills); for the grade 5 cohort, program comparisons were based on average transformed scores for grades 4 and 5 from 2001 and 2002. For a more detailed description of this outcome measure, see Appendix A2.
Staff/teacher training	A Renaissance coach conducted an initial training seminar and provided ongoing assistance to teachers.

Appendix A1.3 Study characteristics: Ysseldyke & Tardrew, 2007

Characteristic	Description
Study citation	Ysseldyke, J., & Tardrew, S. (2007). Use of a progress monitoring system to enable teachers to differentiate mathematics instruction. *Journal of Applied School Psychology, 24*(1), 1–28.
Participants	The initial study sample included 2,397 students (1,319 treatment and 1,078 comparison) in grades 3–10 during the 2001–02 school year. Students were drawn from 125 classrooms (67 treatment and 58 comparison) in 47 schools in 24 states. The elementary school analysis sample in this review included 1,680 students (869 treatment and 811 comparison) in grades 3–5. The grade 3 and grade 5 samples had large differences in baseline test scores that were not controlled for in the analysis. Therefore, only the grade 4 results for 614 students (303 treatment and 311 comparison) are reported.
Setting	The study was conducted in 47 schools in 24 states (Alabama, Arkansas, California, Georgia, Idaho, Illinois, Indiana, Iowa, Kansas, Massachusetts, Michigan, Minnesota, Missouri, Montana, New Mexico, Ohio, Oklahoma, Oregon, Pennsylvania, Tennessee, Texas, Virginia, Washington, and Wisconsin). The authors did not report whether all schools and states were represented in the elementary school sample.
Intervention	Students were taught by teachers using *Accelerated Math*™ during the spring semester of the 2001–02 school year. Teachers assigned to the *Accelerated Math*™ treatment group were asked to use the program with their existing math curriculum.
Comparison	Comparison classrooms were drawn from the same schools as treatment classrooms. Teachers assigned to the comparison group used their usual math curriculum and practices.
Primary outcomes and measurement	Students were pretested in January 2002 and posttested in May 2002 using STAR Math, a computer adaptive math achievement test. For a more detailed description of this outcome measure, see Appendix A2.
Staff/teacher training	Intervention teachers participated in a one-day training session conducted by Renaissance Learning. The training was designed to familiarize teachers with *Accelerated Math*™ and to guide them in integrating it into curriculum and instruction. Of 68 treatment group teachers in the full sample (grades 3–8), 66 attended the training.

Appendix A2 Outcome measures for the math achievement domain

Outcome measure	Description
Terra Nova mathematics subtest	The Terra Nova subtest is a national norm-referenced test that assesses academic performance in math.
Texas Learning Index math scores (based on the Texas Assessment of Academic Skills)	The Texas Assessment of Academic Skills (TAAS) is a criterion-referenced standardized state test that measures problem-solving and critical-thinking skills. The Texas Learning Index (TLI) is an outcome metric, based on student performance on the TAAS, which allows for comparisons between administrations and between grades. The TLI has a common interpretation across grades: a score of 70 or above indicates the student performed at or above grade-level expectations. A student receiving the same score at consecutive grade levels made one year of academic progress. Analyses in the study were based on a transformation of the TLI that was conducted to give the measure a desirable statistical property (a normal distribution).
STAR Math assessment	STAR Math is a computer-adaptive math test that assesses math skills. It combines computation and numeration items with word problems, estimation, statistics, charts and graphs, geometry, measurement, and algebra. STAR scores can appear as scaled scores or normal curve equivalent values. STAR Math was developed by Renaissance Learning.

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

			Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation)[2]					
Outcome measure	Study sample	Sample size (students)	*Accelerated Math*™ group	Comparison group	Mean difference[3] (*Accelerated Math*™ – comparison)	Effect size[4]	Statistical significance[5] (at $\alpha = 0.05$)	Improvement index[6]
Ysseldyke & Bolt, 2007[7]								
STAR Math normal curve equivalent scores	Grades 2–5	587	47.12[8] (21.40)	44.72[9] (24.05)	2.40	0.11	ns	+4
Terra Nova normal curve equivalent scores	Grades 2–5	700	46.59[8] (18.23)	45.43[9] (19.89)	1.16	0.06	ns	+2
Average for math achievement (Ysseldyke & Bolt, 2007)[10]						**0.08**	**ns**	**+3**
Nunnery & Ross, 2007[7]								
2001 and 2002 transformed Texas Learning Index scores	Grade 5 cohort	865	1.26[11] (0.28)	1.21 (0.30)	0.05	0.17	ns	+7
Average for math achievement (Nunnery & Ross, 2007)[10]						**0.17**	**ns**	**+7**
Ysseldyke & Tardrew, 2007[7]								
STAR Math scale scores	Grade 4	614	686.50[12] (85.74)	665.22[13] (85.46)	21.28	0.25	ns	+10
Average for math achievement (Ysseldyke & Tardrew, 2007)[10]						**0.25**	**ns**	**+10**
Domain average for math achievement across all studies[10]						**0.17**	**na**	**+7**

ns = not statistically significant
na = not applicable

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 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 an explanation, 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 for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Ysseldyke and Bolt (2007), corrections for clustering and multiple comparisons were needed, so the significance levels may differ from those reported in the original study. In the cases of Nunnery and Ross (2007) and Ysseldyke and Tardrew (2007), corrections for clustering were needed, so the significance levels may differ from those reported in the original studies.

(continued)

8. The intervention group means from Ysseldyke and Bolt (2007) differ from those presented in the paper. The WWC calculated the intervention group means by adding the impact of the program, derived from a regression that include a control for pretest differences, to the unadjusted comparison group means. The study authors provided the WWC with the impact, unadjusted comparison group mean, and standard deviations for both groups only for elementary schools.
9. The comparison group mean from Ysseldyke and Bolt (2007) 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.
11. Nunnery and Ross (2007, pp. 45–46) computed a transformation of the Texas Learning Index score to give the measure a normal distribution and allow for an analysis of covariance.
12. The intervention group mean from Ysseldyke and Tardrew (2007) equals the unadjusted comparison group mean plus the difference in gains between the treatment and comparison groups from pretest to posttest.
13. The comparison group mean from Ysseldyke and Tardrew (2007) is unadjusted.

Appendix A4 *Accelerated Math*™ 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 *Accelerated Math*™ as having mixed effects for elementary school students. The remaining ratings (no discernible effects, potentially negative effects, negative effects) were not considered, as *Accelerated Math*™ was assigned the highest applicable rating.

Rating received

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

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.

  Met. One study showed substantively important positive effects, and two studies showed indeterminate 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.

  Met. One study showed substantively important positive effects.

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. No studies showed negative effects, but more studies showed indeterminate effects than positive 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 A5 Extent of evidence by domain

Outcome domain	Number of studies	Sample size: Schools	Sample size: Students	Extent of evidence[1]
Math achievement	3	61[2]	2,179[3]	Medium to large

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.
2. For Ysseldyke and Tardrew (2007), the WWC was unable to determine the exact number of schools that were included in the analysis sample for the grades presented in this report.
3. For Ysseldyke and Bolt (2007), this presumes that the students who took the STAR Math test are a subset of those who took the Terra Nova. Otherwise, the number of students could be larger than 2,179.