

Appendix

Appendix A1 Study characteristics: Campuzano, Dynarski, Agodini, & Rall, 2009 (randomized controlled trial)

Characteristic	Description
Study citation	<p>Campuzano, L., Dynarski, M., Agodini, R., & Rall, K. (2009). <i>Effectiveness of reading and mathematics software products: Findings from two student cohorts</i> (NCEE 2009–4041). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.</p> <p>Additional source: Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., Means, B., Murphy, R., Penuel, W., Javitz, H., Emery, D., & Sussex, W. (2007). <i>Effectiveness of reading and mathematics software products: Findings from the first student cohort</i>. Washington, DC: U.S. Department of Education, Institute of Education Sciences.</p>
Participants	<p>The study sample included 1,037 sixth-grade students (547 <i>PLATO® Achieve Now</i>; 490 control) taught by 39 teachers (21 <i>PLATO® Achieve Now</i>; 18 control) in 13 schools across three districts in multiple states across the country during the 2004–05 and 2005–06 school years. Of the study sample, approximately 53% were female (52% <i>PLATO® Achieve Now</i> and 55% control), 74% received free or reduced-price lunch (not reported by intervention status), 42% were Hispanic (not reported by intervention status), and 40% were African-American (not reported by intervention status). Approximately 80% of the teachers in the study were female (81% <i>PLATO® Achieve Now</i>; 78% control) with an average of 11 years of teaching experience (9 years <i>PLATO® Achieve Now</i>; 13 years control) and 33% of whom obtained a master’s degree (24% <i>PLATO® Achieve Now</i>; 44% control).</p>
Setting	<p>The study took place in 13 schools in three districts in multiple states across the country.</p>
Intervention	<p>Students were taught using <i>PLATO® Achieve Now</i> during the 2004–05 and/or 2005–06 school years. <i>PLATO® Achieve Now</i> supplemented standard mathematics instruction for the treatment group. According to the study authors, <i>PLATO® Achieve Now</i> students used the product for independent practice and reinforcement of math skills. Students worked at their own pace on activities identified by the teacher. According to the authors, the recommended usage is 30 minutes per day, four days per week, for at least 10 weeks.</p>
Comparison	<p>Comparison students were taught in traditional classes, with the teachers utilizing any technology products already available to them.</p>
Primary outcomes and measurement	<p>The primary outcome measures in Year 2 of the study were the Stanford Achievement Test–Tenth Edition (SAT–10), the Iowa Test of Basic Skills (ITBS), and the New Mexico Standards Based Assessment (NMSBA). Districts in this study used at least one of these three exams as the outcome measure to obtain pretest and/or posttest scores. The study authors converted the scale scores from these tests to normal curve equivalent (NCE) scores with a range of 1 to 99 and an average of 50 to standardize the measures across tests and cohorts. For a more detailed description of these outcome measures, see Appendix A2.</p>
Staff/teacher training	<p>Vendor training sessions generally took place in host districts, and sometimes host schools, during the summer or early fall of 2004. The initial training lasted about 6 hours and varied by product from 4 hours to about 8 hours. Vendors delivered ongoing support in several modes. Product representatives visited teachers; vendors also provided support through email, telephone help desks, and additional training at schools.</p>

Appendix A2 Outcome measures for the math achievement domain

Outcome measure	Description
Stanford Achievement Test—Tenth Edition (SAT-10)	The SAT-10 is a commercially available standardized test that measures number sense and operations, patterns, relationships, algebra, geometry, measurement, data, statistics, and probability. The study authors converted the scale scores from the SAT-10 math test to normal curve equivalent scores to standardize the measures across tests and cohorts (as cited in Campuzano, Dynarski, Agodini, & Rall, 2009).
Iowa Test of Basic Skills (ITBS)	The ITBS is a group-administered, norm-referenced battery of achievement tests for students in kindergarten through eighth-grade. The tests are ordered by levels ranging from 5 to 14, which correspond to the targeted age of the students. The sixth-grade level of the math test measures number properties and operations, algebra, geometry, measurement, probability and statistics, and estimation. The study authors converted the scale scores from the ITBS math test to normal curve equivalent scores to standardize the measures across tests and cohorts (as cited in Campuzano, Dynarski, Agodini, & Rall, 2009).
New Mexico Standards Based Assessment (NMSBA)	The NMSBA is a criterion-referenced test. The math section measures number and operations, algebra, geometry, measurement, data analysis, and probability. Items were aligned to the New Mexico K-12 content standards, benchmarks, and performance standards. The study authors converted the scale scores from the NMSBA math test to normal curve equivalent scores to standardize the measures across tests and cohorts (as cited in Campuzano, Dynarski, Agodini, & Rall, 2009).

Appendix A3 Summary of study findings included in the rating for the math achievement domain¹

Outcome measure	Study sample	Authors' findings from the study						
		Sample size (teachers/students)	Mean outcome (standard deviation) ²		WWC calculations			
			PLATO [®] Achieve Now group ³	Comparison group ⁴	Mean difference ⁵ (PLATO [®] Achieve Now-comparison)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
Campuzano, Dynarski, Agodini, & Rall, 2009 (randomized controlled trial)⁹								
Math NCE ¹⁰	Grade 6	39/1,037	50.09 (18.44)	50.67 (19.38)	-0.58	-0.03	ns	-1
Domain average for math achievement¹¹						-0.03	na	-1

ns = not statistically significant

na = not applicable

NCE = normal curve equivalent

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. The intervention group value is the control group mean plus the program coefficient from the hierarchical linear model (HLM) analysis.
4. The control group mean is the unadjusted control group mean.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
6. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
7. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
8. 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.
9. 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 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 for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Campuzano, Dynarski, Agodini, & Rall (2009), no corrections for clustering or multiple comparisons were needed.
10. The study authors converted the scale scores from the SAT-10, ITBS, and NMSBA tests to normal curve equivalent units to standardize measures across tests and cohorts. NCE scores have a mean of 50, a standard deviation of 21.06, and a range of 1 to 99. A single overall NCE value is reported by the authors to reflect the average annual treatment effect across all districts.
11. 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 PLATO® Achieve Now rating for the math achievement domain

The WWC rates an intervention's effects in 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 PLATO® Achieve Now 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. The one study did not show a statistically significant or substantively important effect either positive or negative.

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. One study met WWC evidence standards for a strong design.

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.

Met. No studies showed a statistically significant or substantively important negative effect, no studies showed indeterminate effects, and no studies showed statistically significant or substantively important positive 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 study showed a statistically significant or substantively important effect, either positive or negative.

OR

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.

(continued)

Appendix A4 **PLATO® Achieve Now 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. No study showed a statistically significant or substantively important effect, either positive or negative.

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.

Not met. No studies showed a statistically significant or substantively important negative effect.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

Met. No studies showed a statistically significant or substantively important positive effect.

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 met WWC evidence standards for a *strong* design.

Not met. No studies showed a statistically significant or substantively important negative effect.

AND

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

Met. No studies showed a statistically significant or substantively important positive effect.

Appendix A5 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence ¹
		Schools	Students	
Math achievement	1	13	1,037	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.