Appendix

Appendix A1 Study Characteristics: DiLeo, 2007

Characteristic	Description
Study citation	DiLeo, J. (2007). *A study of a specific language arts and mathematics software program: Is there a correlation between usage levels and achievement?* Unpublished doctoral dissertation, Indiana University of Pennsylvania, Indiana, PA.
Participants[1]	In five out of ten schools within a single school district, fifth-grade teachers were randomly assigned to use CompassLearning's *Odyssey Language Arts* or *Odyssey Math* in their classrooms. Random assignment of the thirteen study teachers occurred after students were assigned to classrooms by their principals. The students in the *Odyssey Language Arts* classrooms served as the control group to the students using *Odyssey Math* (the intervention of interest for this WWC review). Two classrooms of students from one school were removed from the analysis sample because the school was a magnet school that had different demographic composition from the other schools. The final analysis sample, after excluding children with missing data, included 4 schools with 7 treatment classrooms (125 students) and 4 comparison classrooms (82 students). The author presented third-grade Pennsylvania System of School Assessment (PSSA) scores to establish baseline equivalence. Fourth-grade students did not take the PSSA in 2005, so third grade was the most recent year available. Post-attrition treatment and control groups had small differences on baseline test scores that were not statistically significant, but that did require the author control the baseline pretest (according to the Elementary Math protocol). Because the differential attrition rates between the treatment and comparison groups at the classroom level were greater than 5%, but the author provided the appropriate demonstration of equivalence and a control for the pretest, the WWC rated this study as meeting evidence standards with reservations. Approximately 7 percent of the students in the analysis sample were non-white, 63 percent of the students qualified for free or reduced-price lunch, and 14 percent of the students had an Individualized Education Program (IEP). An additional school was also presented in the author's analysis but is not included in this report because teachers in that school were not randomly assigned to use either the Language Arts or Math software and could have used both.
Setting	The study was conducted in one school district in central Pennsylvania.
Intervention	The intervention condition consisted of using *Odyssey Math* during the 2005–06 school year in addition to the usual mathematics curriculum, Houghton Mifflin. Teachers were asked to use the software a minimum of 90 minutes per week (the developer-recommended minimum), but usage levels varied across classrooms, at least in part because of access to technology. In two of the schools, students could only access the software during their weekly assigned time in the computer labs. In the other two schools, students had greater access to the software, as it was available during their weekly computer labs, in their classrooms via wireless laptops, and even at home. Students in the *Odyssey Math* condition could use the *Odyssey* software for any subject except *Language Arts*. Some of the students (from two of the schools) had access to the *Odyssey* software in the 4th grade—the year before the study began.
Comparison	The students in the control condition used only the district's Houghton Mifflin curriculum for mathematics. These students used the *Odyssey* software for Language Arts and possibly for other subjects (other than Mathematics). Some of the students (from two of the schools) had access to the *Odyssey* software in the 4th grade—the year before the study began.
Primary outcomes and measurement	Math achievement was measured using the Pennsylvania System of School Assessment (PSSA), which is the standardized assessment used for state accountability. For a more detailed description of this outcome measure, see Appendix A2.
Staff/teacher training	It is not clear how much training the teachers in this study received. However, the district had substantial funds to purchase professional development from CompassLearning for teachers in two of the schools during the 2004–05 school year through an Enhancing Education Through Technology (EETT) grant from the Pennsylvania Department of Education. During the 2005–06 school year, teachers who used *Odyssey* during the previous (2004–05) school year trained teachers using it for the first time in 2005–06.

1. Sample sizes and information about attrition and random assignment were provided to the WWC by the study author.

Appendix A2 Outcome measures for the math achievement domain

Outcome measure	Description
Pennsylvania System of School Assessment (PSSA)[1]	According to the Pennsylvania State Department of Education, the annual PSSA is a standards-based criterion-referenced assessment. The PSSA assesses students' abilities relative to specific standards within each subject and for each grade level. There are specific cut-off scores to determine a student's proficiency level. Students receive designations of below basic, basic, proficient, or advanced, depending on how they score in each tested subject. In addition to achievement categories, a continuous measure of student achievement also is presented in the study. The continuous score is presented in this intervention report.

1. A description of the Pennsylvania System of School Assessment was obtained at the Pennsylvania State Department of Education's website (http://www.pde.state.pa.us/a_and_t/site/default.asp) in December 2008.

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

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Odyssey Math* group	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[3] *(Odyssey Math–* comparison*)*	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
DiLeo, 2007 [7]								
Pennsylvania System of School Assessment	Grade 5	11/207	1583 (239)	1480 (215)	103	0.45	ns	+17
Domain average for math achievement (DiLeo, 2007)[8]						**0.45**	**ns**	**+17**

ns = not statistically significant

1. This appendix reports the 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. The author provided the WWC with standard deviations and means adjusted for pretest differences. Pretest adjusted means are presented for the *Odyssey Math* and comparison groups.
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 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 DiLeo (2007), 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 *Odyssey 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 *Odyssey Math* as potentially positive. It could not achieve a rating of positive because there was only one study. The remaining ratings (mixed, no discernible effects, potentially negative, or negative) were not considered, as *Odyssey Math* 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. *Odyssey Math* had one study that showed 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. *Odyssey Math* 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. *Odyssey Math* had no studies with a strong design.

AND

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

  Met. *Odyssey Math* 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 A5 Extent of evidence

Outcome domain	Number of studies	Sample size		Extent of evidence[1]
		Schools	Students	
Math achievement	1	4	207	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.