WWC Intervention Report

A summary of findings from a systematic review of the evidence

INSTITUTE OF EDUCATION SCIENCES

Primary Mathematics January 2017*

Odyssey® Math

Program Description[1]

Odyssey® Math is a web-based program developed by Compass Learning® for mathematics instruction in grades K–8. The online program includes a mathematics curriculum and formative assessments designed to support differentiated and data-driven instruction. Based on assessment results, the program generates an individualized sequence of mathematics topics and skills—a "learning path." *Odyssey® Math* is often used as a prescriptive tool, where students can start by taking a diagnostic assessment aligned with local or state standards. Teachers can modify learning paths to match their lesson plans or to align them with district scopes and sequences.

Research[2]

The What Works Clearinghouse (WWC) identified three studies of *Odyssey® Math* that both fall within the scope of the Primary Mathematics topic area and meet WWC group design standards. Two studies meet WWC group design standards without reservations, and one study meets WWC group design standards with reservations. Together, these studies included 2,768 students in grades 4–8 in 41 schools.

The WWC considers the extent of evidence for *Odyssey® Math* on the mathematics achievement of primary students to be medium to large for the mathematics achievement domain, the only domain examined for studies reviewed under the Primary Mathematics topic area.[3] (See the Effectiveness Summary on p. 4 for more details of effectiveness for the mathematics achievement domain.)

Effectiveness

Odyssey® Math was found to have potentially positive effects on mathematics achievement for primary students.

Report Contents

This intervention report presents findings from a systematic review of *Odyssey® Math* conducted using the WWC Procedures and Standards Handbook, (version 3.0), and the Primary Mathematics review protocol, (version 3.1).

Table 1. Summary of findings[4]

Outcome domain	Rating of effectiveness	Improvement index (percentile points) Average	Improvement index (percentile points) Range	Number of studies	Number of students	Extent of evidence
Mathematics achievement	Potentially positive effects	+12	+1 to +18	3	2,768	Medium to large

Program Information

Background

Odyssey® Math was developed and is distributed by Compass Learning®. Address: 203 Colorado Street, Austin, TX, 78701. Email: support@Compass Learning.com. Web: www.compasslearning.com/odyssey/ Telephone: (800) 232-9556.

Program details

Odyssey® Math is a web-based program designed to support mathematics instruction. The online program includes a mathematics curriculum and formative assessments that support differentiated and individualized instruction. *Odyssey® Math* can be used as a standalone curriculum or as a supplement to other mathematics curricula, and can be used via desktop, laptop, or mobile device. The curriculum focuses on fundamental math skills in earlier math grades, while in later grades, the curriculum prepares students for middle and high school mathematics. The interactive activities used by the program are designed to allow students to apply ideas, tools, and manipulatives, and build upon previous knowledge.

Odyssey® Math typically begins with students taking a diagnostic assessment. The program emphasizes flexibility in creating appropriate instructional pathways for students based on assessment results. Specifically, the program uses assessment results to create a sequence of skills—"learning paths"—that focus on the knowledge and skills an individual student needs to practice. The program allows teachers to manually modify the learning paths generated by the software. For example, teachers can add or remove tasks to support further practice or align the learning path with lesson plans. In addition to mathematics lessons and activities, the program includes assessment tools that generate data teachers can use for instructional decision making. Teachers can use the program's test bank, build their own items, or import other commercially-available assessments, such as the Northwest Evaluation Association's™ (NWEA™) Measures of Academic Progress® (MAP®), the Scantron Performance Series®, and the Renaissance Standardized Test for the Assessment of Reading® (STAR®).

As of June 2016, *Odyssey® Math* is used by districts and schools across the United States; however, the program is no longer available for purchase. According to the publisher, features of *Odyssey® Math* were incorporated into *Compass Learning Pathblazer® Math* and *Hybridge® Math*, programs currently sold by Compass Learning®. Those programs are different than *Odyssey® Math* and therefore are not within the scope of this WWC review.

Cost

As of June 2016, the program is no longer available for purchase. Cost information for ongoing support is available from the publisher.

Research Summary

The WWC identified nine eligible studies that investigated the effects of *Odyssey® Math* on the mathematics achievement of primary students. An additional 13 studies were identified but do not meet WWC eligibility criteria for review in this topic area. Citations for all 22 studies are in the References section, which begins on p. 5.

Table 2. Scope of reviewed research

Grades	4, 5, 7, 8
Delivery method	Whole class, small group
Program type	Supplement

The WWC reviewed nine eligible studies against group design standards. Two studies are cluster randomized controlled trials that meet WWC group design standards without reservations, and one study uses a quasi-experimental design that meets WWC group design standards with reservations. Those three studies are summarized in this report. The remaining six studies do not meet WWC group design standards.

Summary of studies meeting WWC group design standards without reservations

DiLeo (2007) conducted a cluster randomized controlled trial in a Pennsylvania school district between the 2004–05 and 2005–06 school years. Within study schools, fifth-grade teachers were randomly assigned to use either *Odyssey® Language Arts* or *Odyssey® Math* in their classroom. The *Odyssey® Language Arts* group served as the comparison group to the *Odyssey® Math* intervention group. Teachers assigned to the *Odyssey® Language Arts* group used their core mathematics curriculum (*Houghton Mifflin Mathematics*). Teachers assigned to the *Odyssey® Math* group were asked to use *Odyssey® Math* at least 90 minutes per week as a supplement to the *Houghton Mifflin Mathematics* core curriculum. The original randomly assigned sample included 13 classrooms (seven intervention and six comparison) in five schools. The analytic sample included 11 classrooms (seven intervention and four comparison) in four schools with a total of 207 fifth-grade students (125 intervention and 82 comparison).[5] The study was determined by the WWC to have low attrition; therefore, the study meets WWC group design standards without reservations.

Wijekumar, Hitchcock, Turner, Lei, & Peck (2009) conducted a cluster randomized controlled trial in which 122 fourth-grade teachers in 124 classrooms across 32 schools in Delaware, New Jersey, and Pennsylvania were randomly assigned to either use *Odyssey® Math* or continue using the standard math curriculum.[6] The 60 teachers randomly assigned to *Odyssey® Math* used the program as a partial replacement for the core math curriculum for 60 minutes per week. The 62 teachers assigned to the comparison group continued to use their school's fourth-grade math curriculum. The study investigated the effect of *Odyssey® Math* on student achievement by comparing average classroom scores on the TerraNova Basic Battery math test between *Odyssey® Math* and comparison classrooms during the 2007–08 school year. All outcomes were examined as teacher-level means, and there was no attrition of teachers. In total, pretest and posttest scores across 122 teachers, who together taught 2,456 students, were used in the analysis. The study is a cluster randomized controlled trial with low attrition; therefore, the study meets WWC group design standards without reservations.

Summary of study meeting WWC group design standards with reservations

Cornelius (2013) conducted a quasi-experimental evaluation of the effect of *Odyssey® Math* on the math achievement of 105 seventh- and eighth-grade students in five rural Tennessee schools during the 2010–11 school year. The author identified students who were considered low math performers based on their pretest math scores. These students were assigned to receive an additional 30 minutes of supplemental support as part of one of two groups: (a) a group that received supplemental teacher-directed instruction based on their school's core math curriculum, and (b) a group that used *Odyssey® Math* as a supplemental program to their core math curriculum. The author compared the math achievement of 60 students who used *Odyssey® Math* to 45 students who received teacher-directed instruction. These two groups were similar prior to the study and, therefore, the study meets WWC group design standards with reservations.[7]

Effectiveness Summary

The WWC review of *Odyssey® Math* for the Primary Mathematics topic area includes outcomes in one outcome domain: mathematics achievement. The findings below present the authors' estimates and WWC-calculated estimates of the size and statistical significance of the effects of *Odyssey® Math* on primary students. For a more detailed description of the rating of effectiveness and extent of evidence criteria, see the WWC Rating Criteria on p. 16.

Summary of effectiveness for the mathematics achievement domain

Table 3. Rating of effectiveness and extent of evidence for the mathematics achievement domain

Rating of effectiveness	Criteria met
Potentially positive effects *Evidence of a positive effect with no overriding contrary evidence.*	In the three studies that reported findings, the estimated impact of the intervention on outcomes in the *mathematics achievement* domain was positive and statistically significant in one study, positive and substantively important in one study, and indeterminate in the other study.
Extent of evidence	**Criteria met**
Medium to large	Three studies that included 2,768 students in 41 schools reported evidence of effectiveness in the *mathematics achievement* domain.

Three studies of *Odyssey® Math* that meet WWC group design standards with or without reservations reported findings in the mathematics achievement domain.

DiLeo (2007) reported a positive difference between the *Odyssey® Math* and comparison groups on the mathematics achievement domain. The author did not report the statistical significance of this finding, but the WWC-computed calculations (adjusted for classroom-level clustering) indicated that the results were not statistically significant. The effect size is large enough to be considered substantively important according to WWC criteria (i.e., an effect size of at least 0.25). The WWC characterizes this study finding as a substantively important positive effect.

Wijekumar et al. (2009) reported, and the WWC confirmed, no statistically significant difference between the *Odyssey® Math* and comparison groups on the mathematics achievement domain. The effect size was not large enough to be considered substantively important. The WWC characterizes this study finding as an indeterminate effect.

Cornelius (2013) reported, and the WWC confirmed, a positive and statistically significant difference between the *Odyssey® Math* and comparison groups on the mathematics achievement domain.[8] The WWC characterizes this study finding as a statistically significant positive effect.

Thus, for the mathematics achievement domain, one study showed a statistically significant positive effect, one study showed a substantively important positive effect, and one study showed an indeterminate effect. This results in a rating of potentially positive effects, with a medium to large extent of evidence.

References

Studies that meet WWC group design standards without reservations

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.

Wijekumar, K., Hitchcock, J., Turner, H., Lei, P., & Peck, K. (2009). *A multisite cluster randomized trial of the effects of Compass Learning Odyssey® Math on the math achievement of selected grade 4 students in the mid-Atlantic region. Final report* (NCEE 2009-4068). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://eric.ed.gov/?id=ED507314

Study that meets WWC group design standards with reservations

Cornelius, A. S. (2013). *The effects of response to intervention on the mathematics achievement of seventh and eighth grade students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI 1422044724)

Studies that do not meet WWC group design standards

Adams, T. (2012). *Investigation of the effectiveness of the Compass Learning math program on the mathematics success of urban students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI 1198356258) The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention.

Clariana, R. (2005). *Compass Learning classrooms of the future: An interim report of one-to-one wireless laptop use in upper elementary mathematics: Let them solo*. Austin, TX: Compass Learning, Inc. The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention.

Compass Learning, Inc. (2005a). *Compass Learning Odyssey school effectiveness report: Boone County School Districts (Research report brief)*. Austin, TX: Author. The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention.

Compass Learning, Inc. (2005c). *Odyssey school effectiveness report: Maple Leaf Intermediate School*. Austin, TX: Author. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Compass Learning, Inc. (2007a). *Impact of Compass Learning Odyssey Reading/Language Arts & Mathematics on NWEA RIT scores and Lexile range: Akron Elementary School*. Austin, TX: Author. The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention.

Compass Learning, Inc. (2007b). *Odyssey school effectiveness report: Washoe County School District*. Austin, TX: Author. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Studies that are ineligible for review using the Primary Mathematics Evidence Review Protocol

Brandt, W. C., & Hutchison, C. (2005). *Romulus Community Schools 2002–2005: Summary research and evaluation report*. Austin, TX: Compass Learning, Inc. This study is ineligible for review because it does not use an eligible design.

Additional source:

Brandt, W. C., Hutchison, C., & Learning Point Associates. (2006). *Romulus Community Schools comprehensive school reform evaluation*. Naperville, IL: Learning Point Associates.

Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K–12 classrooms: A meta-analysis. *Educational Research Review, 9*, 88–113. http://files.eric.ed.gov/fulltext/ED527573.pdf. This study is ineligible for review because it does not use an eligible design.

Clariana, R. (2007). *Odyssey school effectiveness report: Pemberton Township School District*. Austin, TX: Compass Learning, Inc. This study is ineligible for review because it does not use an eligible design.

Compass Learning, Inc. (2003). *Manager I and Manager II*. Austin, TX: Author. The study is ineligible for review because it is out of scope of the protocol.

Compass Learning, Inc. (2004). *School effectiveness report: Riverside Middle School, Pendleton, SC*. Austin, TX: Author. This study is ineligible for review because it does not use an eligible design.

Compass Learning, Inc. (2005b). *Compass Learning school effectiveness report: Daniel Boone Area School District*. Austin, TX: Author. This study is ineligible for review because it is out of scope of the protocol.

Compass Learning, Inc. (2006). *Compass Learning Odyssey school effectiveness report: Lillie Burney Elementary School*. Austin, TX: Author. This study is ineligible for review because it does not use an eligible design.

Compass Learning, Inc. (2007c). *Summer school pilot program for third-grade intervention: Tulsa Independent School District*. Austin, TX: Author. This study is ineligible for review because it does not use an eligible design.

Compass Learning, Inc. (2008). *Careful planning & commitment yield outstanding results for Carroll County School*. Austin, TX: Author. This study is ineligible for review because it does not use an eligible design.

Fischer, U., Moeller, K., Cress, U., & Nuerk, H. (2013). Interventions supporting children's mathematics school success: A meta-analytic review. *European Psychologist, 18*(2), 89–113. This study is ineligible for review because it does not use an eligible design.

Lawson-Martin, R. (2005). *Effects of cooperative and individual integrated learning system on attitudes and achievement in mathematics* (Unpublished doctoral dissertation). Florida International University, Miami. Retrieved from http://digitalcommons.fiu.edu/ This study is ineligible for review because it does not use an eligible design.

Milhollin, C. C. (2005). *Integration of an integrated learning system: Does it make a difference? An observational research study* (Unpublished doctoral dissertation). Valdosta State University, GA. This study is ineligible for review because it does not use an eligible design.

Wright, W. (2006). Catching up in math? The case of newly-arrived Cambodian students in a Texas Intermediate School. *Texas Association for Bilingual Education Journal, 9*(1), 1–22. This study is ineligible for review because it does not use an eligible design.

Appendix A.1: Research details for DiLeo (2007)

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.

Table A1. Summary of findings

Meets WWC group standards without reservations

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
Mathematics achievement	11 classrooms/207 students	+17	No

Setting

The study was conducted in 11 fifth-grade classrooms across four schools in a single school district in south central Pennsylvania. The study occurred in the 2005–06 school year.

Study sample

At the beginning of the study, 13 fifth-grade teachers in five schools in a single school district were randomly assigned to use either *Odyssey® Math* or *Odyssey® Language Arts* in their classrooms. As confirmed through an author query, random assignment of the teachers occurred after students were assigned to classrooms by their principals. For this review, students in the *Odyssey® Math* classrooms served as the intervention group, while students in the *Odyssey® Language Arts* classrooms served as the comparison group. Two classrooms were excluded from the analysis because the school was a magnet school with a demographic composition that differed from the other schools. After excluding students with missing data, the analytic sample included four schools, seven intervention classrooms (with 125 students), and four comparison classrooms (with 82 students).[9] Approximately 7% of the students in the analytic sample were non-White, 63% of the students qualified for free or reduced-price lunch, and 14% of students had an Individualized Education Plan.

Intervention group

Students in the intervention group used *Odyssey® Math* during the 2005–06 school year as a supplement to their core mathematics curriculum, *Houghton Mifflin Mathematics*. Teachers were asked to use the software a minimum of 90 minutes per week, but usage levels varied across classrooms, 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 at home. At two of the schools, some students may have had access to the Odyssey software as fourth graders—the year before the study began. At least one fourth grade class in each of these two schools had the Odyssey software, but it is unclear how many students from those fourth grade classes are included in the intervention group.

Comparison group

Students in the comparison group used the district's core mathematics curriculum, *Houghton Mifflin Mathematics*. The comparison students used the *Odyssey® Language Arts* software and may have used *Odyssey®* software in other subjects, but they were not permitted to use *Odyssey® Math*. At two of the schools, some students may have had access to the Odyssey software as fourth graders—the year before the study began. At least one fourth grade class in each of these two schools had the Odyssey software, but it is unclear how many students from those fourth grade classes are included in the comparison group.

Outcomes and measurement

The primary outcome measure was the math scale score on the Pennsylvania System of School Assessment (PSSA). According to the Pennsylvania State Department of Education, the PSSA is a standards-based criterion-referenced assessment administered annually. The math score of the PSSA assesses students' mathematics abilities relative to specific standards for each grade level. The math scale scores from the third-grade PSSA assessment were used as the baseline assessment, and math scale scores from the fifth-grade PSSA were used as the posttest.

The study included two assessments that are not eligible for this review: scores on the 4Sight math and reading assessments, both of which were administered to students as mid-year benchmark tests. The 4sight reading assessments are not an eligible outcome in the primary mathematics topic area, and the 4Sight math assessments were not used in an eligible analysis; therefore, it is not included in this review. For a more detailed description of the eligible outcome measure, see Appendix B.

Support for implementation

The report does not specify how much training intervention teachers received. The district purchased teacher professional development from Compass Learning® in two of the study schools during the year prior to the study (2004–05) through an Enhancing Education Through Technology grant from the Pennsylvania Department of Education. During the study year (2005–06), teachers who used *Odyssey®* during the previous school year (2004–05) trained their colleague teachers who were using the software for the first time (in 2005–06).

Appendix A.2: Research details for Wijekumar et al. (2009)

Wijekumar, K., Hitchcock, J., Turner, H., Lei, P., & Peck, K. (2009). *A multisite cluster randomized trial of the effects of Compass Learning Odyssey® Math on the math achievement of selected grade 4 students in the mid-Atlantic region. Final report* (NCEE 2009-4068). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Table A2. Summary of findings — **Meets WWC group standards without reservations**

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
Mathematics achievement	122 teachers/2,456 students	+ 1	No

Setting

The study was conducted in 32 elementary, intermediate, and charter schools in Delaware, New Jersey, and Pennsylvania in the 2007–08 school year. To be eligible for the study, schools had to have at least two fourth-grade classrooms, not track students based on academic performance, use the same math curricula within fourth-grade classrooms, have more than one fourth-grade teacher, have adequate technology to implement the intervention, and have no evidence of use of *Odyssey® Math* within the past 2 years.

Study sample

The analytic sample included 32 schools. All fourth-grade teachers within each school participated in the study; the analytic sample comprised 60 teachers in the intervention group and 62 teachers in the comparison group. Data from 1,223 intervention students and 1,233 comparison students contributed to the classroom-level means used in the analytic sample. The baseline student sample included 1,403 intervention students and 1,451 comparison students. Approximately 7% of the baseline student sample were English learners, about 25% belonged to a racial or ethnic minority group, and about 18% of the students were eligible for free or reduced-price lunch.

Intervention group

Students in the intervention group used *Odyssey® Math* for 60 minutes per week between October 2007 and April 2008 as a partial replacement of their schools' core math curriculum. Teachers used *Odyssey® Math's* online assessments and tools to improve students' math skills.

Comparison group

Students in the comparison group continued to use their school's core math curriculum. The core math curricula used by these schools included *Everyday Math®*, *Scott Foresman, Harcourt Brace,* and *Saxon Math*. Participating teachers were encouraged to spend a similar amount of daily and weekly instruction time on mathematics. Five students in the comparison group switched groups mid-year and received the *Odyssey® Math* curriculum. These students were analyzed as part of the comparison group, their originally assigned condition, thereby maintaining the integrity of the randomization.

Outcomes and measurement

The primary outcome measure was student scale scores on the math subtest of the TerraNova Basic Battery (grade 4, level 14). The assessment was administered as a pretest between September and October 2007 and as a posttest between April and May 2008. The subtest scores were aggregated and analyzed by classroom.

The study also examined the software usage patterns of intervention group classrooms and the data collected from classroom observations of both intervention and comparison group teachers. These data were used to assess implementation fidelity and are not considered eligible outcomes in the mathematics achievement domain. For a more detailed description of the eligible outcome measure, see Appendix B.

Support for implementation

Two professional development training workshops taught by Compass Learning® were offered to intervention teachers and to interested school administrators. The first workshop was offered in four regional locations in August 2007 and provided 5–6 hours of training. In total, 37 of the 60 teachers attended the August workshop, and 23 intervention group teachers attended a makeup day. The second workshop was offered on a single day in January 2008, and all 60 intervention group teachers attended. Both of these workshops provided training on the *Odyssey® Math* software. The workshops were followed by one-on-one coaching sessions with the intervention teachers in their classrooms and involved 1–2 hours of coaching per teacher. These workshops and coaching sessions were offered in addition to the standard professional development opportunities available to teachers.

Appendix A.3: Research details for Cornelius (2013)

Cornelius, A. S. (2013). *The effects of response to intervention on the mathematics achievement of seventh and eighth grade students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database.

Table A3. Summary of findings — **Meets WWC group design standards with reservations**

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
Mathematics achievement	5 schools/105 students	+18	Yes

Setting

The study was conducted in seventh- and eighth-grade classrooms in five schools located in one rural west Tennessee school district. The study occurred in the 2010–11 school year.

Study sample

The study sample consisted of seventh- and eighth-grade students with low performance based on their scores on the STAR® Math Assessment, with students scoring at or below the 10th percentile (based on the national norming sample) identified as "at risk". These students were assigned to receive either teacher-directed math instruction or *Odyssey® Math* as a supplemental program to their core math curriculum. The analysis compared the STAR® scores of 60 students who were taught using *Odyssey® Math* to 45 students who were taught using teacher-directed instruction. Among the 105 students in the analytic sample, 66% were African American and 28% were Caucasian.

Intervention group

Students in the intervention group received 30 minutes of supplemental instruction using *Odyssey® Math* computer-assisted instruction in addition to their core math curriculum.[10] The core mathematics curricula differed across the study schools and included *Accelerated Math*™, *Glencoe*, and *SRA Corrective Math*.

Comparison group

Students in the comparison group received 30 minutes of supplemental teacher-directed instruction in addition to their core math curriculum. The core math curricula differed across the study schools and included *Accelerated Math*™, *Glencoe*, and *SRA Corrective Math*. The teacher-directed instruction was based on the core math curricula used in the schools.

Outcomes and measurement

The primary outcome measure was student scale scores on the STAR® Math Assessment. In the school district, the STAR® Math Assessment was used as a benchmark to monitor student progress in math. The STAR® Assessment was administered at 8-week intervals in the fall, winter, and spring of the school year; however, the author used only the scores for the fall 2010 and spring 2011 tests.

The study included one assessment that is not eligible for this review: math scores on the Tennessee Comprehensive Assessment Program (TCAP). The TCAP was not used in an eligible analysis; therefore, it is not included in this review. For a more detailed description of the eligible outcome measure, see Appendix B.

Support for implementation

The report did not describe teacher training or implementation support.

Appendix B: Outcome measures for the mathematics achievement domain

Mathematics achievement	
Pennsylvania System of School Assessment (PSSA)	Scale scores on the PSSA provide a continuous measure of student achievement (as cited in DiLeo, 2007). The 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. Specific cut-off scores at each grade determine a student's proficiency level.
STAR® Math Assessment	STAR® Math is a commercially available, computer-adaptive mathematics testing program created by *Renaissance Learning*™. The adaptive component of the STAR® assessment uses student's responses on previous questions to determine the level of difficulty of subsequent questions. The STAR® assessments were used by the school district as benchmark measures during the school year and were administered in the fall, winter, and spring. The fall and spring scores were used in this report. The scale scores of the STAR® assessments can be used to monitor student progress and as a placement tool (as cited in Cornelius, 2013).
TerraNova Math Assessment (Form A)	The fourth-grade math subtest of the TerraNova Basic Battery consists of 57 selected-response questions and takes approximately 70 minutes to administer. Form A of the Basic Battery was administered at pretest and posttest, and the scale scores were used in this report. The norms of the TerraNova Basic Battery were updated in 2005 and are representative of the general population of K–12 students, including students with disabilities and English language learners. The math subtest has an internal consistency (Kuder-Richardson coefficient) of .93 and a Cronbach's alpha of .91 (as cited in Wijekumar et al., 2009).

Appendix C: Findings included in the rating for the mathematics achievement domain

			Mean (standard deviation)		WWC calculations			
Outcome measure	**Study sample**	**Sample size**	**Intervention group**	**Comparison group**	**Mean difference**	**Effect size**	**Improvement index**	***p*-value**
DiLeo (2007)[a]								
Pennsylvania System of School Assessment (PSSA)	Grade 5	11 classes/ 207 students	1,583.00 (239.15)	1,480.00 (215.04)	103.00	0.45	+17	nr
Domain average for mathematics achievement (DiLeo, 2007)						**0.45**	**+17**	**Not statistically significant**
Wijekumar et al. (2009)[b]								
TerraNova Math Assessment (Form A)	Grade 4	122 teachers/ 2,456 students	648.29 (38.69)	647.50 (38.18)	0.78	0.02	+1	.54
Domain average for mathematics achievement (Wijekumar et al., 2009)						**0.02**	**+1**	**Not statistically significant**
Cornelius (2013)[c]								
STAR® Math Assessment	Grades 7–8	5 schools/ 105 students	673.72 (68.70)	640.22 (75.20)	33.50	0.46	+18	< .05
Domain average for mathematics achievement (Cornelius, 2013)						**0.46**	**+18**	**Statistically significant**
Domain average for mathematics achievement across all studies						**0.31**	**+12**	**na**

Table Notes: For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. The WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of each study's domain average was determined by the WWC. Some statistics may not sum as expected due to rounding. nr = not reported, na = not applicable.

[a] For DiLeo (2007), the means and standard deviations used to calculate the effect size were obtained from an author query. The means reported in the table are adjusted for the pretest based on an ANCOVA that was provided by the authors. The author did not report the statistical significance of this result; therefore, the WWC computed the statistical significance using a correction for clustering, and resulted in a WWC-computed *p*-value of .15 for the PSSA. The WWC does not find the result to be statistically significant. The study is characterized as having a substantively important positive effect because the estimated effect is positive and larger than 0.25 standard deviations.

[b] For Wijekumar et al. (2009), the WWC did not need to make corrections for clustering or multiple comparisons, or to adjust for baseline differences. The *p*-value presented here was reported in the original study. Although the study is a cluster randomized controlled trial, the effect sizes are reported using a student level standard deviation. These student-level effect sizes were calculated using the level-2 coefficient from a hierarchical linear model divided by the pooled student-level posttest standard deviation. The standard deviation reported for the intervention group is the pooled student-level posttest standard deviation. This study is characterized as having an indeterminate effect because the estimated effect is neither statistically significant nor substantively important.

[c] For Cornelius (2013), pre- and posttest standard deviations were obtained through an author query. The WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The effect size is based on an ANCOVA analysis that was reported in the original study, which included the STAR® Math assessment pretest as a covariate. The *p*-value range presented here was reported in the original study. Please see the WWC Procedures and Standards Handbook (version 3.0) for more information. This study is characterized as having a statistically significant positive effect because the estimated effect is positive and statistically significant. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

Endnotes

* Due to the 2015 restructuring of the Mathematics topic area from three areas (Elementary, Middle, and High School) to two areas (Primary and Secondary Mathematics), this report is considered a new report, rather than an updated report. The information in this report includes the research examined in two prior reports, which are discussed in Endnote 2.

[1] The descriptive information for this program was obtained from a publicly available source: the publisher's website (www.compasslearning.com, downloaded September 2015). The WWC requests publishers review the program description sections for accuracy from their perspective. The program description was provided to the publisher in October 2015, and the WWC incorporated feedback from the publisher. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review.

[2] The WWC previously released reports on *Odyssey® Math* under the Middle School Mathematics topic area in June 2009 and the Elementary School Mathematics topic area in August 2009. The prior reports were prepared using the WWC Procedures and Standards Handbook (version 1.0), along with the Middle School Mathematics review protocol (version 1.0) and the Elementary School Mathematics review protocol (version 1.0), respectively. In June 2015, the WWC restructured the reviews of research on math interventions into two areas instead of three. These two review areas are Primary Mathematics (which includes interventions in which math is presented through multi-topic materials and curricula, typically used in grades K–8), and Secondary Mathematics (which includes interventions that are organized by math content area [e.g., Algebra, Geometry, and Calculus], typically taught in grades 9–12). These two areas replaced the prior Elementary School Math, Middle School Math, and High School Math areas, which were organized by student grade level. The WWC is updating and replacing intervention reports written under the prior topic areas.

The literature search for the current report reflects documents publicly available by January 2016. This report has been updated to include reviews of six studies that were not included in the prior reports. Of the additional studies, three were not within the scope of the review protocol for the Primary Mathematics topic area, one was within the scope of the review protocol but did not meet WWC group design standards, one study meets WWC groups design standards without reservations, and one meets WWC group design standards with reservations; the findings from these studies are summarized in this report. A complete list and disposition of all studies reviewed are provided in the references.

The report includes reviews of all previous studies that met WWC group design standards with or without reservations and resulted in a revised disposition for one study. DiLeo (2007) received a rating of *meets WWC group design standards without reservations*, where it had previously received a rating of *meets WWC group design standards with reservations*. The revised disposition is due to additional reviewer guidance on whether sample loss should be considered attrition. Previously, the exclusion of a magnet school (and its associated comparison classrooms) was treated as attrition, resulting in high cluster-level attrition. Under the current reviewer guidance, the magnet school is treated as an exogenous subgroup, so that its exclusion from the analysis does not affect attrition.

The studies in this report were reviewed using the Standards from the WWC Procedures and Standards Handbook (version 3.0) and the Primary Mathematics review protocol (version 3.1). The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.

[3] Please see the Primary Mathematics topic area review protocol (version 3.1) for more information about the outcome domain.

[4] For criteria used in the determination of the rating of effectiveness and extent of evidence, see the WWC Rating Criteria on p. 16. These improvement index numbers show the average and range of individual-level improvement indices for all findings across the three studies that met WWC group design standards with and without reservations. The WWC was able to calculate student-level effect sizes for all three of these studies.

[5] A single magnet school with two comparison classrooms was excluded from the analysis by the authors because the school was substantially different from the other schools in the sample, which were not magnet schools. Because the school's status as a magnet school was defined prior to random assignment, the randomized controlled trial design is considered intact by the WWC, and the exclusion of this school and its two comparison classrooms is not counted as attrition.

[6] A total of 33 schools were originally part of the study's random assignment design. However, after randomization occurred, the authors learned that one school was already using *Odyssey® Math*. This school was excluded from the study because it was ineligible, since it had been using the intervention prior to random assignment. Hence, the randomized controlled trial design is considered intact by the WWC, and the exclusion of this school is not counted as attrition.

[7] In the report, the author also identified a third group of students who were considered higher math performers; these students were not eligible to receive either form of the supplemental instructional support provided to the other two groups of lower achieving students. Instead, students in this third group received only regular classroom instruction. Students were assigned to this third group

based on a strict cutoff for test scores. The study compared the math achievement of these higher achieving students to the group of lower achieving students who used *Odyssey® Math*. This comparison of *Odyssey® Math* students to the higher-performing students does not meet group design standards due to lack of baseline equivalence.

[8] The author reported statistically significant results based on both an ANOVA and ANCOVA. Both results showed positive and statistically significant differences between the intervention and comparison groups. The author-reported ANOVA results indicate a substantively important effect size (i.e., an effect size above 0.25). The author did not report an effect size for the ANCOVA results in the original study; however, the WWC-calculated effect sizes for this analysis are also substantively important. This intervention report uses the ANCOVA results because they may be more precise due to controlling for baseline covariates.

[9] A sixth school is excluded from this WWC review because teachers in that school were not randomly assigned to use either the *Odyssey® Language Arts* or *Odyssey® Math* software. The author provided results that exclude this school in response to a WWC query.

[10] The author did not indicate whether the 30 minutes of supplemental instruction was daily or weekly.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2017, January). *Primary Mathematics intervention report: Odyssey® Math.* Retrieved from http://whatworks.ed.gov

WWC Rating Criteria

Criteria used to determine the rating of a study

Study rating	Criteria
Meets WWC group design standards without reservations	A study that provides strong evidence for an intervention's effectiveness, such as a well-implemented RCT.
Meets WWC group design standards with reservations	A study that provides weaker evidence for an intervention's effectiveness, such as a QED or an RCT with high attrition that has established equivalence of the analytic samples.

Criteria used to determine the rating of effectiveness for an intervention

Rating of effectiveness	Criteria
Positive effects	Two or more studies show statistically significant positive effects, at least one of which met WWC group design standards for a strong design, AND No studies show statistically significant or substantively important negative effects.
Potentially positive effects	At least one study shows a statistically significant or substantively important positive effect, AND No studies show a statistically significant or substantively important negative effect AND fewer or the same number of studies show indeterminate effects than show statistically significant or substantively important positive effects.
Mixed effects	At least one study shows a statistically significant or substantively important positive effect AND at least one study shows 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, OR At least one study shows a statistically significant or substantively important effect AND more studies show an indeterminate effect than show a statistically significant or substantively important effect.
Potentially negative effects	One study shows a statistically significant or substantively important negative effect and no studies show a statistically significant or substantively important positive effect, OR Two or more studies show statistically significant or substantively important negative effects, at least one study shows a statistically significant or substantively important positive effect, and more studies show statistically significant or substantively important negative effects than show statistically significant or substantively important positive effects.
Negative effects	Two or more studies show statistically significant negative effects, at least one of which met WWC group design standards for a strong design, AND No studies show statistically significant or substantively important positive effects.
No discernible effects	None of the studies shows a statistically significant or substantively important effect, either positive or negative.

Criteria used to determine the extent of evidence for an intervention

Extent of evidence	Criteria
Medium to large	The domain includes more than one study, AND The domain includes more than one school, AND The domain findings are based on a total sample size of at least 350 students, OR, assuming 25 students in a class, a total of at least 14 classrooms across studies.
Small	The domain includes only one study, OR The domain includes only one school, OR The domain findings are based on a total sample size of fewer than 350 students, AND, assuming 25 students in a class, a total of fewer than 14 classrooms across studies.

Glossary of Terms

Attrition Attrition occurs when an outcome variable is not available for all participants initially assigned to the intervention and comparison groups. The WWC considers the total attrition rate and the difference in attrition rates across groups within a study.

Clustering adjustment If intervention assignment is made at a cluster level and the analysis is conducted at the student level, the WWC will adjust the statistical significance to account for this mismatch, if necessary.

Confounding factor A confounding factor is a component of a study that is completely aligned with one of the study conditions, making it impossible to separate how much of the observed effect was due to the intervention and how much was due to the factor.

Design The design of a study is the method by which intervention and comparison groups were assigned.

Domain A domain is a group of closely related outcomes.

Effect size The effect size is a measure of the magnitude of an effect. The WWC uses a standardized measure to facilitate comparisons across studies and outcomes.

Eligibility A study is eligible for review and inclusion in this report if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design.

Equivalence A demonstration that the analysis sample groups are similar on observed characteristics defined in the review area protocol.

Extent of evidence An indication of how much evidence supports the findings. The criteria for the extent of evidence levels are given in the WWC Rating Criteria on p. 16.

Improvement index Along a percentile distribution of individuals, the improvement index represents the gain or loss of the average individual due to the intervention. As the average individual starts at the 50th percentile, the measure ranges from –50 to +50.

Intervention An educational program, product, practice, or policy aimed at improving student outcomes.

Intervention report A summary of the findings of the highest-quality research on a given program, product, practice, or policy in education. The WWC searches for all research studies on an intervention, reviews each against design standards, and summarizes the findings of those that meet WWC design standards.

Multiple comparison adjustment When a study includes multiple outcomes or comparison groups, the WWC will adjust the statistical significance to account for the multiple comparisons, if necessary.

Quasi-experimental design (QED) A quasi-experimental design (QED) is a research design in which study participants are assigned to intervention and comparison groups through a process that is not random.

Randomized controlled trial (RCT) A randomized controlled trial (RCT) is an experiment in which eligible study participants are randomly assigned to intervention and comparison groups.

Rating of effectiveness The WWC rates the effects of an intervention in each domain based on the quality of the research design and the magnitude, statistical significance, and consistency in findings. The criteria for the ratings of effectiveness are given in the WWC Rating Criteria on p. 16.

Single-case design A research approach in which an outcome variable is measured repeatedly within and across different conditions that are defined by the presence or absence of an intervention.

Glossary of Terms

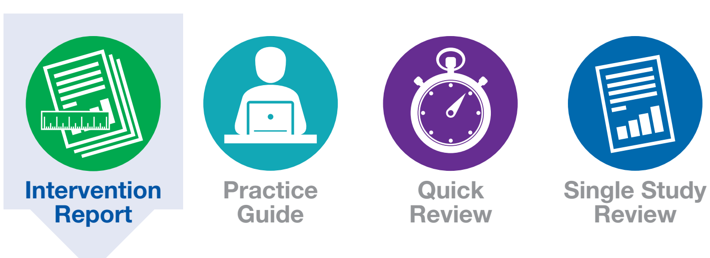
Standard deviation The standard deviation of a measure shows how much variation exists across observations in the sample. A low standard deviation indicates that the observations in the sample tend to be very close to the mean; a high standard deviation indicates that the observations in the sample tend to be spread out over a large range of values.

Statistical significance Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ($p < .05$).

Substantively important A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.

Systematic review A review of existing literature on a topic that is identified and reviewed using explicit methods. A WWC systematic review has five steps: 1) developing a review protocol; 2) searching the literature; 3) reviewing studies, including screening studies for eligibility, reviewing the methodological quality of each study, and reporting on high quality studies and their findings; 4) combining findings within and across studies; and, 5) summarizing the review.

Please see the WWC Procedures and Standards Handbook (version 3.0) for additional details.



An **intervention report** summarizes the findings of high-quality research on a given program, practice, or policy in education. The WWC searches for all research studies on an intervention, reviews each against evidence standards, and summarizes the findings of those that meet standards.

This intervention report was prepared for the WWC by Mathematica Policy Research under contract ED-IES-13-C-0010.