

WWC Review of the Report “A Multisite Cluster Randomized Trial of the Effects of CompassLearning Odyssey[®] Math on the Math Achievement of Selected Grade 4 Students in the Mid-Atlantic Region”^{1,2}

The findings from this review do not reflect the full body of research evidence on *Odyssey[®] Math*.³

What is this study about?

The study examined whether exposure to *Odyssey[®] Math*, a web-based mathematics curriculum and assessment tool, improved mathematics achievement of fourth-grade students.

The study analyzed data from 2,456 fourth-grade students in 122 classrooms in 32 elementary schools in Delaware, New Jersey, and Pennsylvania.

The study randomly assigned all fourth-grade classrooms in 32 elementary schools to intervention or comparison conditions. Intervention classrooms used *Odyssey[®] Math* for 60 minutes each week during the entire school year as a partial substitute for the regular curriculum; comparison classrooms used their school’s standard mathematics curriculum for the total math instructional time.

The study assessed the effectiveness of *Odyssey[®] Math* by comparing the mathematics achievement of students in the intervention and comparison groups in the spring of the implementation year.

Features of *Odyssey[®] Math*

Odyssey[®] Math, published by CompassLearning, is a web-based K–8 mathematics curriculum and assessment tool designed to enable teachers to differentiate student instruction and make data-driven decisions. *Odyssey[®] Math* can be used as a standalone curriculum or as a partial substitute to other mathematics curricula.

Each *Odyssey[®] Math* module contains learning activities for students that include narrative descriptions of how to solve problems, practice tasks, quizzes, and feedback. Modules also contain math tools and assessments. Specific learning activities and difficulty levels can be selected by the software or teacher.

What did the study find?

The study found no discernible effects of *Odyssey[®] Math* on mathematics achievement in the spring of the implementation year. The estimated effect size of 0.02 is not statistically significant or substantively important.

WWC Rating

The research described in this report meets WWC evidence standards without reservations

Strengths: This study is a well-implemented randomized controlled trial.

Appendix A: Study details

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

Setting The study was conducted in 122 classrooms in 32 elementary schools in Delaware, New Jersey, and Pennsylvania.

Study sample School districts and charter schools in Delaware, the District of Columbia, Maryland, New Jersey, and Pennsylvania not using *Odyssey® Math* were invited to participate in the study. Eligibility criteria for the study required that schools:

- had at least two fourth-grade classrooms with different teachers using the same curriculum,
- did not track students into classrooms based on academic performance,
- had at least one computer per student in the classroom or a computer lab, and
- had not used *Odyssey® Math* in recent years.

Thirty-two schools in Delaware, New Jersey, and Pennsylvania were eligible and agreed to participate. On average, the student population of the sample schools included 18.7% minority students, and 36.3% of students were eligible for free or reduced-price lunch. At baseline, about half of the students in the sample were female, and about 7% were English language learners. Intervention and comparison classrooms did not differ significantly on a pre-intervention measure of math achievement or other characteristics, including student socioeconomic status, percentage of English language learners, student race/ethnicity, student gender, or teacher experience and participation in professional development. The original sample included 1,448 students in 60 intervention classrooms and 1,492 students in 62 comparison classrooms. The analytic sample included 1,223 intervention students and 1,233 comparison students, with no loss of classrooms.

Intervention group Intervention classrooms were asked to replace their school's standard mathematics curriculum and instructional methods with *Odyssey® Math* software for 60 minutes per week during the entire school year. The time spent on *Odyssey® Math* was expected to be integrated into the overall math instructional time to avoid confounding the amount of instructional time with the use of *Odyssey® Math*. Teachers were not instructed on what part of their regular math curriculum to replace with *Odyssey® Math*, but could substitute *Odyssey® Math* for any combination of the following: traditional practice tasks (for example, hands-on activities using a ruler), assessment, or whole instructional modules. In practice, students spent approximately 38 minutes each week on *Odyssey® Math* in intervention classrooms. Although considerably less than recommended, CompassLearning reported that these usage times were typical of routine implementation.

Comparison group

Comparison classrooms used their school's standard mathematics curriculum for the total math instructional time. Standard curricula used by study schools were *Everyday Mathematics* (10 schools), *Scott Foresman Math* (seven schools), *Harcourt Brace Math* (five schools), or *Saxon Math* (five schools). Depending on the specific curriculum used in the school, instruction differed somewhat across comparison classrooms. However, in general, instruction was primarily teacher led, often using real-world examples for presentation and student practice. To supplement the regular curriculum, comparison classrooms in 12 schools used *Study Island* software, and three additional schools used other existing curriculum supplements. Thus, 47% of participating schools reported use of software in comparison group classrooms.

Outcomes and measurement

For both the fall (September or October) pretest and the spring (April or May) posttest, students took the TerraNova Second Edition Basic Battery math subtest, form A. Analyses used scores that were scaled on level 14 (grade 4). For a more detailed description of this outcome measure, see Appendix B.

Support for implementation

Intervention teachers participated in a large-group training session in *Odyssey[®] Math* in August before the implementation year and in another session in January of the implementation year. In addition, teachers received three individual in-class coaching sessions over the course of the year. Large-group sessions covered using the software and incorporating it into lesson plans. In-class coaching included learning management and logistics, developing assessments and reports, and tailoring assignments to individual students based on *Odyssey[®] Math* software.

Identification

This study is an Institute of Education Sciences (IES)-funded study that was conducted by Regional Educational Laboratory Mid-Atlantic.

Appendix B: Outcome measure for the mathematics achievement domain

Mathematics achievement

TerraNova Second Edition Basic Battery math subtest, form A

The TerraNova Second Edition Basic Battery is a norm-referenced standardized test of reading/language arts and math. Form A of the Battery's math subtest is appropriate for administration as a standalone assessment. Internal consistency, as measured by the Kuder–Richardson Formula 20 coefficient, is 0.93; the Cronbach's alpha for pre- and posttest on a standardized national sample is 0.91. The fourth-grade subtest's 57 selected-response items are well aligned with the National Council of Teachers of Mathematics (NCTM) 2008 standards, the National Assessment of Educational Progress (NAEP) 2008 conceptual framework, and state and local curricula. The math subtest provides a comprehensive measure of mathematics competency, including items that assess estimation and mental computation.

Appendix C: Study findings for the mathematics achievement domain

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Mathematics achievement								
<i>TerraNova Second Edition Basic Battery math subtest, form A</i>	Grade 4	122 classrooms/ 2,456 students	648.29 (38.69)	647.50 (38.18)	0.79	0.02	1	0.54
Domain average for mathematics achievement						0.02	1	Not statistically significant

Table Notes: Positive results for mean difference, effect size, and improvement index favor the intervention group; negative results favor the comparison group. The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the change (measured in standard deviations) in an average student's outcome that can be expected if the student is given the intervention. The improvement index is an alternate presentation of the effect size, reflecting the change in an average student's percentile rank that can be expected if the student is given the intervention. The statistical significance of the study's domain average was determined by the WWC; a study is characterized as not statistically significant when univariate statistical tests are reported for each outcome measure and each of the effects within the domain are not statistically significant.

Study Notes: Wijekumar et al. (2009) used a three-level hierarchical linear modeling (HLM) analysis, which accounts for clustering at the classroom and school levels. The post-test means reported here were adjusted for the classroom pretest means by the authors in the original article. The authors reported the pooled standard deviation for the unadjusted student-level posttest score; the effect size was 0.02, regardless of whether the pooled or comparison group standard deviation was used to standardize the difference. No corrections for clustering or multiple comparisons were needed. The p-values presented here were reported in the original study.

Endnotes

¹ Single study reviews examine evidence published in a study (supplemented, if necessary, by information obtained directly from the author[s]) to assess whether the study design meets WWC evidence standards. The review reports the WWC's assessment of whether the study meets WWC evidence standards and summarizes the study findings following WWC conventions for reporting evidence on effectiveness. The WWC rating applies only to the summarized results, and not necessarily to all results presented in the study. This study was reviewed using the Elementary School Math review protocol, version 2.0.

² Absence of conflict of interest: The Regional Educational Labs were provided technical assistance by Mathematica Policy Research, which also operates the WWC. For this reason, this study was reviewed by staff from subcontractor organizations.

³ A systematic review of *Odyssey[®] Math* was conducted by the WWC using the Middle School Math review protocol and the report was released in 2009.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2012, May). WWC *review of the report: A multisite cluster randomized trial of the effects of CompassLearning Odyssey[®] Math on the math achievement of selected grade 4 students in the Mid-Atlantic region*. Retrieved from <http://whatworks.ed.gov>.

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.
Improvement index	Along a percentile distribution of students, the improvement index represents the gain or loss of the average student due to the intervention. As the average student starts at the 50th percentile, the measure ranges from -50 to +50.
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 subjects 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 investigators randomly assign eligible participants into intervention and comparison groups.
Single-case design (SCD)	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.
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% ($p < 0.05$).
Substantively important	A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.

Please see the [WWC Procedures and Standards Handbook \(version 2.1\)](#) for additional details.