

# What Works Clearinghouse™



July 2014

## WWC Review of the Report “Longitudinal Evaluation of a Scale-Up Model for Teaching Mathematics with Trajectories and Technologies”<sup>1</sup>

The findings from this review do not reflect the full body of research evidence on *Technology-enhanced, Research-based, Instruction, Assessment, and professional Development (TRIAD)*.

### What is this study about?

The study examined the effects of *Technology-enhanced, Research-based, Instruction, Assessment, and professional Development (TRIAD)*, a math intervention for preschoolers that combines a curriculum, a software-based teaching tool, and in-person teacher professional development. The intervention was designed for young children, particularly those at risk of low math achievement. The study also included an assessment of whether continuing the intervention through kindergarten improved math achievement at the end of kindergarten.<sup>2</sup> To measure these effects, two versions of the intervention were delivered: (a) *TRIAD* no follow-through (*TRIAD-NFT*), where children only received *TRIAD* in preschool; and (b) *TRIAD* follow-through (*TRIAD-FT*), where children received *TRIAD* in both preschool and kindergarten.

Forty-two schools from Buffalo, NY, and Boston, MA were included in this study. Within each city, schools were grouped by the previous year’s fourth grade mathematics state assessment scores and then randomly assigned to either the *TRIAD-NFT* group (14 schools), the *TRIAD-FT* group (12 schools), or a comparison group that implemented neither intervention (16 schools). From each preschool classroom, researchers randomly selected up to 15 children to participate. Researchers assessed the math ability of these children at the start of preschool (prior to intervention), at the end of preschool (after 1 year of study participation), and at the end

of kindergarten (after 2 years of study participation). Using these completed assessments, 963 students were included in the final analysis. To measure the intervention’s effects, the researchers separately compared outcomes at the end of kindergarten for those in *TRIAD-NFT* and *TRIAD-FT* to those from the comparison group. The researchers further compared the end-of-kindergarten outcomes between *TRIAD-NFT* and *TRIAD-FT* to assess whether there were any differences due to the continuity of the *TRIAD* intervention into kindergarten.

### Features of *Technology-enhanced, Research-based, Instruction, Assessment, and professional Development (TRIAD)*

*TRIAD* is a branded math intervention that combines a curriculum (*Building Blocks*), a software-based teaching tool (*Building Blocks Learning Trajectories*), and in-person teacher professional development.<sup>3</sup> *TRIAD* is designed to help teachers tailor instruction to children’s individual needs. Two versions of *TRIAD* were implemented in this study: one that was implemented in preschool only (*TRIAD* no follow-through), and one that was implemented in both preschool and kindergarten (*TRIAD* follow-through). There were no differences in how these two versions of *TRIAD* were implemented in preschool. In kindergarten, the *TRIAD-FT* teachers used *Building Blocks Learning Trajectories* and received in-person professional development on the use of formative assessment; however, these teachers did not use the *Building Blocks* curriculum. There were no differences in the experiences of the *TRIAD-NFT* group and the comparison group during kindergarten.

What did the study find?

At the end of kindergarten, the authors found that both the *TRIAD-NFT* and *TRIAD-FT* interventions had a statistically significant positive effect on student performance on the Research-based Elementary Math Assessment (REMA) when compared to students attending comparison schools, and the WWC confirmed these findings. The authors found no difference in REMA performance when comparing students attending *TRIAD-NFT* schools to those attending *TRIAD-FT* schools.

WWC Rating

***The research described in this report meets WWC group design standards with reservations***

This study is a randomized controlled trial with unknown levels of study attrition, and thus, this study does not pass the attrition standard.<sup>4</sup> However, the study demonstrates the equivalence of the analytic samples for each comparison at baseline.

### Appendix A: Study details

Sarama, J., Clements, D. H., Wolfe, C. B., & Spitler, M. E. (2012). Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies. *Journal of Research on Educational Effectiveness*, 5(2), 105–135.

#### Additional source:

Clements, D. H., Sarama, J., & Liu, X. (2008). Development of a measure of early mathematics achievement using the Rasch model: The research-based math assessment. *Educational Psychology*, 28(4), 457–482.

**Setting** This study was conducted in preschool and kindergarten classrooms in 42 schools that were located in Buffalo, NY, and Boston, MA.

**Study sample** Researchers originally recruited schools to implement a school-level cluster randomized controlled trial in the late 2000s. Within each city, participating schools were grouped according to the mathematics state test scores of the fourth-grade students who attended the school the previous year. The schools were then randomly assigned to either implement *TRIAD-NFT*, *TRIAD-FT*, or act as a comparison school. Once the school year began, consent from families in each preschool classroom was requested. The researchers then randomly sampled up to 15 students from each classroom to be included in the study. This resulted in a sample of 5–15 students from each classroom which was included in the study.

The final analysis sample included 42 schools and 963 students that were close to evenly distributed across the three treatment arms. Fifty-three percent of the students in the study were African American, 22% were Hispanic, and 19% were White. Overall, the sample was economically disadvantaged, with 82% receiving a lunch subsidy. Approximately 13% of the students were English language learners.

**Intervention group** There were two intervention groups included in the study: (a) *TRIAD* no follow-through (*TRIAD-NFT*) implemented the *TRIAD* intervention in the preschool year only, and (b) *TRIAD* follow-through (*TRIAD-FT*) implemented the *TRIAD* intervention in both preschool and kindergarten. *TRIAD* is a math intervention that combines a curriculum, a software-based teaching tool, and in-person professional development for teachers. The overarching goal of these combined components is to help teachers individualize their teaching to each child's needs. The follow-through component of the intervention is meant to educate kindergarten teachers to build upon the gains made by students experiencing *TRIAD* in preschool. The *TRIAD* curriculum, *Building Blocks*, is a stand-alone curriculum that includes a software component for children with activities and built-in assessments.<sup>5</sup> The software-based teaching tool, *Building Blocks Learning Trajectories*, is a resource for teachers to improve their implementation of individualized lesson planning for each student. The in-person professional development components provide hands-on instruction on how to individualize teaching for students and incorporate the *Building Blocks* and *Building Blocks Learning Trajectories* components of the intervention. In kindergarten, *TRIAD-FT* teachers used *Building Blocks Learning Trajectories* and received in-person professional development; however, they did not use the *Building Blocks* curriculum. Instead, *TRIAD-FT* teachers incorporated the district's kindergarten curriculum with the other components of *TRIAD*.

### Comparison group

Students in the comparison condition did not have teachers implementing the *TRIAD* program. No additional details were provided, but schools could not participate in the study if they had any previous experience implementing either *TRIAD* or *Building Blocks* before the study began.

### Outcomes and measurement

The student-level outcome analyzed in this study is the Research-based Elementary Math Assessment (REMA). This assessment was created by two of the study authors (Clements, Sarama, & Liu, 2008) and provides a general measure of math knowledge based on the Rasch model.

### Support for implementation

*TRIAD-FT* kindergarten teachers were trained for 32 hours over seven sessions throughout the follow-through year. Professional development participation, measured as a percentage of hours attended, ranged from 16% to 100% with a mean of 63%. The focus of the professional development for kindergarten teachers was on formative assessment. Teachers were also trained on studying student trajectories using the *Building Blocks Learning Trajectories* program. In schools implementing the *TRIAD-FT* intervention, professional development staff also met with preschool and kindergarten staff at each school in order to facilitate the sharing of information. Training of preschool teachers focused on the introduction of the new curriculum, but additional detail on the professional development for *TRIAD* preschool teachers was not provided in the study.

### Reason for review

This study was identified for review by the WWC because it was supported by a grant (R305K050157) to the State University of New York at Buffalo (Principal Investigator: Douglas Clements) from the National Center for Education Research (NCER) at the Institute of Education Sciences (IES).

Appendix B: Outcome measure for the mathematics achievement domain

Mathematics achievement	
Research-based Elementary Math Assessment (REMA)	The REMA is an assessment that captures general math knowledge as a latent trait using item response theory in an interview format. The study authors reported a reliability of .92–.94 for the overall assessment. The primary outcome used for the analysis is measured at the end of kindergarten (Clements, Sarama, & Liu, 2008).

**Table Notes:** The study authors also included a teacher-level outcome measure: Classroom Observation of Early Mathematics Environment and Teaching (COEMET). This outcome was excluded in this WWC report because it is not eligible for review under the Early Childhood Education review protocol, version 2.0.

### 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								
REMA	TRIAD-NFT vs. comparison	30 schools/618 students	52.64 (4.44)	51.61 (4.69)	nr	nr	nr	.03
Domain average for TRIAD-NFT vs. comparison						nr	nr	Statistically significant
REMA	TRIAD-FT vs. comparison	28 schools/630 students	53.47 (4.68)	51.61 (4.69)	1.86	0.40	+15	.00
Domain average for TRIAD-FT vs. comparison						0.40	+15	Statistically significant
REMA	TRIAD-FT vs. TRIAD-NFT	26 schools/678 students	53.22 (4.68)	52.64 (4.44)	nr	nr	nr	.13
Domain average for TRIAD-FT vs. TRIAD-NFT						0.40	+15	Not statistically significant

**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 student outcomes, representing the average change expected for all students 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 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. nr = not reported.

**Study Notes:** No corrections for clustering or multiple comparisons were needed. The authors controlled for preschool pretest measures and block indicators to reflect the original assignment process, and the *p*-values presented here are based on these models as they were reported in the study. However, for those contrasts that have an "nr" under "WWC calculations," the WWC could not reconcile the effect size estimates from the models as reported in the study with WWC-calculated effect size estimates using additional statistics reported in the study. Specifically, the impact estimates described in prose on page 116 could not be translated into meaningful effect sizes that aligned with the unadjusted descriptive statistics on page 117 of the article (or any other metric of the outcome that the single study review team considered). As a result, the WWC does not report mean difference, effect size, or improvement index statistics from the information presented in the study in this table, since the magnitude of these statistics cannot be credibly identified given the information in the article (note: these statistics were not used to determine the level of evidence for these contrasts). In the article, the authors calculated the effect size for the TRIAD-NFT vs. comparison as  $g = .21$  and the effect size for the TRIAD-FT vs. TRIAD-NFT as  $g = .14$  (and the direction and statistical significance of the study-reported impacts were used to determine the level of evidence from these contrasts, since the direction and significance of the impacts was deemed to be credible by the WWC, even though the magnitude of the reported impact estimates were not replicable with other information in the article).

Unlike the other two contrasts shown in this table, the TRIAD-FT vs. comparison subgroup did not require a pretest adjustment due to small baseline differences. For this final comparison, the WWC was able to perform an effect size estimate based on reported statistics from the study. For this comparison, the WWC calculated the program group mean using a difference-in-differences approach (see WWC Handbook) by adding the impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. Please see the WWC Procedures and Standards Handbook, version 3.0 for more information. Both TRIAD-NFT and TRIAD-FT are characterized as having a statistically significant positive effect because they each have a positive and statistically significant effect for at least one measure when compared to the comparison group, and no effects are negative and statistically significant. For more information, please refer to the WWC Standards and Procedures Handbook, version 3.0, page 25.

### Endnotes

<sup>1</sup> Single study reviews examine evidence published in a study (supplemented, if necessary, by information obtained directly from the authors or related studies) to assess whether the study design meets WWC group design standards. The review reports the WWC's assessment of whether the study meets WWC group design standards and summarizes the study findings following WWC conventions for reporting evidence on effectiveness. This study was reviewed using the Early Childhood Education review protocol, version 2.0. The WWC rating applies only to study outcomes that were eligible for review under this topic area. The reported analyses in this single study review are only for those eligible outcomes that met WWC group design standards without reservations or met WWC group design standards with reservations, and do not necessarily apply to all results presented in the study.

<sup>2</sup> There was one teacher-related outcome included in the study that is not described in this WWC report, since the protocols used for this review clarify that the focus should be on child and student outcomes. See the table notes in Appendix B for more information.

<sup>3</sup> Details about the *TRIAD* intervention can be obtained from the official website: <http://triad-research.du.edu/>

<sup>4</sup> Information required to assess attrition from random assignment to the analysis sample was not available in the primary or supplemental studies.

<sup>5</sup> Details about the *Building Blocks* curriculum can be obtained from the official website: <http://www.ubbuildingblocks.org/>

### Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2014, July). *WWC review of the report: Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies*. Retrieved from <http://whatworks.ed.gov>

### Glossary of Terms

<b>Attrition</b>	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.
<b>Clustering adjustment</b>	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.
<b>Confounding factor</b>	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.
<b>Design</b>	The design of a study is the method by which intervention and comparison groups were assigned.
<b>Domain</b>	A domain is a group of closely related outcomes.
<b>Effect size</b>	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.
<b>Eligibility</b>	A study is eligible for review if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design.
<b>Equivalence</b>	A demonstration that the analysis sample groups are similar on observed characteristics defined in the review area protocol.
<b>Improvement index</b>	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.
<b>Multiple comparison adjustment</b>	When a study includes multiple outcomes or comparison groups, the WWC will adjust the statistical significance to account for the multiple comparisons, if necessary.
<b>Quasi-experimental design (QED)</b>	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.
<b>Randomized controlled trial (RCT)</b>	A randomized controlled trial (RCT) is an experiment in which investigators randomly assign eligible participants into intervention and comparison groups.
<b>Single-case design (SCD)</b>	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.
<b>Standard deviation</b>	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 are spread out over a large range of values.
<b>Statistical significance</b>	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% ( $p < 0.05$ ).
<b>Substantively important</b>	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 \(version 3.0\)](#) for additional details.