

Abstract Title Page
Not included in page count.

Title:

Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies: Persistence of effects three years after the treatment

Authors and Affiliations:

Douglas H. Clements, PhD, University of Denver

Julie Sarama, PhD, University of Denver

Carolyn Layzer, PhD, Abt Associates

Fatih Unlu, PhD, Abt Associates

Christopher B. Wolfe, PhD, Indiana University-Kokomo

Mary Elaine Spitler, PhD, University at Buffalo

Abstract Body

Limit 4 pages single-spaced.

Background / Context:

Education needs generalizable models to scale up evidence-based practices and programs and longitudinal research evaluating the persistence of the effect of their implementation. This is particularly important given the “deep, systemic incapacity of U.S. schools, and the practitioners who work in them, to develop, incorporate, and extend new ideas about teaching and learning in anything but a small fraction of schools and classrooms” (see also Berends, Kirby, Naftel, & McKelvey, 2001; Cuban, 2001; Elmore, 1996, p. 1; Tyack & Tobin, 1992). We synthesized research to create a model for the scale of successful interventions, called TRIAD (*Technology-enhanced, Research-based, Instruction, Assessment, and professional Development*). Learning trajectories are at the core of the components of this intervention, guiding the curriculum, the instruction, the formative assessment, and the professional development. Our implementation of this model with the *Building Blocks* pre-K math curriculum has shown positive effects on pre-K mathematics (ES=.72) and on kindergarten language competencies (ES=.13, .16, .29, .36), and—particularly with follow through in the kindergarten and first grade years—on end-of-first-grade mathematics achievement (ES=.28 for those with just pre-K TRIAD, .51 for those who also experienced the TRIAD Follow-Through component in Kndg. and 1st grade). Although there was no evidence that the Building Blocks intervention was differentially effective for schools with different percentages of students with free or reduced lunch or English Language Learners nor for individual children with or without IEPs (Authors, 2008, 2009, 2011), there was evidence that the intervention was differentially effective for one ethnic/racial comparison: African-American children learned less than other children in the same control classrooms and more than other children in the same TRIAD classrooms up to first grade.

TRIAD’s theoretical framework (Sarama, Clements, Starkey, Klein, & Wakeley, 2008) is an elaboration of the *Network of Influences* model (Sarama, Clements, & Henry, 1998), illustrated in Figure 1. It is consistent with, but extends in levels of detail, such theories as diffusion theory and the overlapping spheres of influence (Rogers, 2003; Showers, Joyce, & Bennett, 1987). It applies to the preschool intervention and, recursively, to the longitudinal intervention—the follow through treatment—and its evaluation (see the lower right corner of Fig. 1). The TRIAD model involves 10 research-based guidelines for scaling up (space constraints prohibit full description, but see Sarama et al., 2008, or TRIADscaleup.org).

Purpose / Objective / Research Question / Focus of Study:

Lasting effectiveness can be categorized as *sustainability* or *persistence*. We use sustainability to mean the length of time an innovation continues to be implemented with fidelity, the topic of a different TRIAD study (2013). We use *persistence* to mean the continuation of the effects of an intervention in individual children after the end of research-project-based support. To study persistence, we designed and evaluated the effectiveness of TRIAD’s two treatments—with and without following through—5 years after the pre-K component and 3 years after the end of even the follow-through treatment.

Setting:

The study took place in pre-K to first grade classrooms in two urban school districts, the Buffalo

Public School system in Buffalo, NY and the Boston Public School system in Boston, MA (a third site, in Nashville, TN/Vanderbilt University, did not have a Follow-Through intervention).

Population / Participants / Subjects:

Participants were all students of 106 (72 treatment) prekindergarten classrooms. Students enrolled at randomly assigned schools who returned parental consents were eligible. Student participants were four-year-olds (51% Female) of mixed ethnicity (53% African-American, 21% Hispanic, 19% White, 3.7% Asian Pacific, 1.8% Native American, and .6% Other). Most (82.33%) received free or reduced lunch, 13.5% had limited English proficiency, and 10% had an IEP.

Intervention / Program / Practice:

We created a research-based model to meet the aforementioned scale-up challenge in the area of early mathematics, with the intent that the model generalize to other subject matter areas and other age groups. The specific goal of our implementation of the TRIAD (*Technology-enhanced, Research-based, Instruction, Assessment, and professional Development*) model is to increase math achievement in young children, especially those at risk, by means of centering aspects of the curriculum—mathematical content, pedagogy, technology, and assessments—on a common core of learning trajectories. For pre-K, this was facilitated by our introduction of the *Building Blocks* pre-K curriculum, designed on our learning trajectories. The Follow-Through treatment was more difficult, involving training teachers on the learning trajectories separately, and then on how such knowledge could be used to teach their regular mathematics curriculum (*Investigations in Number, Data, and Space*) more effectively. We used the software application, *Building Blocks Learning Trajectories (BBLT)*, which provides scalable access to the learning trajectories via descriptions, videos, and commentaries. We also offered teachers supplementation of their curriculum with the *Building Blocks Software*, also based on learning trajectories (but, unlike the print materials, the software progresses to 3rd grade). The *professional* development was also limited to only 5 days of training starting *during* the year of data collection (the 15 days of pre-K training started a full year before data collection).

Research Design:

In a CRT design, schools within each district were publicly assigned to one of three treatment groups using a randomized block design (using a table of random numbers, with blind pointing to establish the starting number). Data were analyzed with hierarchical linear modeling (HLM, Raudenbush, Bryk, Cheong, & Congdon, 2006; Raudenbush & Liu, 2003).

Data Collection and Analysis:

For legacy data, all assessments were completed each year of the treatment, including the Classroom Observation of Early Mathematics Environment and Teaching (COEMET) and child outcomes in math (Research-based Elementary Math Assessment, REMA; literacy and language assessments were also administered, but are not the focus of this report). For this study, mathematics achievement data were collected in the children's fourth-grade year, both in the fall and spring, using the REMA 3-5.

Although we have explored a model that pools data across Fall 2011 and Spring 2012, we determined that it was better to run two separate models for Fall 2011 and Spring 2012 as (i) the relationship between covariates included in the model and test scores seem to be different across

the two time points (especially the relationship between being African-American and math achievement scores) and (ii) doing so does not lead to a decrease in precision (which would be the primary reason for using the pooled model). Therefore results reported here are from two-level cross-sectional HLMs estimated separately for the Fall 2011 and Spring 2012 data, nesting students within schools and controlling for randomization block indicators, student-level indicator for being African-American, and school-level percentages of students eligible for free/reduced price lunch and students who are LEP.

Findings / Results:

Table 1 presents the estimated impacts for the follow through (hereafter denoted by FT) and the original TRIAD condition without the follow through component (hereafter denoted by nFT) on the Fall 2011 and Spring 2012 test scores. All impact estimates are displayed in effect size units, which were calculated by dividing the impact estimates by the pooled standard deviation of the test scores. (The pooled standard deviation of the Fall 2011 test scores is 0.84 and the pooled standard deviation of the Spring 2012 test score is 1.03.) Table 1 suggests that there is no persistent impact of FT or nFT (when compared with students placed in the control group originally) at either time point.

To test the heterogeneity in the impacts on Fall 2011 and Spring 2012 test scores, we modified the cross-section two-level HLMs described above to include interactions between the FT and nFT indicators and measures for three subgroups: African-American (which was available at the student-level), school-level percent eligible for free/reduced price lunch, and school-level percent LEP. Note that these interactions are introduced to the impact model separately; that is, three separate models are estimated for the three subgroups for each time point. We observed a notable result only for African-Americans, which are presented in Table 2 below. These analyses suggest that none of the subgroup-specific impact estimates are statistically significant. In Spring 2012, however, the difference between the impact of FT for African-Americans (0.15 ES, p-val: 0.28) and the impact for non -African-Americans (-0.17, p-val: 0.20) was 0.32 of a standard deviation and statistically significant at the $p < 0.10$ level. The same difference in the impacts in Fall 2011 was smaller, 0.16, and not statistically significant.

Conclusions:

TRIAD was effective during the treatment period. The TRIAD implementation included a complete intervention in pre-K, and the impact was strong at pre-K. Although there was no evidence that the Building Blocks intervention was differentially effective for schools with different percentages of students with free or reduced lunch or English Language Learners nor for individual children with or without IEPs (Authors, 2008, 2009, 2011), there was evidence that the intervention was differentially effective for one ethnic/racial comparison: African-American children learned less than other children in the same control classrooms and African-American children learned more than other children in the same Building Blocks classrooms up to first grade.

The TRIAD effect did not persist through three years of non-follow through for the full sample. No full-sample effects were found in the children's fourth-grade year, three years after the end of even the follow-through treatment.

The TRIAD effect may have persisted for African-American children. For African-Americans, the effect of FT seems to be stable in both Fall 2011 (fall of 4th grade) and Spring 2012 – about 0.15 of a standard deviation but it is not statistically significant at either time point. When this impact (i.e., the difference between FT and CTRL for African-Americans) is compared with the impact for non African-Americans (i.e., the difference between FT and CTRL for non African-Americans), the differential is small and not significant in Fall 2011 but it is larger – 0.32 of a standard deviation – and significant at $p < 0.10$. It may be that the Building Blocks intervention is particularly effective in ameliorating the negative effects of low expectations for African-American children’s learning of mathematics (see National Mathematics Advisory Panel, 2008).

The TRIAD follow-through component was important for persistence of effects. By the end of first grade, the TRIAD-NFT group was no longer significantly higher than the control group (ES = .17). The TRIAD Follow-Through group outperformed the control group (ES = .47) and the TRIAD-NFT group (ES = .26). Further, only the comparison for the FT, relative to the control, African-Americans was (marginally) significant by the end of the children’s fourth-grade year. Thus, the Follow-Through treatment had "value added." Multiple studies have reported that preschool gains “fade.” This is often reported without adequate attention to the follow-up—more frequently, the *lack* of follow-up—planned and implemented for these children. “It is unrealistic, given our knowledge of development, to expect short-term early interventions to last indefinitely, especially if children end up attending poor quality schools. It is magical thinking to expect that if we intervene in the early years, no further help will be needed by children in the elementary school years and beyond” (Brooks-Gunn, 2003). Although this might appear to be an issue of effective “educational engineering,” *the issue has momentous policy implications.* Interpretations of this “fade” often call for *decreased* funding and attention to preschool (Fish, 2003, 2007). Although this may appear reasonable—“If effects fade out, why fund that intervention?”—We believe this mistakenly treats initial effects of interventions as independent of the future school contexts. That is, they theoretically reify the treatment effect as an entity that should persist unless it is "weak" or evanescent, susceptible to fading. Instead, we believe children’s trajectories must be studied as they experience different educational courses. Treatment effects are relative, both in contrasting experimental and control groups and, longitudinally, to the nature of educational experiences these groups receive subsequently. *The fact that at least some effects last for African-American children 3 years after the end of all interventions support the need for continued follow through interventions in mathematics for at-risk children.*

Appendices

Not included in page count.

Appendix A. References

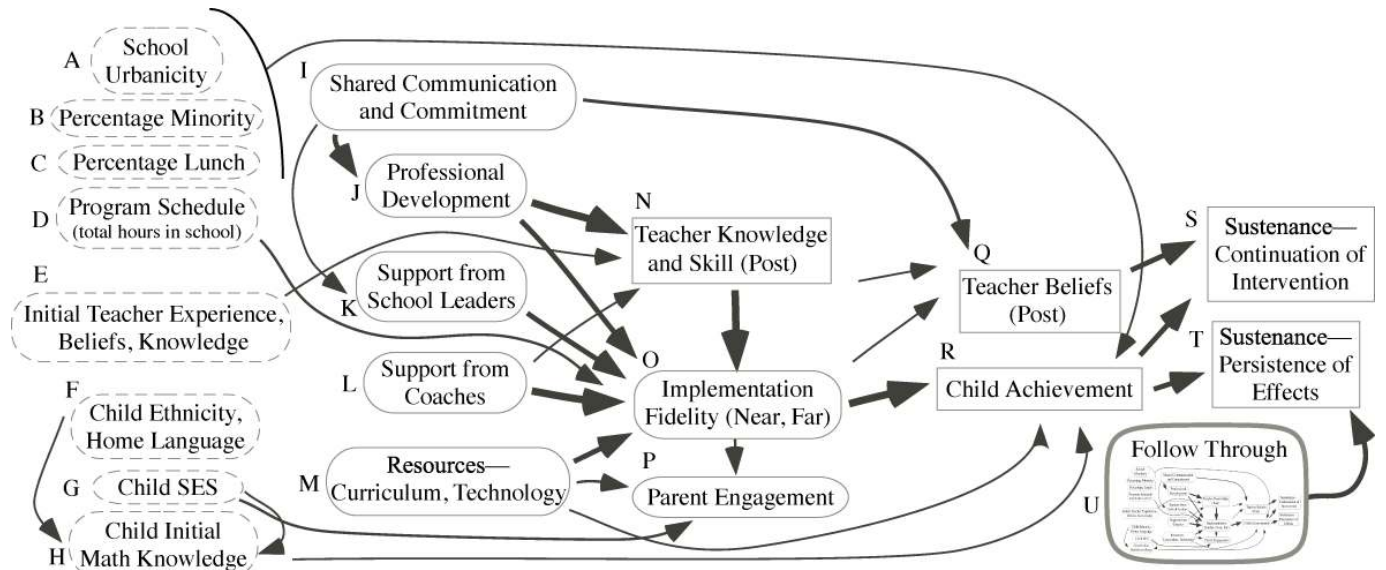
References are to be in APA version 6 format.

References

- Berends, Mark, Kirby, Sheila Nataraj, Naftel, Scott, & McKelvey, Christopher. (2001). *Implementation and performance in New American Schools: Three years into scale-up*. Santa Monica, CA: Rand Education.
- Brooks-Gunn, Jeanne. (2003). Do you believe in magic? What we can expect from early childhood intervention programs. *Social Policy Report*, 17(1), 1; 3-14.
- Cuban, Larry. (2001). *Oversold and underused*. Cambridge, MA: Harvard University Press.
- Elmore, Richard F. (1996). Getting to scale with good educational practices. *Harvard Educational Review*, 66, 1-25.
- Fish, Reva. (2003). *Effects of attending prekindergarten on academic achievement*. (Masters Thesis), University of Buffalo, State University of New York, Buffalo, NY.
- Fish, Reva. (2007). *Relationship between education intensity in kindergarten and grade 1 and the academic benefits of attending preschool*. (Doctoral dissertation), University of Buffalo, State University of New York, Buffalo, NY.
- National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington D.C.: U.S. Department of Education, Office of Planning, Evaluation and Policy Development.
- Raudenbush, Stephen W., Bryk, Anthony S., Cheong, Y. F., & Congdon, R. (2006). *HLM: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.
- Raudenbush, Stephen W., & Liu, X. F. (2003). *Optimal design*. Lincolnwood, IL: Scientific Software.
- Rogers, Everett M. (2003). *Diffusion of innovations* (Fourth ed.). New York: The Free Press.
- Sarama, Julie, Clements, Douglas H., & Henry, Julie Jacobs. (1998). Network of influences in an implementation of a mathematics curriculum innovation. *International Journal of Computers for Mathematical Learning*, 3, 113-148.
- Sarama, Julie, Clements, Douglas H., Starkey, Prentice, Klein, Alice, & Wakeley, Ann. (2008). Scaling up the implementation of a pre-kindergarten mathematics curriculum: Teaching for understanding with trajectories and technologies. *Journal of Research on Educational Effectiveness*, 1, 89-119.
- Showers, Beverly, Joyce, Bruce, & Bennett, Barrie. (1987). Synthesis of research on staff development: A framework for future study and a state-of-the-art analysis. *Educational Leadership*, 45(3), 77-87.
- Tyack, David, & Tobin, W. (1992). The "grammar" of schooling: Why has it been so hard to change? *American Educational Research Journal*, 31, 453-479.

Appendix B. Tables and Figures

Figure 1: Revised Network of Influences Theoretical Framework including Follow-Through*



* For this study, note that the Follow-Through model in the lower right-hand corner is simply a copy of the same *Network of Influences* framework for upper grades. *Contextual variables in dotted ovals* include the school (A-D), teacher (E), and child (F-H) factors. For example, child socioeconomic status, or SES (G), impacts children’s initial math knowledge (H), which influences children’s achievement (R)—an outcome variable indicated by the solid rectangle. *Implementation variables in solid ovals* are features that the project can encourage and support, but cannot control absolutely. For example, **heavy arrows** from professional development (J), to teacher knowledge (N), to implementation fidelity (O), to child achievement (R), indicate the strong effects in that path. Support from coaches (L) also has a strong effect on implementation fidelity, while other factors (J, K, M) are influential, but to a moderate degree (not all small effects are depicted). Relationships are further described in the following section

Table 1: Impact Estimates for the Full Sample (effects displayed in effect size unit)

	Fall 2011		Spring 2012	
	Impact Est	P-Value	Impact Est	P-Value
FT vs CTRL	.047	.80	-.035	.79
nFT vs CTRL	-.089	.62	-.081	.52

Table 2: Subgroup Impact Estimates and Test for Differential Impacts: African-Americans
(effects displayed in effect size unit)

	HLM for Fall 2011						HLM for Spring 2012					
	African Amer.		Not African Amer.		Differential Impact AA vs. not AA		African Amer.		Not African Amer.		Differential Impact AA vs. not AA	
	Impact Est	p	Impact Est	p	Diff.	p	Impact Est	p	Impact Est	p	Diff.	p
FT vs CTRL	.150	.46	-.004	.98	.155	.44	.149	.28	-.167	.20	.316	.07
nFT vs CTRL	-.021	.91	-.102	.58	.080	.67	.038	.78	-.125	.33	.162	.33