Abstract Title Page

Title:

Longitudinal Evaluation of a Scale-up Model for Teaching Mathematics with Trajectories and Technologies: Mechanisms of Persistence of Effects

Author:

Douglas H. Clements (with Julie Sarama, Mary Elaine Spitler, Christopher B. Wolfe), University at Buffalo, The State University of New York

Abstract Body

Background / Context:

Although the successes of some research-based educational practices have been documented, equally recognized is 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). There may be no more challenging educational and theoretical issue than scaling up educational programs across a large number of diverse populations and contexts in the early and primary educational system in the U.S., avoiding the dilution and pollution that usually plagues such efforts to achieve broad success.

Our TRIAD model (Sarama, Clements, Starkey, Klein, & Wakeley, 2008), including the *Building Blocks* curriculum, have significantly and substantially increased preschooler's mathematical competence, both in previous studies (Clements & Sarama, 2008, g = 1.07) and in our present, largest implementation (Clements, Sarama, Spitler, Lange, & Wolfe, in press, g = .72). The present study was the first to evaluate the effects of this model longitudinally.

Longitudinal evaluation is uniquely important for the preschool and primary years. Lasting effectiveness can be categorized as *sustainability* or *persistence.* We use sustainbility to mean the length of time an innovation continues to be implemented with fidelity, the topic of a different TRIAD study (2010). We use persistence to mean the continuation of the effects of an intervention in individual children. Some studies indicate that early interventions can have lasting effects. For example, several studies have shown positive and long-lasting effects of preschool experience (Broberg, Wessels, Lamb, & Hwang, 1997; Gray, Ramsey, & Klaus, 1983; Magnuson & Waldfogel, 2005; Montie, Xiang, & Schweinhart, 2006). However, there is considerable empirical research and resultantant (practical) assertions that "preschool gains fade" in the primary grades. For example, in one study of six cohorts, gains in preschool weakened as children progressed through the primary grades, disappearing by fourth grade (Fish, 2003). Other studies show a similar fade (Natriello, McDill, & Pallas, 1990; Preschool Curriculum Evaluation Research Consortium, 2008; Turner & Ritter, 2004; U.S. Department of Health and Human Services — Administration for Children and Families, 2010).

Although an ostensible reason is that early effects are themselves evanescent, we believe that a *contradictory explanation of the mechanisms at work is more cogent.* That is, we hypothesize that most present educational contexts (e.g., curricula, minimal requirements, teaching practices) are unintentionally and perversely aligned *against* early interventions. Kindergarten and first grade curricula assume little or no mathematical competence, so only low-level skills are taught. Most teachers are required to follow such curricula rigidly and remain unaware that some of their students have already mastered the material they are about to "teach" (Clements & Sarama, 2009; National Research Council, 2009; Sarama & Clements, 2009). Further, biases may negatively affect the subsequent school experiences of at-risk preschoolers. For example, kindergarten teachers rated Head Start children's math ability as lower than that of other children, even though there was direct assessments showed no such diffferences (U.S. Department of Health and Human Services — Administration for Children and Families, 2010). Teachers are held accountable for getting the largest number of students to pass minimal competency assessments, engendering the belief that higher performing students are "doing fine." *Within this context and without continual, progressive support, early gains are lost. In this way, we believe the present U.S. educational system unintentionally but insidiously re-opens the gap between students from low- and higher-resources communities.*

For these reasons, we designed and evaluated the effectiveness of TRIAD's Follow-Through treatment, testing our hypothesis that such follow through is the "missing piece" in many early interventions whose evaluations have found less positive effects.

TRIAD's theoretical framework (Sarama, et al., 2008) is an elaboration of the *Network of Influences* model (Sarama, Clements, & Henry, 1998), illustrated in Figure 1 (please insert figure 1 here). 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 UBTRIAD.org).

Purpose / Research Questions

Our overarching research question was: ***What are the long-range (persistence of) effects of the TRIAD intervention, with and without follow through, on achievement?*** The present research included two experimental groups (and one control group). In both, pre-K teachers participated in the intervention. In the *Follow-Through* experimental group, teachers in grades K and 1 were taught about the pre-K intervention and ways to build upon it. Do both experimental groups outperform those in the comparison group in math achievement on the average, at the end of kindergarten and first grade? Do children in the experimental TRIAD Follow-Through (TRIAD-FT) group on the average outperform children in the TRIAD (non-follow through, TRIAD-NFT) experimental group (the value added question)?

Setting:

The study took place in pre-K 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).

Participants:

In the Buffalo Public Schools, all schools whose pre-K teachers had not previously been involved in *Building Blocks* (e.g., Clements & Sarama, 2007; Clements & Sarama, 2008; Sarama, 2004; Sarama & Clements, 2002) or TRIAD (Sarama, et al., 2008) research or development projects were included. Extending the evaluation to distal sites— essential for generalizing to the target population of all U.S. pre-Ks—all Boston, MA schools that were not adopting a new pre-K curriculum that year and whose principals agreed to participate were included. For the pre-K study, there was only 5% attrition (from control, 12 from Buffalo and 6 from Boston from TRIAD, 37 in Buffalo and 15 from Boston schools, for a total of 70), most moved out of state, but some were ill for the entire posttest period, leaving a total of 1305 children with complete data on both pretest and posttest). Analyses revealed no significant difference between those who left and remained in mean pretest achievement (F= 2.09(2), p = .148); further, the small effect size, ES < .01, indicates that any effect on the findings was negligible. By the end of the kindergarten and first grade years, the populations remaining in the original schools included the following: TRIAD-FT—K 348, 1st 272; TRIAD-NFT—K 335, 1st 242; and control—K 290, 1st, 257 (we will also present analyses of the intent to treat groups, which had much smaller attrition).

Intervention:

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 (*T*echnology-enhanced, *R*esearch-based, *I*nstruction, *A*ssessment, and professional *D*evelopment) 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 the learning trajectories separately, and then 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 (see Fig. 2—please insert Figure 2 about here). 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 Follow Through 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 ordered on the basis of their average scores on state-based math achievement tests and then 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 Collection and Analysis:

All assessments were completed each year, 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). Data were analyzed with hierarchical linear modeling (HLM, Raudenbush, Bryk, Cheong, & Congdon, 2000; Raudenbush & Liu, 2003). All level-2 (school) predictors were centered on their group means. All interactions were computed on mean-centered transformations of the variables involved. Effect sizes were computed for significant main effects by dividing the regression coefficient by the pooled posttest standard deviation.

Findings:

Kindergarten. Consistent with the results from pre-K (in which the experimental treatments were identical), both experimental groups outperformed the control group in math achievement at the end of kindergarten (TRIAD-NFT ES = .34, $p <. 01$; TRIAD-FT ES = .55, $p < .01$). However, contrary to our hypothesis, the TRIAD-FT did not statistically significantly outperform the TRIAD group (ES = .175, p = p>.05). Of the tested moderators (child-level-racial/ethnic group, gender, IEP status; school level-percentage of children who are English Language Learners and free/reduced lunch), only one was significant. Children identified as African American children in the TRIAD-FT group outperformed those in the other two groups when controlling for pre-K pretest (Control coeff: -.208, p < .0001; BB: -.214, p < .0001).

First grade. The TRIAD-NFT group was no longer significantly higher in math achievement at the end of first grade (ES = .17, *ns).* The TRIAD-FT group continued to outperform the control group (TRIAD-FT ES = .47, $p < .01$). At this grade, consistent with our hypothesis, the TRIAD-FT significantly outperformed the TRIAD-NFT group-the Follow Through treatment had statistically significantly "value added" (ES = .25, $p <. 05$). African American and Hispanic children displayed the lowest scores at first grade, but at this grade, the interaction was no longer significant. There was not a significant main effect of gender, but there was a significant interaction between gender and the TRIAD-FT group. Girls performed lower than boys in the control group, about the same as boys in the TRIAD-NFT group, and slightly better than boys in the TRIAD-FT group.

Mediators of the Follow Through treatment. We tested COEMET variables as mediators. In Kindergarten, the Number of SMAs (Specific Math Activities) mediated the effect of the TRIAD-FT compared to Control treatment (IE: .081, CI: .006 - .195). Also, the Classroom Culture subscale mediated the effects of the TRIAD-FT compared to the TRIAD-NFT treatment (IE: .139, CI: .027 - .297), even when controlling for pre-K posttest as well as pretest-the "value added" condition (IE: .047; CI: .015 - .134). In First Grade, our present findings have yielded no significant mediators, but analyses are continuing.

Conclusions:

This is the first study to report on the longitudinal effects of an implementation of the TRIAD model of scale up, which has proven effective at scale in increasing preschoolers' mathematics achievement in multiple studies (Clements & Sarama, 2008; Clements, et al., in press; Sarama, et al., 2008). Here we address the persistence of effects on individuals over time (sustainability of the main pre-K intervention is addressed in a separate study).

The TRIAD implementation included a complete intervention in pre-K, but not in the subsequent two years. This design allowed us to assess the mechanisms for the persistence of effects (or lack thereof) of the full TRIAD intervention under two conditions, with and without follow through.

At the end of Kindergarten, both experimental groups outperformed the control group in math achievement at the end of kindergarten. However, contrary to our hypothesis, the TRIAD Follow Through did not statistically significantly outperform the TRIAD Non-Follow Through (TRIAD-NFT) group. Although it is encouraging that the gains of both TRIAD groups persisted, these findings did not support our hypothesis that Follow Through would be effective and necessary for this persistence. Nevertheless, the effect size of the TRIAD Follow Through group vs. the control group (.55) was greater than that of the TRIAD-NFT group (.34), so the trends were consistent with our expectations.

Further, those trends did continue and all hypotheses were then supported. That is, 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, "value added" (ES = .26). 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. We designed and evaluated the effectiveness of TRIAD’s Follow-Through treatment, testing our hypothesis that such follow through is the “missing piece” in many early interventions whose longitudinal evaluations have found less positive effects. We agree that, “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.

Even in our TRIAD Follow-Through treatment, multiple factors impeded implementation, including teacher's views that district rules and "fidelity police" demanded following scripts and schedules exactly—*and would not allow formative assessment or curriculum contraction.* These factors were especially present at Kindergarten, when a new edition of the mathematics curriculum (unfortunately) coincided with the TRIAD implementation for that grade. Factors such as these appear to have led to the lack of significant differences between the TRIAD-NFT group and the TRIAD Follow Through group at that grade. First grade teachers had already implemented their new edition the previous year, and that may have helped them be more receptive to modifications and pedagogical strategies, such as formative assessment based on learning trajectories, that the TRIAD Follow Through Intervention emphasized.

These factors indicate that the Follow Through condition lacked elements of the TRIAD model (see descriptions in Sarama, et al., 2008). That is, the lack of a shared vision of teaching, and especially the constraints on school leaders' support of the innovation appeared to prevent learning trajectories from standing at the core of the first grade, and especially kindergarten, teachers' curriculum and teaching.

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. 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 in Kindergarten. 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). In first grade, similar changes in expectations may have accounted for girls' (vs. boys') better performance in the TRIAD groups vs. the control group.

There are five basic recommendations. (1) Curriculum and policy should ensure that children, especially those living in poverty, should be provided with research-based, focused early mathematical interventions which can increase their knowledge of multiple mathematical concepts and skills (including, but also going beyond number). (2) It is essential that preschool mathematics interventions be continued into the primary grades. (3) The Curriculum Research Framework (Clements, 2007) upon which the curriculum was based has been repeatedly empirically supported and may serve as a useful guide to policy makers, curriculum and software developers, and administrators. (4) The learning trajectories at the core of the curriculum and TRIAD model may constitute a useful construct in future research, curriculum development, and professional development efforts. (5) This is the first study of this kind of which we are away. We need more research on the conditions that children from early interventions enter in the primary school years.

Appendixes

References

Berends, M., Kirby, S. N., Naftel, S., & McKelvey, C. (2001). *Implementation and performance in New American Schools: Three years into scale-up*. Santa Monica, CA: Rand Education.

Broberg, A. G., Wessels, H., Lamb, M. E., & Hwang, C. P. (1997). Effects of day care on the development of cognitive abilities in 8-year-olds: A longitudinal study. *Developmental Psychology, 33*, 62–69.

Brooks-Gunn, J. (2003). Do you believe in magic? What we can expect from early childhood intervention programs. *Social Policy Report, 17*(1), 1; 3-14.

Clements, D. H. (2007). Curriculum research: Toward a framework for 'research-based curricula'. *Journal for Research in Mathematics Education, 38*, 35–70.

Clements, D. H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the *Building Blocks* project. *Journal for Research in Mathematics Education, 38*, 136-163.

Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal, 45*, 443-494.

Clements, D. H., & Sarama, J. (2009). *Learning and teaching early math: The learning trajectories approach*. New York: Routledge.

Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (in press). Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. *Journal for Research in Mathematics Education*.

Cuban, L. (2001). *Oversold and underused*. Cambridge, MA: Harvard University Press.

Elmore, R. F. (1996). Getting to scale with good educational practices. *Harvard Educational Review, 66*, 1-25.

Fish, R. (2003). *Effects of attending prekindergarten on academic achievement.* Unpublished Masters Thesis, University of Buffalo, State University of New York, Buffalo, NY.

Fish, R. (2007). *Relationship between education intensity in kindergarten and grade 1 and the academic benefits of attending preschool.* Unpublished Doctoral dissertation, University of Buffalo, State University of New York, Buffalo, NY.

Gray, S. W., Ramsey, B. K., & Klaus, R. A. (1983). The early training project 1962-1980. In Consortium for Longitudinal Studies (Ed.), *As the twig is best…Lasting effects of preschool programs* (pp. 33-69). Mahwah, NJ: Lawrence Erlbaum Associates.

Magnuson, K. A., & Waldfogel, J. (2005). Early childhood care and education: Effects on ethnic and racial gaps in school readiness. *The Future of Children, 15*, 169-196.

Montie, J. E., Xiang, Z., & Schweinhart, L. J. (2006). Preschool experience in 10 countries: Cognitive and language performance at age 7. *Early Childhood Research Quarterly, 21*, 313–331.

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.

National Research Council. (2009). *Mathematics in early childhood: Learning paths toward excellence and equity*. Washington, DC: National Academy Press.

Natriello, G., McDill, E. L., & Pallas, A. M. (1990). *Schooling disadvantaged children: Racing against catastrophe*. New York: Teachers College Press.

Preschool Curriculum Evaluation Research Consortium. (2008). *Effects of Preschool Curriculum Programs on School Readiness (NCER 2008-2009)*. Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education, U.S. Government Printing Office.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. (2000). HLM5: Hierarchical linear and nonlinear modeling. Lincolnwood, IL: Scientific Software International.

Raudenbush, S. W., & Liu, X. F. (2003). *Optimal design*. Lincolnwood, IL: Scientific Software.

Rogers, E. M. (2003). *Diffusion of innovations* (Fourth ed.). New York: The Free Press.

Sarama, J. (2004). Technology in early childhood mathematics: *Building Blocks*™ as an innovative technology-based curriculum. In D. H. Clements, J. Sarama & A.-M. DiBiase (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 361-375). Mahwah, NJ: Lawrence Erlbaum Associates.

Sarama, J., & Clements, D. H. (2002). *Building Blocks* for young children's mathematical development. *Journal of Educational Computing Research, 27*(1&2), 93-110.

Sarama, J., & Clements, D. H. (2009). *Early childhood mathematics education research: Learning trajectories for young children*. New York: Routledge.

Sarama, J., Clements, D. H., & Henry, J. J. (1998). Network of influences in an implementation of a mathematics curriculum innovation. *International Journal of Computers for Mathematical Learning, 3*, 113-148.

Sarama, J., Clements, D. H., Starkey, P., Klein, A., & Wakeley, A. (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, B., Joyce, B., & Bennett, B. (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.

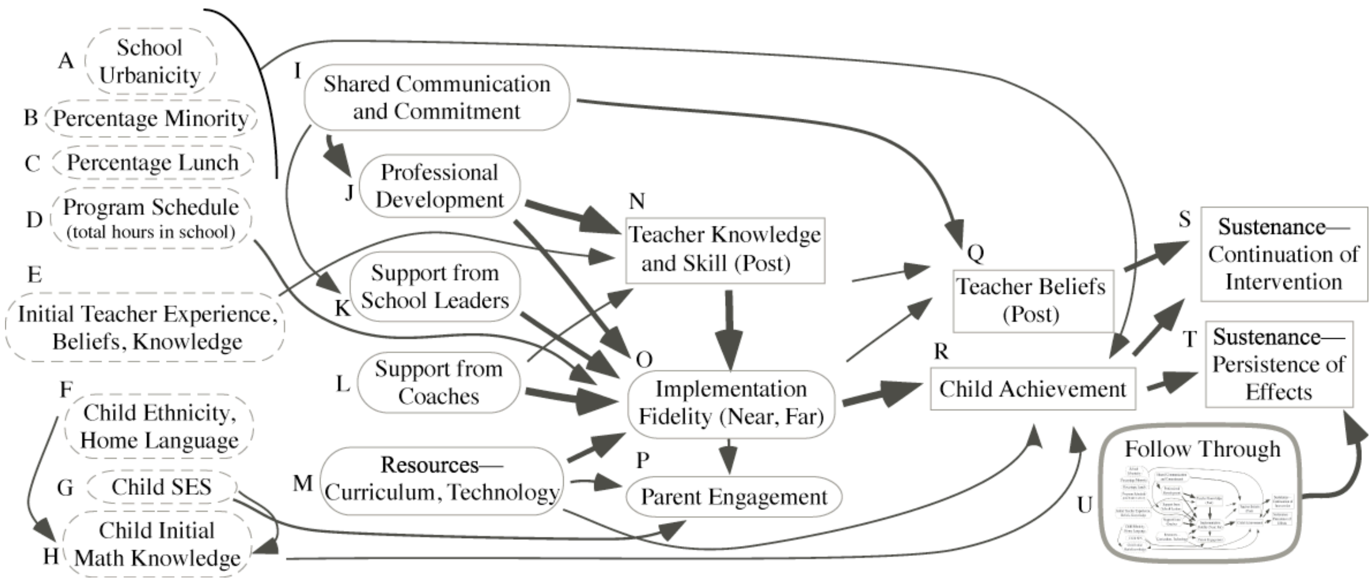
Turner, R. C., & Ritter, G. W. (2004, April). *Does the impact of preschool childcare on cognition and behavior persist throughout the elementary years?* Paper presented at the American Educational Research Association, San Diego, CA.

Tyack, D., & Tobin, W. (1992). The "grammar" of schooling: Why has it been so hard to change? *American Educational Research Journal, 31*, 453-479.

U.S. Department of Health and Human Services — Administration for Children and Families. (2010). *Head Start impact study. Final report*. Washington, D.C.: Office of Planning, Research and Evaluation, U.S. Department of Health and Human Services.

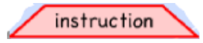
Appendix B. Tables and Figures

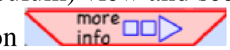
Figure 1: Revised Network of Influences *Theoretical Framework including Follow Through**



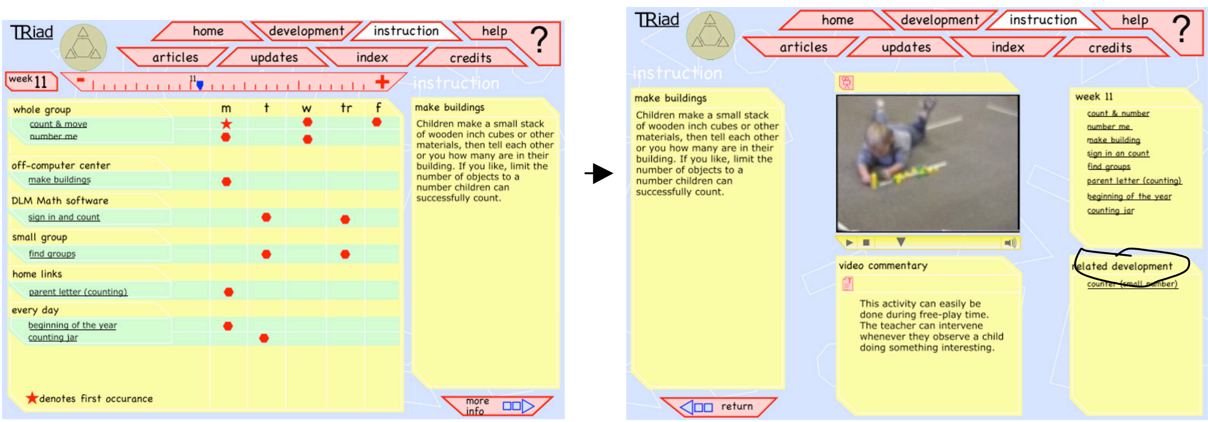
* For this study, note that the Follow Through model in the lower right-hand corner is simple a copy of the same *Network of Influences* framework for upper grades. *Contextual variables* in dotted ovals include the school (A-D), teacher (E), and child (F-H) factors. For example, child socioeconomic 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.

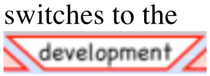
Figure 2 Building Blocks Learning Trajectories (BBLT) Web Application

BBLT provides scalable access to the learning trajectories via descriptions, videos, and commentaries. Each aspect of the learning trajectories—*developmental progressions* of children's thinking and connected *instruction*—are linked to the other. For example, teachers might choose the 

(curriculum) view and see the screen on the left, below. Clicking on a specific activity provides a description. Clicking on 

slides the screen over to reveal descriptions, several videos of the activity "in action," notes on the video, and the level of thinking in the learning trajectory that activity is designed to develop, as shown below on the right. (See UBTRIAD.org for a demonstration.)



Clicking on the related developmental level, or child's level of thinking, ringed above, switches to the 

view *of that topic and that level of thinking*. This likewise provides a description, video, and commentary on the developmental level—the video here is of a clinical interview task in which a child displays that *level of thinking*. Teachers can also study a development view, studying clinical interviews of children at each level of thinking, and, if desired, link back to activities.