Innovations in Early Mathematics: Final Report

April 2021

SRI Education™
A DIVISION OF SRI INTERNATIONAL

erikson
early math collaborative
INNOVATIONS

Acknowledgments

This study was supported by a 2010 i3 (Investing in Innovation) Development grant with additional support from CME Group Foundation. We would like to thank the teachers, administrators, and support staff at the Chicago Public Schools and the 16 elementary schools that participated in our research study. Without their commitment to learning and engaging in an innovative approach to professional development of teachers in prekindergarten to third-grade classrooms, this study would not have been possible. A special thanks also to the families and their children who participated in the study and helped us learn more about what educational supports students need to learn foundational math concepts. We are proud to call colleagues the Erikson faculty and staff who provided leadership, professional development, and coaching to teachers, and those whose behind-the-scenes work made everything possible. We are grateful to the partnership with Erikson Institute and SRI International that began on this i3-funded project and continues almost 10 years later.

September 2015 (posted to website April 30, 2021)

SRI International
Erika Gaylor
Xin Wei
Donna Spiker

Digital Promise
Ximena Dominguez

Erikson Institute
Jennifer McCray
Jie-Qi Chen
Erin Reid
Yinna Zhang
Bilge Cerezci
Suzanne Budak

Suggested citation:

Gaylor, E., Wei, X., Spiker, D., Dominguez, X., McCray, J., Chen, J.-Q., Reid, E., Zhang, Y., Cerezci, B., & Budak, S. (2021). *Innovations in Early Mathematics: Final report.* SRI International.

Contents

List of Exhibits

Executive Summary

In 2010, The Early Math Collaborative (the Collaborative) at Erikson Institute received a 5-year, $5 million grant from the Investing in Innovation (i3) Fund of the U.S. Department of Education (the Department). Supplemented by an additional $1 million grant from CME Group Foundation (http://www.cmegroupfoundation.org/), this grant was used to design, implement, and assess Innovations in Early Mathematics, an intensive professional development (PD) program addressing one of the Department's first priorities—helping students reach or exceed state learning standards. To achieve this goal, the initiative focused on improving early learning opportunities by working with teachers who serve high-needs children in prekindergarten through third grade (PK-3) in the Chicago Public Schools (CPS).

Under a subcontract from Erikson Institute, SRI served as the independent evaluator for the *Innovations in Early Mathematics Professional Development Program: PreK to Third Grade (Innovations in Early Mathematics Professional Development)* project. SRI evaluated the impact of teachers' participation in a professional development intervention to improve teachers' knowledge, attitudes, and instructional strategies in math across prekindergarten (preK) to third-grade classrooms. In this final report, we describe the rationale for and description of the project, the conceptual approach that guided the work, the evaluation design and methodology, and the findings. The executive summary highlights key features of the intervention, a brief description of the evaluation design and measures, and key findings related to teacher and child outcomes.

Intervention Objectives, Approach, and Design

Professional development (PD) is generally designed to change three aspects of teaching—dispositions toward subject matter (beliefs, confidence), content knowledge, and skills around classroom practice—in order to strengthen teachers' capacity to support and improve learning in students. However, the PD in math available to early childhood teachers is limited and largely ineffective. Curriculum-specific training tends to be shallow, emphasizing activities and neglecting deep conceptual understanding. Further, the dominant model of in-service PD is a 1-day workshop; continuing support is rarely available for teacher implementation of new practices. Lastly, PD effectiveness is limited by its lack of a conceptual framework that specifies dimensions of teacher change and guides program design. Given these issues, the goal of the Innovations in Early Mathematics Professional Development project was to develop and test the impact of a whole-teacher PD intervention for preK to third-grade teachers on teacher and student outcomes. Analogous to the "whole child" approach in early childhood education, the Whole Teacher Approach that guided the PD program suggests that in order to make real changes in teaching, PD needs to address gaps in *teacher knowledge*, but must be designed to consciously involve *teaching practice* in learning, and to address *teachers' attitudes* about teaching and content throughout implementation.

The *Innovations in Early Mathematics Professional Development* project was a 4-year, school-wide intervention that included four types of teacher PD experiences: learning labs during the school year, summer institutes, coaching, and grade-level meetings, each aimed at improving teachers' positive dispositions, math content knowledge, and instructional practice. A fifth type (leadership academies) targeted administrators such as principals and preschool program directors. The most intensive PD occurred in the first 2 years. During the 3rd and 4th years of the PD intervention, the focus changed to efforts to build capacity at the school level in creating a learning community around teaching and learning math.

The evaluation consisted of two studies conducted by SRI International: 1) an implementation study to examine fidelity to the model; and 2) an impact study to examine effects on both children and teachers.

Implementation Study

The primary research question for the implementation study was: What is the fidelity of implementation for each component of the professional development intervention? Fidelity data were collected on the major intervention components (learning labs, summer institutes, coaching, grade-level meetings, and leadership academies) in both Year 1 and Year 2. Approximately 110 lead and 40 assistant teachers participated in at least one PD component in any given implementation year. Teacher numbers fluctuate within and across years due to staff turnover and missing data.

Fidelity of implementation findings were analyzed both for Year 1 and Year 2. The findings showed that overall in Year 1, high fidelity of implementation was observed at the teacher level on all four teacher-related components, while in Year 2, high fidelity of implementation was observed at the teacher level on two of the three components: grade-level meetings attended and learning labs participation.

Impact Study

SRI designed and conducted a quasi-experimental study that included a matched comparison group of nonparticipating schools and teachers (business-as-usual comparison). Participation in the research component of the project was voluntary; as a result, fewer teachers provided data for the impact evaluation. At baseline, the evaluation sample included approximately 220 preK through third-grade teachers across 16 schools in the Chicago Public School district, with 115 teachers across the eight schools selected to participate in the PD intervention and 105 teachers in the business-as-usual schools. Analytic sample sizes vary due to staff turnover and missing data. The evaluation examined impacts of the PD intervention on teachers' dispositions (attitudes, beliefs, and confidence), instructional practice, and content knowledge as well as student math outcomes.

Compared to teachers receiving business-as-usual PD, participation in Erikson's Whole-Teacher PD led to observed changes in *teacher practice* after 1 year (ES = 0.65) which were sustained at the 4-year follow-up observations (ES = 1.01) for remaining teachers. Intervention teachers also reported higher confidence in their ability to use high-quality math instruction practices after completing the intensive 2-year PD compared to business-as-usual teachers (ES = 0.51) and at the 4-year follow-up wave (ES = 0.71). Intervention teachers demonstrated higher math content knowledge on the number sense component of the knowledge measure compared to business-as-usual teachers at the 3-year follow-up wave (ES = 0.48).

Limited impacts on student learning were observed. However, there were small impacts on the preK student subsample on a measure of mathematical knowledge and skills. That is, there was a significant treatment by age interaction for the Woodcock-Johnson-III Applied Problems, showing that younger (but not older) children who received 2 years of the intervention demonstrated a greater rate of growth in scores relative to their peers in the comparison group, even after controlling for baseline demographics and test scores. In other words, after 2 years of intervention, children who were 3 and 4 years old when Innovations began, benefited from their teachers' participation in ways that older children (60, 72, 84, and 96 months old) did not. Finally, this study did not find a significant difference between intervention and comparison children on third-grade math scores at the school level.

Conclusions

From 2011 to 2015, the Innovations in Early Mathematics project reached administrators in eight elementary schools and teachers in 106 classrooms serving an estimated 3,000 students each year with its unique approach to providing professional development in mathematics instruction to prekindergarten through third-grade teachers and school administrators. The findings for teacher outcomes suggest high-quality PD that focuses on teachers' instructional practices and pedagogical content knowledge, as well as their dispositions toward math, can lead to significant changes in teacher behavior across the early elementary grades. While the overall impact analysis did not demonstrate a significant effect on children's learning outcomes at the school level, children who were in preschool when the intervention began significantly outstripped children in the comparison condition on one of the two math achievement measures used. This innovative project showed promise for developing materials and engaging teachers in professional development that not only affected their teaching practice but also produced changes in knowledge and confidence in teachers.

Introduction

In 2010, The Early Math Collaborative (the Collaborative) at Erikson Institute received a 5-year, $5 million Development grant from the Investing in Innovation (i3) Fund of the U.S. Department of Education (the Department). Supplemented by a $1 million grant from CME Group Foundation (http://www.cmegroupfoundation.org/), this grant was used to design, implement, and assess *Innovations in Early Mathematics*, an intensive professional development (PD) program addressing the Department's **Absolute Priority 3**—helping students reach or exceed state learning standards. To achieve this goal, the initiative focused on the Department's **Competitive Priority 5—**improving early learning opportunities—by working with teachers who serve high-needs children in prekindergarten through third grade (PK-3) in the Chicago Public Schools (CPS).

In this final report, we discuss the implementation and impact of this program to improve early math achievement, including:

- A description of the Investing in Innovation (i3) grant program
- The rationale for and design of the *Innovations* project
- The details of implementation for each of the 4 years of implementation
- An explanation of the evaluation design and methodology for the implementation and impact studies
- A description of original measure development for the three teacher outcomes of interest
- A detailed section of main findings about participating schools, teachers, and children
- A discussion of the implications of the findings for professional development for teachers in prekindergarten to third-grade classrooms, and how to support and improve math achievement in young children

Description of the Investing in Innovation (i3) Program

The Investing in Innovation (i3) Fund was established under section 14007 of the American Recovery and Reinvestment Act of 2009 (ARRA). From 2010 to 2015, a total of six cohorts were funded by the U.S. Department of Education, with over a billion dollars spent altogether. In 2010, the first year of the fund's existence, nearly 1,500 organizations across the nation applied for grants and only 49 were awarded. The Collaborative's proposal to develop, implement, and assess *Innovations* was ranked eighth among all 1,500 applicants and a $5 million grant was awarded.

According to the Department, the i3 program was designed

> to provide competitive grants to applicants with a record of improving student achievement and attainment in order to expand the implementation of, and investment in, innovative practices that are demonstrated to have an impact on improving student achievement or student growth, closing achievement gaps, decreasing dropout rates, increasing high school graduation rates, or increasing college enrollment and completion rates (U.S. Department of Education, 2009).

There have been some changes to the program requirements over the years; here we describe the rules as issued in 2010.

The i3 program provided three types of grants: Scale-up, Validation, and Development. Innovations was funded by a Development grant, which provides funding to support new, high potential, and relatively untested practices. Development grants did not need to have the same level of evidentiary support as the Scale-up or Validation grants, nor did they need to plan to "go to scale" in a specific geographic region, but they had to have the potential to be further developed and scaled.

The i3 Fund competition also identified eight priorities to be addressed that included four absolute priorities and four competitive priorities. Every proposal had to address at least one of the absolute priorities, under which it would then apply. The four absolute priorities included:

- support effective teachers and school leaders;
- improve the use of data;
- complement the implementation of high standards and high-quality assessments; and
- turn around persistently low-performing public schools.

Innovations addressed Absolute Priority 3, "complement the implementation of high standards and high quality assessments," by attempting to help more students meet or exceed state learning standards in mathematics.

Competitive priorities identified other key goals to be addressed for which additional points would be awarded during the proposal evaluation process. These included:

- improve early learning opportunities;
- support college access and success;
- address the unique needs of students with disabilities and limited English-proficient students; and
- serve schools in rural local education agencies (LEAs).

Innovations addressed Competitive Priority 5, “improve early learning opportunities,” by working to improve teaching effectiveness in grades PK-3.

All applications for i3 grants were required to focus on high-need students. Entities that were eligible for an i3 grant included: (1) LEAs; or (2) nonprofit organizations in partnership with (a) one or more LEAs, or (b) a consortium of schools. Erikson’s Innovations project partnered with the Chicago Public Schools, meeting the second criterion for eligibility.

To leverage the i3 funds and ensure the creation of additional community linkages and buy-in, the Department proposed a fund matching requirement for applicants: entities applying for a grant had to obtain matching funds or in-kind donations of at least 20% of the federal grant award amount. The Collaborative’s $5 million grant was augmented by a $1 million grant from CME Group Foundation (http://www.cmegroupfoundation.org/).

The i3 applicants were also required to include plans for an independent evaluation funded by their grant. The Collaborative selected SRI International to be the external evaluator for the project.

Rationale for the Innovations in Early Mathematics Project

The Collaborative’s i3 project was designed to address three significant problems confronting mathematics education in Illinois: (1) the mathematics achievement gap between low-income, minority students and their more advantaged peers; (2) the low level of competence in math teaching of early childhood and early elementary teachers; and (3) the limited and ineffective in-service opportunities for early childhood and elementary teachers to develop mathematics competencies.

Achievement Gap in Mathematics. Mathematics knowledge is vital to the economic well-being of society and the practical life success of its members. Although the Illinois Learning Standards for Mathematics (Illinois State Board of Education, 2002)—the state standards in place at the time the Innovations project was proposed—are based on research and incorporate recommendations of the National Council of Teachers of Mathematics (NCTM; National Council of Teachers of Mathematics, 2000), NAEP results from 2009 showed that Illinois had one of the five largest math achievement gaps in the United States (Kerachsky, 2009). National data showed this gap appearing as early as kindergarten and widening during elementary school (Flanagan, McPhee, & Mulligan, 2009), indicating the importance of early intervention (National Research Council, 2009). Although mathematics intervention during preschool can be effective (e.g., McCray & Chen, 2012), gains generally fade out during primary school years if not supported by continued excellent and aligned instruction (Heckman & Masterov, 2004), suggesting the importance of high-quality teaching well into elementary school. Further, large-scale studies show that early mathematics achievement predicts not only later mathematics achievement but also later reading achievement, suggesting the broad influence of early math across school subjects (Claessens, Duncan, & Engel, 2009).

Low Level of Math Competence of PK-3 Teachers. Despite emphasis on benchmarks and performance standards, research indicates that PK-3 teachers lack the content knowledge to make use of these pedagogical aids (Ginsburg, Lee, & Boyd, 2008; National Research Council, 2009). In Illinois, early childhood teacher certification for PK-3 is general; early childhood teachers typically acquire less math knowledge than teachers at higher grade levels in pre-service training. Because teachers tend to teach what they know (Darling-Hammond & Bransford, 2005; Sarama & DiBiase, 2004), this limited preparation in mathematics leads to inadequate and ineffective classroom practice. A 2008 study of Chicago early childhood classrooms found that 90% of teachers conducted literacy-related activities daily, but only 21% carried out mathematics activities (Chicago Program Evaluation Project, 2008).

Limited and Ineffective In-service PD for PK-3 Teachers. The PD in math available to early childhood teachers is very limited, and largely ineffective (National Research Council, 2009). Training related to learning standards mainly focuses on introducing the standards rather than helping teachers learn how to use them. Curriculum-specific training tends to be similarly shallow, emphasizing activities and neglecting deep conceptual understanding. Further, the dominant model of in-service PD is a 1-day workshop; continuing support is rarely available for teacher implementation of new practices (Sarama & DiBiase, 2004). Lastly, PD effectiveness is limited by its lack of a conceptual framework that specifies dimensions of teacher change and guides program design (Chen & Chang, 2006).

The Innovations project, then, proposed a PD model to address this lack of high-quality math-focused training for PK-3 teachers. By bringing to bear two unique caches of thinking—the Collaborative's *Big Ideas of Early Mathematics* combined with Erikson Institute's original framework for professional development, the *Whole Teacher Approach*—the Collaborative hoped to improve math teaching practices within schools in a profound and sustainable way, ensuring that more high needs students would meet or exceed state learning standards.

Innovations in Early Mathematics: Professional Development Intervention for PreK to Third-Grade Teachers

Conceptually, Innovations was grounded in two bodies of thinking developed at Erikson Institute: the *Big Ideas of Early Mathematics* and the *Whole Teacher Approach*. The understanding of Innovations' designers was that early childhood and elementary teachers were both substantially underprepared in early math and generally underconfident about teaching it. For this reason, effective PD had to address math content in great depth, and it had to bring a unique, wholistic approach to adult learning that could overcome years of math anxiety and uncertainty.

The Big Ideas of Early Mathematics

In 2014, the Collaborative published a book with Pearson, *The Big Ideas of Early Mathematics: What Teachers of Young Children Need to Know* (Early Math Collaborative). The book's content was based on years of study, development, and testing that began in 2007 at Erikson Institute. Collaborative faculty and senior program developers undertook extensive study of cognitive developmental research findings with implications for how young children's mathematical thinking develops. As former early childhood teachers and experienced facilitators of adult learning, the Collaborative's staff brought a unique educational and applied perspective to these topics. Together, this author team developed 26 Big Ideas and used them extensively in professional development with preschool and kindergarten teachers over a period of 5 years. The significant shifts their work enabled in both teaching practices and child outcomes were a key piece of evidence that convinced the Department to award the grant that funded Innovations, and the team, in turn, was convinced of the power of the *Big Ideas* for informing effective PD. For Innovations, the *Big Ideas* were expanded upward to include content knowledge relevant for math teaching in Grades 1–3.

The *Big Ideas of Early Mathematics* provided Innovations with a sorely needed, curriculum-agnostic focus on math content. Essentially, they represent a distillation of key pedagogical content knowledge relevant for math teaching during early childhood and early elementary schooling. The idea, for example, that "quantity is an attribute of a set, and we use numbers to name specific quantities" is one such *Big Idea*, generative of future math thinking, and appropriately central to early childhood math-related teaching (see Appendix A for a sample Big

Idea). When teaching is built around such an idea, it addresses key understandings children need opportunities to explore and discover, creating rich, connected, conceptual knowledge that provides a foundation for children's later math learning.

Through the work with Innovations, the Collaborative has identified several dozen such ideas, organized by topic areas, and these provided a content focus for each stage of our ongoing PD. During the first year of our work, for example, we focused exclusively on number and operations, covering number sense, counting and cardinality, addition, subtraction, grouping (multiplication), algorithms, and fractions. During a given month, each piece of the intervention incorporated one such set of content, connecting it to existing curricula-based work at the schools. For example, when the Big Ideas of counting and cardinality were the topic, our learning labs included adult learning experiences with non-base 10 counting, while our individualized coaching focused on strategies for using a rekenrek counting frame for classroom attendance. By organizing our PD around the *Big Ideas*, teachers were "marinated" in a few key pieces of pedagogical content knowledge at a time, seeing their implications both for their own understanding and for their mathematics teaching. This was a means of ensuring teachers gained a deep content understanding that would allow them to implement curricula with confidence and in ways that would improve students' chances of meeting and exceeding the standards.

The Whole Teacher Approach

The other body of thinking that was central to the design of *Innovations* was the *Whole Teacher Approach*, a conceptual framework developed by researchers at Erikson Institute, including Innovations' Principal Investigator, Dr. Jie-Qi Chen (Chen & Chang, 2006a; 2006b). Analogous to the "whole child" approach in early childhood education, which posits that teachers of young children must simultaneously address cognitive, physical, and social emotional development, the *Whole Teacher Approach* suggests that effective PD is similarly integrative. That is, in order to make real changes in teaching, PD must not only address gaps in teacher knowledge, but must be designed to consciously involve teaching practice in learning, and to address teachers' attitudes about teaching and content throughout implementation (see Appendix B). This certainty that the best PD addresses knowledge, practice, and attitudes in synergistic ways offered a conceptual framework for the effective design of professional development (Chen & McCray, 2012), and guided not only the intervention, but also the evaluation of the effects of *Innovations* on teacher and child outcomes (Blank, de las Alas, & Smith, 2008).

Through experiences utilizing the *Whole Teacher Approach* in the design and implementation of PD over several years, by 2010, the Collaborative team identified at least two levels on which it has specific types of implications. First, the Approach suggests the importance of a comprehensive set of different types of intervention inputs used in combination. For example, if addressing knowledge is important, then there must be some way to help teachers engage directly with math content as learners: Innovations includes learning labs (a workshop-like setting) for this purpose. If utilizing practice settings is also crucial, this implies the need for coaching, or other school-based services that allow PD facilitators to recruit classroom teaching itself in the service of educating the teacher: Innovations included both individual coaching and coach-facilitated grade-level meetings. The variety of intervention inputs, in combination, allowed Innovations to thoroughly implement the knowledge-practice-attitude framework of the *Whole Teacher Approach.*

The *Approach*, however, also lends itself to the development of key principles for adult education that have implications for every point of contact between PD facilitators and teachers. For Innovations, these included:

- Use parallel processing;
- Establish a learning community;
- Analyze video of children and classrooms; and
- Model reflective practice.

USE PARALLEL PROCESSING

Teaching teachers is essentially a cognitively layered task: the teacher must take her own experience of the PD and somehow transform it into a learning experience s/he will facilitate for others. For this reason, as PD providers, it is tempting to simply “tell” or “show” the teacher how to do her job. Unfortunately, this “simple solution” will seriously neglect the development of her own understanding of the content, a particularly devastating effect in mathematics, where better content knowledge is so urgently needed. Guided by the tenet of the Whole Teacher Approach that knowledge development among teachers is key, the Innovations project committed to providing rich, real, adult learning experiences, related to the mathematical pedagogical content understandings teachers needed.

To accomplish this, the developers presented our adult learners with authentic learning tasks related to a *Big Idea* they might explore with children. Such adult learning activities were set at a high enough level of complexity that teachers experienced the cognitive dissonance of a novice learner and were therefore thoroughly engaged with the content. This is a distinct experience from asking teachers to simply enact an activity they might use with the children in their classroom, or telling them what to do, since it engages them in true struggle and problem-solving related to the content they will eventually teach. This engagement of adult learners in constructing their own understanding *parallels* the kinds of experiences they should be providing for children, rather than simply modeling or mimicking them. It takes more time, since it must be followed by analysis and discussion that links it to children's developmental trajectories as well as to teaching practices, but the developers believed this approach is ultimately more effective.

ESTABLISH A LEARNING COMMUNITY

It is important to establish a *learning community*, meaning a PD classroom with shared understandings that create a safe, open learning environment, even if the group is only temporary. Explicit efforts can be made to create an environment in which adult learners feel safe to share their mathematical ideas honestly and openly—a particularly important notion, since many early childhood education teachers lack confidence in math. The aim is that by doing math together in a new way, participants can begin to repair past humiliations, failures, and fears. There are many small choices that, when combined, help to establish such a learning community.

For example, it can help to set group norms, such as "be present," and to define them clearly, as in "if you must take a phone call, please do so in the hallway so the rest of us can continue to focus on our shared learning experiences." Also key to building a positive group identity are name tags and icebreaker activities, since they fulfill participants' need to feel known. It can also be useful to be strategic about seating—sometimes it helps to sit next to someone new; sometimes the planned activities are better if participants are grouped by their roles or experience.

Such elements, though they seem simple, are crucial to creating a space in which deep learning happens. Through explicitly identifying these elements and honoring their implementation with rigor, the developers create a setting that allows the kind of teaching that eschews answer-getting for building understanding and making sense of the world through math. And given our emphasis on parallel processing (see above) it is important that teachers have such an experience of *how we do math*, so they know how to value it and are willing to do what is necessary to recreate the experience for their own students. Without these kinds of efforts, the developers believe it is not possible to shift teachers' attitudes about math and math teaching, a key element of the *Whole Teacher Approach.*

ANALYZE VIDEO OF CHILDREN AND CLASSROOMS

Since adults often gather to learn in settings away from young children, *analyzing video of children and classrooms* is crucial to bringing learning and teaching to life for participants. Examining authentic video vignettes allows participants to apply new understandings of what math learning and teaching look and sound like. Video can help teachers become better observers of children, studying what they say and do to improve their analysis of what is understood and meant. This is crucial to practice, when decisions must be made on the fly about when to pursue a question further and when to move on to the next point, for example.

When the videos are focused on teacher-led activities or teacher-child interactions, this has the further advantage of making teaching practice public, allowing the group, as professionals, to analyze, discuss, and debate the myriad of teaching choices available in any educational moment. As are the practices of medicine and law, teaching is a complicated art, enacted in the moment and in response to a constant inflow of new information. Video allows us to capture practice and study it in a way no other medium can do.

Further, it can boost the creation of routines of continuous improvement, something sorely missing from most school environments, by providing ways to come together and analyze practice with colleagues. Most teachers have very limited experiences using video of teaching in this way, so they need guided experiences in its use. Innovations used video in each of its PD settings, giving teachers multiple opportunities to learn about its strengths for helping them improve their teaching practice.

MODEL REFLECTIVE PRACTICE

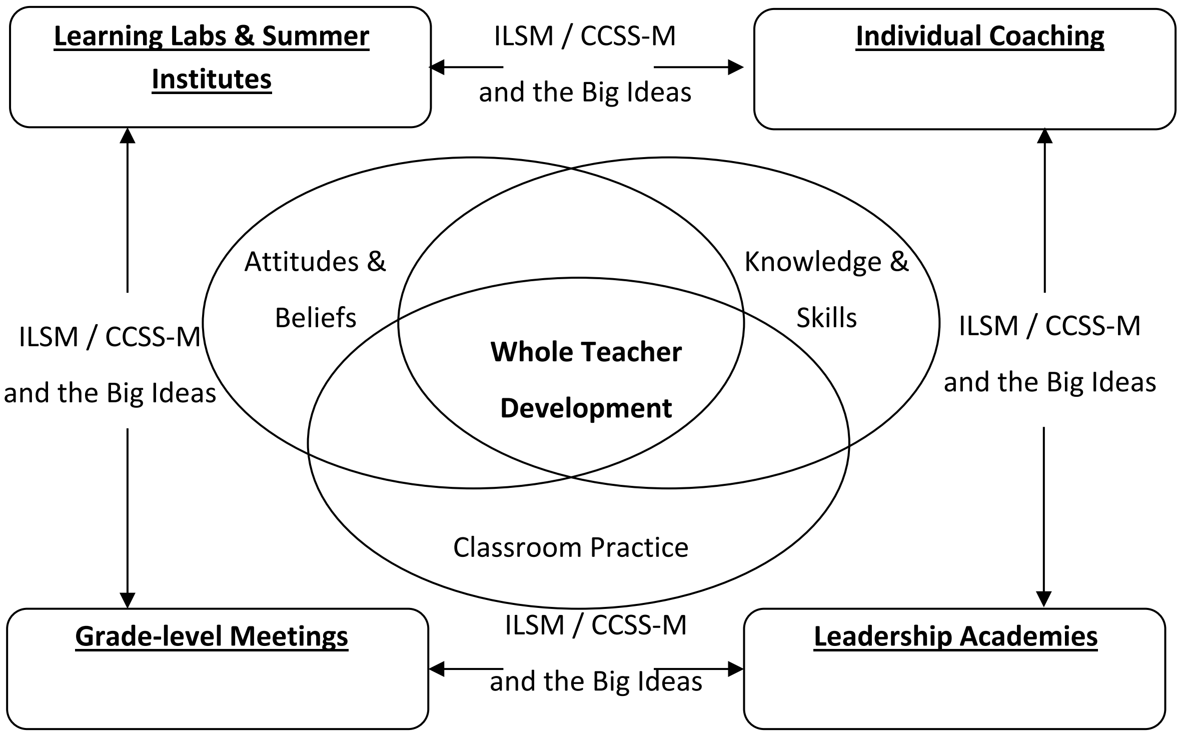
Effective teacher education and professional development require that facilitators of the adult learning experience model reflective practice. This means that developers set goals for teachers' learning, plan how and when to collect data on our own work (this may include work samples or activity photos, as well written reflections from participants), and collaborate with colleagues to analyze our work and make corrections and improvements. It is also crucial to thoroughly plan *the how* as well as *the what* for the entire session; such planning makes reflection much more useful and precise, since it is easier to assess whether goals were met when they are clearly and explicitly set. It is also sometimes useful to make instructional choices transparent to participants, especially since developers are working with teachers and talking about teaching. These elements combined not only improve our facilitation dramatically but are also a powerful incentive to the establishment of a learning community, since they help to create a safe and predictable environment.

Description of the Intervention Components

The intervention was designed to be ongoing and comprehensive, and included five specific components: learning labs, summer institutes, individual coaching, grade-level meetings, and leadership academies. This combination of activities allowed for the following key activities: in-depth study of the mathematics content by teachers and administrators; supported, on-site attention to individual teaching practices; guided collaboration among teachers at school sites; and a focus on the management and administrative techniques that would support changes to teaching. Implementation involved all PK-3 teachers at eight Chicago Public Schools and took place over a 4-year period, with the first 2 years meant to create shifts in teaching practices, and the second 2 years focused on sustainability. School administrators and specialist teachers were also involved, in an attempt to create school-wide change in math teaching that was universally understood and supported.

Exhibit 1 shows the relationship between the conceptual framework that guided Innovations and the program components. The Whole Teacher Approach, its ideas about addressing and utilizing teacher knowledge, practice, and attitudes in an integrated fashion, are at the core, guiding every piece of the implementation. The intervention inputs—learning labs/summer institutes, individualized coaching, grade-level meetings, and leadership academies—anchor the intervention, and are connected by content, specifically the standards and the Big Ideas.

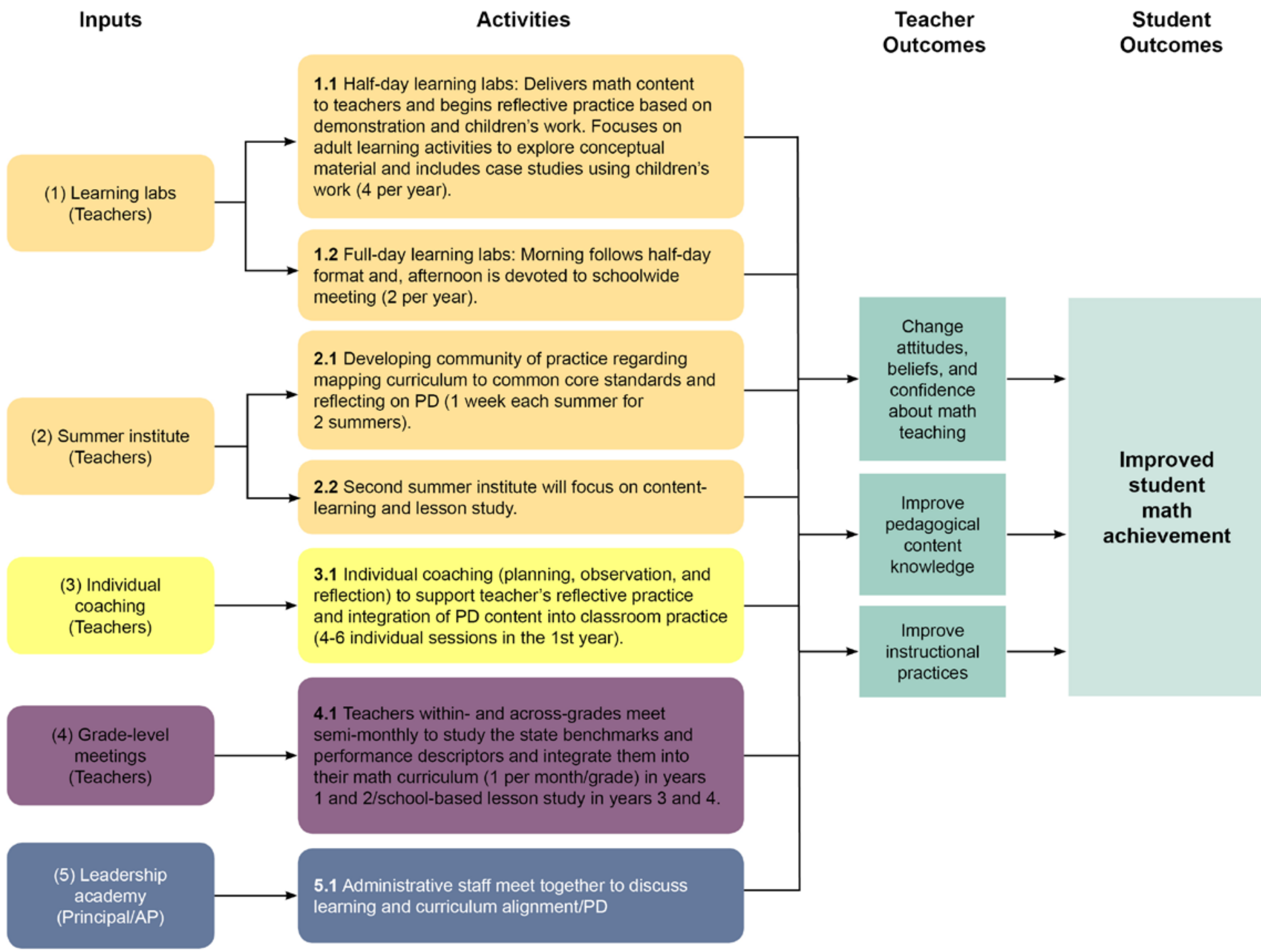
Exhibit 1. Program Components and Conceptual Framework



The logic model in Exhibit 2 illustrates how the developers expected the intervention to work. Intervention inputs are linked to specific activities, which generate improvements in the three teacher outcomes—math teaching-related knowledge, practices, and attitudes—and these, in turn, result in improved student math achievement. The inputs and activities hypothesized to impact the teacher outcomes are described more fully below.

Operationally, the program included five components: learning labs, summer institutes, individual coaching, grade-level meetings, and leadership academies. Learning labs and summer institutes are sometimes thought of as one component, as in Exhibit 1 above, since their delivery and purposes were extremely similar, but they are examined separately in terms of the Implementation Study presented below. Learning labs, summer institutes, and coaching had been developed and used previously as intervention inputs with more than 250 preK and K teachers in the Chicago Public Schools (CPS) between 2007 and 2010. The last two PD components, the grade-level meetings and leadership academies, were newly developed to attempt to take advantage of the school-wide nature of the Innovations project. Together, the components constituted varied learning contexts that promoted a range of social interaction and study designed to activate the Whole Teacher Approach.

Exhibit 2. Logic Model for Innovations in Early Mathematics



Learning labs: Held at Erikson Institute, these were interactive learning sessions in which PK-3 teachers gained knowledge of the math content and processes in the ILSM/CCSS-M and focused on developing a deep understanding of the *Big Ideas*. In *learning labs*, teachers from all participating schools met together with Innovations instructors outside of work time, forming connections across school cultures. Teachers received a small stipend for participating in these sessions, which were either a half day after school or a full day on a Saturday. All of our learning labs had a similar structure:

(1) *Adult Learning Activity: Greeting*—Short mixer activity centered around a mathematical question or task to:

- Break the ice and begin to build learning community
- Turn on brain and warm up "math muscles"

(2) *Adult Learning Activity: Investigation*—A mathematical problem task for adult learners (not like high school or college classes, but also not adults pretending to be children) to:

- Engage adults in math for math's sake
- Have fun doing math
- Activate prior mathematical knowledge
- Build new mathematical understandings

(3) *What are the Big Ideas?*—A discussion of the math concepts underlying the adult learning activities to:

- Put math from adult learning activities into the framework of the Big Ideas of Early Mathematics
- Consider what those Big Ideas mean in other contexts

(4) *What does children's learning look like?*—An examination of the typical learning progression for the math concepts, usually with video examples to:

- Consider what words and actions are evidence of children's math thinking
- Put math ideas into broader context of children's development
- Connect the content and Big Ideas to the standards

(5) *Implications for Teaching*—A conversation about what all these ideas mean for doing math with children in the classroom, usually with video examples or case studies to:

- Analyze teachers' instructional decisions
- Discuss how teacher words and actions foster children's thinking

- Consider connections between math-focused activities and other learning opportunities, especially how good books can be good starting points for doing math
- Put math into the context of real life in a community of young children

The labs were supplemented by full-day sessions offered as **summer institutes** during the June–August period. Summer institutes followed the same format as the learning labs offered during the year.

Individual coaching: Between learning labs, teachers worked with a math coach in their classrooms to plan and reflect on their teaching. Innovations' coaching relationships typically began with a focus on using Erikson's three-stage cognitive coaching cycle, which involves a planning meeting, an observation, and a reflection meeting. Through this technique, coaches help teachers to establish their own goals for improving their practice. Based on these goals, teacher and coach review a lesson plan and determine what the coach will be looking for during a given observation. A detailed data collection plan is developed and utilized by the coach during the observation. These techniques are meant to shift control of the coaching process to the teacher: the coach is there to facilitate development of the teacher's practice and to support her own efforts in this regard. Whenever possible, video was also gathered, so that teacher and coach could study the teaching that occurred together.

Later in the coaching relationship, other techniques were utilized, such as strategizing together how to mathematize classroom routines. Sometimes coaches co-taught lessons, or modelled their implementation and followed this with discussion. Briefer, less formal coaching visits were also used as relationships solidified. Through these means, c*oaching* focused on the individual needs of teachers for learning and development and brought shifts in teaching practice into classrooms, turning the school into a setting for continuous improvement and bringing the "practice" emphasis of the Whole Teacher Approach to the fore.

Grade-level meetings: At schools, teachers within (and sometimes across) grades met semi-monthly with their coach to study the standards and the Big Ideas and integrate them into their math curriculum. Using classroom videos and other means, these *school-based learning groups* became communities of learners, in which co-construction of knowledge, shared learning, and collaboration were the norm (National Research Council, 2009). They provided a venue for planning among teachers from the same grade, discussion of curriculum alignment between grades, and group analysis of student work. These meetings were envisioned as important vehicles for the sustainability of the intervention, since it was hoped teachers would learn how to work together to improve practice on an ongoing basis. During the last 2 years of implementation, teachers learned to utilize a group protocol for looking at student work together (see Appendix C).

Leadership academies: School administrators met quarterly to learn about the math their teachers were studying, share ideas with peers about instructional leadership for change, and become better prepared to observe and support excellent math teaching. In *leadership academies*, the PD team harnessed the school structure, enabling principals to support the shifts in teaching required by the work.

Through these intervention inputs, Innovations utilized beyond-school, within-school, and individualized learning experiences to simultaneously promote development of math teaching-related *attitudes, practices, and knowledge,* connecting each element of the PD experience through the *Big Ideas* content.

Teachers attend an Innovations "Kick-off" Meeting at Erikson Institute, learning about the Whole Teacher Approach, the intervention components, and the Big Ideas of Early Mathematics.

Initial Preparation and Implementation Activities

Below we describe the initial preparation activities in Year 1 that included hiring and training coaches, recruiting schools, and developing and extending PD materials. The instructor team held focus groups on the math curricula used at each school and met with district math specialists to identify issues of concern and enhance training plans. New written materials to cover the 1−3 grade span and supplement existing materials focused on preK and kindergarten were developed and piloted in short-term PD sessions with other teachers. Videotapes of students and classroom teaching were created for PD use and also piloted. Eight coaches were hired and trained in both our coaching model and in our early mathematics *Big Ideas* PD curriculum. Administrative staff prepared enrollment and registration materials, and developed an online system for recording and reporting on intervention activities delivered at schools by coaches. Finally, a kickoff meeting was held with all teachers and administrators to orient them to the program and its evaluation.

Implementation of Professional Development to Teachers and Principals

On average, 110 lead and 40 assistant teachers participated in at least one PD component each year of the project, except for the last year when one intervention school was closed by the district. Therefore, teacher numbers vary due to staff turnover and missing data, as well as the addition of new teachers each year. Below we describe how each of the PD components shown in Exhibit 2 were implemented in the participating schools across 4 years and what was learned throughout the process of implementation.

YEAR 1 OF IMPLEMENTATION: 2011–12

During the 2nd year of our project (or Year 1 of implementation), the project launched both intervention and evaluation components. A total of 150 teachers participated in learning labs, and of them, 110 had major responsibility for a classroom and so received individualized coaching as well.

Learning Labs and Summer Institutes (Year 1)

Two full-day and four half-day learning labs were held at Erikson Institute. At these events, all PK-3 teachers from the eight intervention schools were invited to work with our senior instructors and coaches. Through adult learning activities, video analysis, hands-on practice with math models and materials, and collaborative efforts in problem solving and curriculum design, teachers studied math content. During the 2011–12 school year, the labs focused exclusively on number and operations, and included number sense, counting and cardinality, addition, subtraction, grouping (multiplication), and algorithms (see Appendix D).

At an early learning lab, teachers work together to count a large set of beads, coming up with their own methods for grouping and tracking items counted. Later, they will use these beads to construct "rekenreks"—small counting frames that emphasize fives, tens, and number combinations up to 20—for use in their classrooms.

For 4 full days in early summer of 2012, teachers attended labs on their own time to study fractions (see Appendix D). Also during this time, teachers studied the Common Core State Standards for Mathematics, focusing on the practice standards. In collaboration with Erikson staff, the teachers developed the "Standards for Mathematical Practice Toolkit," a document that summarizes each of the eight practice standards, describing the ways that PK-3 children might demonstrate proficiency in each one. It also specifies those things that teachers can do to support students' proficiency in each practice in their classroom (see Appendix E).

Coaching (Year 1)

Each of the eight schools was assigned a team of two or three coaches. During Year 1, coaches emphasized our cognitive coaching model, completing three to six complete cycles with each teacher in grades PK-3. Each cycle consists of a planning conversation, the observation of a math teaching lesson, and a reflecting conversation. Whenever possible, math lessons were videotaped so coaches and teachers could review them together.

Grade-level Meetings (Year 1)

Also at the schools, coaches led monthly grade-level meetings focused on math teaching. At these meetings, teachers conducted curriculum analysis, planned lessons together, and studied student work.

Leadership Academy Meetings (Year 1)

Our program supervisor convened four half-day meetings of the principals and assistant principals of the eight schools. At these meetings, principals received program updates about teacher participation and services delivered at their school, studied some of the mathematics their teachers were learning, and worked to develop observation skills for supporting teachers' efforts to enhance their math teaching.

Additional Activities to Support School-based Learning (Year 1)

Throughout the project, additional activities were sometimes developed in response to direct requests from schools and teachers for the implementation team's help. We describe those activities here, but they are not included as part of the main intervention components when describing fidelity of implementation. In Year 1 of implementation, the only such activity was a **Family Math Night.** Each school held an event for families, making them aware of our project work and inviting them to participate in their children's math learning. At these events, teachers led parents and children through joint activities they had designed in coordination with coaches.

Addressing Challenges (Year 1)

Because a greater number of teachers enrolled in the project than anticipated, the developers increased the number of coaches. Training efforts for these new coaches were extensive. There also was one school that took on additional literacy-focused PD from an outside source during our intervention timeframe. This was something the developers had hoped to avoid, so that teachers would have enough time and energy to participate fully in our intensive intervention. Through negotiations and with assistance at the network level of CPS, the developers were able to work out a plan for this school's involvement. The plan appeared to work, since there was marked positive change in math teaching at this school.

> *I've become a lot more thoughtful in my practice, and my understanding of what children can do and how they can think mathematically has just been opened. I think before we really focused on counting, patterns, maybe some shapes, but it was very basic, and we really didn't get the kids thinking. After working with Erikson, we started to look at numbers in a much more in-depth way. Instead of focusing so much on can they count, it's been more do they even understand what the number 1, 2, up to 5 and 6, what that actually means? So, I find that my practice has become a lot more thoughtful.*

YEAR 2 OF IMPLEMENTATION: 2012–13

During the 3rd year (Year 2 of implementation) of our project, the developers continued our intervention and evaluation efforts. A total of 146 teachers attended learning labs, and of them 105 received individualized coaching.

Learning Labs (Year 2)

One full-day and five half-day learning labs were held at Erikson Institute during the school year, followed by a 1-day "summer spectacular" in August, in preparation for the year to come. Additionally, 2 full days of induction activities were provided in the fall for new teachers. During the 2012–13 school year, the labs focused on geometry, measurement, and data analysis, building on the intensive study of number and operations completed the year before. Throughout the year, emphasis was placed on incorporating the Common Core State Standards for Mathematics, focusing especially on the practice standards. See Appendix D for more detail on the content covered in the learning labs in the 2nd year.

Coaching (Year 2)

During the 2012–13 school year, coaching continued as before, but the developers implemented the use of the dimensions of our observation tool (High-Impact Strategies for Early Mathematics) as a format for setting goals for each teacher. This format provided specific questions that could be addressed during coaching, such as "to what degree does this lesson evoke a positive mathematical learning community?" or "to what extent am I ensuring that most of the math during this lesson is done by the children as opposed to by me?" Coaches reported the use of the dimensions helped increase the specificity and effectiveness of their individualized work with teachers.

> *The first year of Erikson, it was actually my first year of teaching, so I kind of just jumped into the program, like put everything that I thought about math behind and just jumped into their strategies and everything they were teaching. And for me, it was like it made me comfortable teaching math, because I hated it. I was like, "I don't know what I'm doing," and it just made me comfortable and made me confident to teach the kids.*

Grade-level Meetings (Year 2)

Also at the schools, coaches continued to lead monthly grade-level meetings focused on math teaching. Where possible these were expanded to include cross-grade-level meetings focused on between-grade alignment of curricula.

Leadership Academy Meetings (Year 2)

The program supervisor convened four half-day meetings of the principals and assistant principals of the eight schools, continuing the work that had begun with them the previous year. At the beginning of this year, the developers shared data on services delivered and began to overview our outcome findings with principals. These efforts helped principals to focus their efforts around supporting the work and increase their sense of urgency about the intervention.

Additional Activities to Support School-based Learning (Year 2)

Again, additional activities were developed in response to direct requests from schools and teachers for the implementation team's assistance. These activities are described below, but are not included as part of the main intervention components when describing fidelity of implementation.

- **Family Math Nights.** As in the previous year, each school held an event for families, making them aware of our project work and inviting them to participate in their children's math learning. At these events, teachers led parents and children through joint activities they had designed in coordination with our coaches.
- **"Share Fairs."** At each school, coaches helped teachers organize an event during which teachers shared the new types of teaching practice successfully implemented with their

students. At some schools, this meant that the PK-3 teachers presented to the entire faculty (up to eighth grade), while at others, the event was limited to the PK-3 teachers. At these events, teachers used manipulatives, posters, and examples of student work to describe changes in their math teaching during the school year.

- **Preparation for Lesson Study.** Interested schools identified three to five teachers to participate as a "Lesson Study" team. These teachers attended a Lesson Study Conference in Chicago along with our senior instructors and coaches. At the conference, teachers were introduced to the Lesson Study protocol, and experienced the teaching and group discussion of a "public lesson." During the summer, each school's Lesson Study team attended a week-long intensive workshop to begin their lesson planning for a public lesson to be held at their own school. Lesson Study experts from the Lesson Study Alliance provided supervision for these activities, and teams from four of our schools developed beginning lesson plans. These activities comprised another effort to generate school-based leadership among teachers, so they can continue to work together to enhance their own professional development after our intervention is over.

Addressing Challenges (Year 2)

The biggest challenge the developers faced during Year 2 was the potential closing of three of our intervention and one of our comparison schools. Our LEA, the Chicago Public Schools, generated a list of more than 150 schools they felt should be closed due to underutilization, and four of our schools were on that list. During the ensuing public discussion, the PD team worked hard to find appropriate means to ensure that decisionmakers and the public understood the work engaged in at these schools, and the disastrous effect that closure would have, not only on the intervention but also on our understanding of its evaluation results. Erikson's president issued a letter describing our i3 initiative, and members of our team attended many public meetings to represent our project. The intervention schools were eventually removed from the list of potential closures, and the CPS superintendent credited the decision, in part, to her awareness of the amount of effort and money that had already gone into these schools' improvement. Unfortunately, the decision was made to close one of our comparison schools after the 2012−13 school year was completed. Since Year 2 was the last year of data collection for our child-level impact analysis, the project was still able to gather the child data. However, the distress experienced at all four of these schools, by administrators, faculty, and students, may have had a negative impact on the effect of the intervention in the schools and on the teachers.

YEAR 3 OF IMPLEMENTATION: 2013–14

During the 4th year (Year 3 of implementation) of the project, the team continued our intervention and evaluation efforts. A total of 153 teachers attended at least one learning lab, and of those, 128 received coaching.

Learning Labs (Year 3)

As in previous years, the developers provided 2 days of induction in the fall for teachers who were new to the intervention schools during the 2013–14 school year. Additionally, one half-day learning lab was held at Erikson Institute during the school year, followed by a half-day event at the school year's end featuring poster presentations by teacher groups, entitled "Shifts Happen." At this event, teachers documented changes they had made in their own teaching practices and presented artifacts of this work to one another. Guests from the Chicago Public Schools, our i3 matching grant funder, the CME Group Foundation, and Erikson's Board of Trustees also attended, and a keynote address was given by Professor William Ayers of the University of Illinois at Chicago. Finally, a full-day event to begin the new school year was held in August 2014 for interested teachers.

Teachers at a learning lab work together to prepare a poster explaining their thinking about patterns to the larger group.

Coaching (Year 3)

During the 2013–14 school year, the developers began to tailor our coaching intervention more to each teacher-coach relationship. In some instances, this meant the use of more frequent, less formal coach visits of shorter duration, called "pop-ins." The developers continued to use the dimensions of our observation tool (High-Impact Strategies for Early Mathematics) as a format for setting goals for each teacher.

Grade-level Meetings (Year 3)

During Year 3, grade-level meetings took on new significance in the project. The developers introduced a structured protocol for the shared study of student work (see Appendix C). Grade-level groups chose and implemented common math assessment tasks, spending time together anticipating student responses. Teachers brought selected samples of work from their students to the following meeting and used our protocol to analyze them. These meetings, focused on task analysis and selection and detailed examination of students' responses, became powerful mechanisms for helping teacher groups become small professional learning communities, focused on improving their own mathematical knowledge and math teaching practice. It was our hope that these experiences would enhance sustainability of the intervention at the schools.

Leadership Academy Meetings (Year 3)

Our program and coach supervisors convened four half-day meetings of the principals and assistant principals of the eight schools, continuing the work that had begun with them the previous year.

Additional Activities to Support School-based Learning (Year 3)

Again, additional activities were conducted in response to direct requests from schools and teachers. These activities are described below, but are not included as part of the main intervention components when describing fidelity of implementation.

- **Family Math Nights.** Each school held an event for families, making them aware of our project work and inviting them to participate in their children's math learning. At these events, teachers led parents and children through joint activities they had designed in coordination with our coaches.
- **Lesson Study Activities.** In the summer of 2013, four of our eight Innovations schools had sent a "Lesson Study Team" of three to five teachers to a week-long intensive workshop to begin their planning for a public lesson to be held at their own school. Lesson Study experts from the Lesson Study Alliance provided supervision for these activities. During the school year, each of these teams planned and conducted a lesson, which was observed by other teachers and discussed afterward with support from our coaches and instructors. At each school, plans were made to continue all or some of the Lesson Study activities during Year 4–for example, by joint planning and the use of videotape rather than live observation.

Addressing Challenges (Year 3)

The most significant challenge the developers faced during this year was continuing to enhance mechanisms for sustainability at each of the schools. In some cases, and at some schools, a strong group of teachers made it easy to implement and support ongoing rigorous study of math teaching at grade-level meetings. At other schools, it was more difficult to develop a critical mass of teachers with enthusiasm for trying new things and pushing their teaching practice to new levels. To address these challenges, the PD team made unique arrangements at each school. For example, at some sites, the developers combined grade-level groups, creating a Grades 1 to 2 group to study math, while at others, the developers split a single grade-level group into two groups of three or four teachers each. Coaches and supervisors worked together to develop plans for each situation that were most likely to create thriving and sustainable groups.

YEAR 4 OF IMPLEMENTATION: 2014−15

During the 5th year (Year 4 of implementation) of our project, the developers continued the intervention and evaluation efforts. A total of 98 teachers received classroom-based coaching.

Learning Labs (Year 4)

As in previous years, the PD team provided 2 days of induction in the fall for teachers who were new to the intervention schools during the 2014–15 school year. Additionally, one half-day learning lab was scheduled to be held at Erikson Institute during the school year (this was canceled, see “challenges” below), followed by a half-day event at the school year’s end featuring workshop presentations by teacher groups entitled “Shifts Happen II.” During the workshops, teachers described changes they had made in their own teaching practices, and presented artifacts of this work to one another. School administrators and guests from the Chicago Public Schools, our i3 matching grant funder, the CME Group Foundation, and Erikson’s Board of Trustees also attended.

At "Shifts Happen II" in the spring of 2015, a teacher shares a poster with her colleagues describing how she has changed her math teaching practices.

Coaching (Year 4)

During the 2014–15 school year, the PD team continued tailoring our coaching intervention more to each teacher-coach relationship. In some instances. the developers moved to doing group coaching within a grade level as a way to build sustainability going forward. The developers continued to use the dimensions of our observation tool (High-Impact Strategies for Early Mathematics) as a format for setting goals for each teacher.

Grade-level Meetings (Year 4)

During Year 4 of implementation, monthly grade-level meetings using our structured protocol for the shared study of student work continued.

Leadership Academy Meetings (Year 4)

Our program and coach supervisors convened four half-day meetings of the principals and assistant principals of the eight schools, continuing the work that had begun with them the previous year.

Additional Activities to Support School-based Learning (Year 4)

Additional activities conducted in response to direct requests from schools and teachers but not included as part of the main intervention components analyzed as part of fidelity are described below.

- **Family Math Nights.** Each school held an event for families, making them aware of our project work and inviting them to participate in their children's math learning. At these events, teachers led parents and children through joint activities they had designed in coordination with our coaches.

- **Lesson Study Activities.** At four of our schools, teachers within a grade level regularly planned a lesson together and then observed one of the teachers implement it with his or her students. Other teachers were also invited to observe and participating in the debriefing discussion that followed.

Addressing Challenges (Year 4)

The major challenge for Year 4 of implementation was the fact that three of our eight intervention schools changed principals during the summer and early fall of 2014. One of these three schools decided to discontinue participation in the project due to new requirements that the CPS central office was imposing on the school. The other two schools with new principals continued to be a part of the project, but it was challenging to have these two new administrators be as committed to the project as their predecessors had been. Fortunately, the PD team had established strong relationships with teachers at those schools and worked intensively with these administrators to help them see the value of being part of this project. These efforts made a positive difference, but overall the change in leadership still had a negative effect on degree of implementation school wide. In addition, due to the extreme winter weather in Chicago, the school calendar changed; scheduled PD dates became regular attendance days to make up for snow days. That meant that one of the Innovations learning labs that was scheduled for this school year had to be canceled.

Evaluation Design and Methodology

This section focuses on the design of two studies conducted by SRI International for the i3 requirement to have an independent evaluation: an implementation study to examine fidelity and an impact study to examine effects on both children and teachers. Below we describe each of these studies. Subsequently, detailed descriptions of the sample, methods, analyses, and findings are presented. Finally, we also describe a series of exploratory analyses that were conducted to further examine effects.

Implementation Study

To examine the implementation of the Innovations project, the developers and research team collaboratively developed a priori a matrix to identify fidelity of implementation for each component and then collected data on each component at the teacher- and school-level to measure fidelity. Data were collected on the four major components that targeted teachers (learning labs, summer institutes, coaching, and grade-level meetings) aimed at increasing teachers' math knowledge, beliefs, and practices, as well as data on participation in the leadership academy. The primary research question for the implementation study was: What is the fidelity of implementation of the professional development intervention in the PK-3 intervention classrooms defined primarily as attendance at the learning labs, summer institutes, and grade-level meetings and completion of coaching cycles? We defined fidelity at the teacher level and the school level, as well as at the project level as required by the grant.

Exhibits 3 to 8 show the criteria used for identifying the fidelity of implementation at the individual teacher level, the school level, and at the project level (Exhibits 4−5 for Year 1; Exhibits 6−8 for Year 2). The developer, PD, and evaluation team defined these levels of implementation fidelity a priori to reflect the extent to which the teachers received the PD and in a dosage of sufficient amount hypothesized to lead to changes in teachers' knowledge, beliefs, and instructional practice.

Teacher-level fidelity was defined as the amount, or "dosage," of PD services actually received by each teacher. Based on what was available to teachers across the year, we set criteria for ratings of *low*, *adequate,* and *high* dosage, assuming that an adequate amount would be enough to create significant change in teacher-level outcomes. Each teacher was assigned scores for each PD component; scores were calculated for Years 1 (fall 2011 to spring 2012) and 2 (fall 2012 to spring 2013). We focused fidelity measurement on Years 1 and 2 as the PD was expected to be more individualized in Years 3 and 4 with the hope that the schools would take on more ownership of the learning and meeting in the latter years.[1] Components measured at the teacher level included:

- Learning lab attendance
- Summer institute attendance
- Coaching cycles completed
- Grade-level meetings attended

Each school also received fidelity scores for each intervention component during Years 1 and 2. In general, school-level fidelity scores were based on the percentage of teachers within that school that achieved at least adequate fidelity at the teacher level. Two new components were also added for school-level fidelity: number of grade-level meetings *held* at the school, and number of leadership academy meetings attended by an administrator. For this reason, components measured at the school level included:

- Learning lab attendance
- Coaching cycles completed
- Grade-level meetings, including
  - Held at school
  - Attended by teachers
- Leadership academy attendance

Attendance at learning labs, leadership academy meetings, and the summer institute was tracked through the use of sign-in sheets at the events. Coaching cycles and grade-level meetings, which were held at the school, were tracked by project coaches. Coaches carried tablets, and via Wi-Fi, entered attendance and other data directly into a database as they delivered services.

[1] New teachers were provided with initial PD to promote sustainability of the practices and knowledge at the school. The new teachers then continued with the same PD activities as other teachers.

Exhibit 3. Measuring Fidelity of Implementation for Teachers: Innovations in Early Mathematics (Year 1)

Inputs		Indicators	Sources	Fidelity Scale	Criterion for Adequate/High Fidelity of Implementation
Teacher					
1	Learning labs	Hours attended (across 6 learning labs); ranges from 0 to 22 hours	Attendance logs	0 = 0-5 hours 1 = 6-14 hours 2 = 15-18 hours 3 = 19-22 hours	Low = 0-1 Adequate = 2 High = 3
2	Summer institutes	Hours attended (across 4 learning labs); ranges from 0 to 20 hours	Attendance logs	0 = 0-4 hours 1 = 5-9 hours 2 = 10-14 hours 3 = 15-20 hours	Low = 0-1 Adequate = 2 High = 3
3	Coaching	Number of coaching cycles completed (planning, observation, and reflection); ranges from 0 to 6 cycles	Coaching logs	0 = 1 or fewer coaching cycles 1 = 2 coaching cycles 2 = 3 coaching cycles 3 = 4 or more coaching cycles	Low = 0-1 Adequate = 2 High = 3
4	Grade-level meetings	Number of grade-level meetings attended (up to 6 per year)	Attendance logs	0 = 0-1 meetings 1 = 2-3 meetings 2 = 4-5 meetings 3 = 6 or more meetings	Low = 0-1 Adequate = 2 High = 3

Note. A rating of “adequate” or “high” would be considered meeting the fidelity threshold.

Exhibit 4. Measuring Fidelity of Implementation for Schools: Innovations in Early Mathematics (Year 1)

Inputs		Indicators	Sources	Fidelity Scale	Criterion for Adequate/High Fidelity of Implementation
School					
1	Learning labs	Percentage of teachers who had at least adequate attendance	Attendance logs	0 = 0%-24% of teachers 1 = 25%-69% of teachers 2 = 70%-89% of teachers 3 = 90% or more of teachers	Low = 0-1 Adequate = 2 High = 3
2	Summer institute	Percentage of teachers who had at least adequate attendance	Attendance logs	0 = 0%-24% of teachers 1 = 25%-69% of teachers 2 = 70%-89% of teachers 3 = 90% or more of teachers	Low = 0-1 Adequate = 2 High = 3
3	Coaching	Percentage of teachers who completed 3 or more coaching cycles (planning, observation, and reflection)	Coaching logs	0 = 0%-24% of teachers 1 = 25%-59% of teachers 2 = 60%-89% of teachers 3 = 90% or more of teachers	Low = 0-1 Adequate = 2 High = 3
4	Grade-level meetings	Number of grade-level meetings held (across grades)	Attendance logs	0 = 0-10 meetings 1 = 11-17 meetings 2 = 18-24 meetings 3 = 25 or more meetings	Low = 0-1 Adequate = 2 High = 3
		Percentage of teachers who attended at least 4 grade-level meetings across school year	Attendance logs	0 = 0%-24% of teachers 1 = 25%-59% of teachers 2 = 60%-89% of teachers 3 = 90% or more of teachers	Low = 0-1 Adequate = 2 High = 3
5	Leadership academy	Number of 3-hour meetings attended by school principal or designee; ranges from 0 to 4	Attendance logs	0 1 2 3 4	Low = 0-2 Adequate = 3 High = 4

Note. A rating of “adequate” or “high” would be considered meeting the fidelity threshold.

Exhibit 5. Measuring Fidelity of Implementation for the Project: Innovations in Early Mathematics (Year 1)

Inputs		Indicators	Sources	Fidelity Scale	Criterion for Adequate/High Fidelity of Implementation
Project					
1	Learning labs	Percentage of schools who had at least adequate attendance (scored adequate or high)	Attendance logs	0 = 0-3 schools (0%-38%) 1 = 4-5 schools (50%-63%) 2 = 6-7 schools (75%-88%) 3 = 8 schools (100%)	Low = 0-1 Adequate = 2 High = 3
2	Summer institute	Percentage of schools who had at least adequate attendance (scored adequate or high)	Attendance logs	0 = 0-3 schools (0%-38%) 1 = 4-5 schools (50%-63%) 2 = 6-7 schools (75%-88%) 3 = 8 schools (100%)	Low = 0-1 Adequate = 2 High = 3
3	Coaching	Percentage of schools who had at least adequate level of coaching (scored adequate or high)	Coaching logs	0 = 0-3 schools (0%-38%) 1 = 4-5 schools (50%-63%) 2 = 6-7 schools (75%-88%) 3 = 8 schools (100%)	Low = 0-1 Adequate = 2 High = 3
4	Grade-level meetings	Percentage of schools who had at least adequate level in both areas of grade-level meetings (scored adequate or high)	Attendance logs	0 = 0-3 schools (0%-38%) 1 = 4-5 schools (50%-63%) 2 = 6-7 schools (75%-88%) 3 = 8 schools (100%)	Low = 0-1 Adequate = 2 High = 3
5	Leadership academy	Percentage of schools who had at least adequate attendance (scored adequate or high)	Attendance logs	0 = 0-3 schools (0%-38%) 1 = 4-5 schools (50%-63%) 2 = 6-7 schools (75%-88%) 3 = 8 schools (100%)	Low = 0-1 Adequate = 2 High = 3

Note. A rating of “adequate” or “high” would be considered meeting the fidelity threshold.

Exhibit 6. Measuring Fidelity of Implementation for Teachers: Innovations in Early Mathematics (Year 2)

Inputs		Indicators	Sources	Fidelity Scale	Criterion for Adequate/High Fidelity of Implementation
Teacher					
1	Learning labs	Hours attended (across 4 learning labs); ranges from 0 to 20 hours	Attendance logs	0 = 0-3 hours 1 = 4-12 hours 2 = 13-16 hours 3 = 17-20 hours	Low = 0-1 Adequate = 2 High = 3
2	Summer institute	Hours attended (across 4 learning labs); ranges from 0 to 20 hours	Attendance logs	Canceled due to change in CPS calendar to make up days from Chicago Teachers Union (CTU) strike	N/A
3	Coaching	Number of coaching cycles completed (planning, observation, and reflection); ranges from 0 to 8 cycles	Coaching logs	0 = 2 or fewer coaching cycles 1 = 3 coaching cycles 2 = 4 coaching cycles 3 = 5 or more coaching cycles	Low = 0-1 Adequate = 2 High = 3
4	Grade-level meetings	Number of grade-level meetings attended (up to 10 per year)	Attendance logs	0 = 0-2 meetings 1 = 3-4 meetings 2 = 5-6 meetings 3 = 7 or more meetings	Low = 0-1 Adequate = 2 High = 3

Exhibit 7. Measuring Fidelity of Implementation for Schools: Innovations in Early Mathematics (Year 2)

Inputs		Indicators	Sources	Fidelity Scale	Criterion for Adequate/High Fidelity of Implementation
School					
1	Learning labs	Percentage of teachers who had at least adequate attendance	Attendance logs	0 = 0%-24% of teachers 1 = 25%-69% of teachers 2 = 70%-89% of teachers 3 = 90% or more of teachers	Low = 0-1 Adequate = 2 High = 3
2	Summer institute	Percentage of teachers who had at least adequate attendance	Attendance logs	Canceled due to change in CPS calendar to make up days from CTU strike	N/A
3	Coaching	Percentage of teachers who completed 4 or more coaching cycles (planning, observation, and reflection)	Coaching logs	0 = 0%-24% of teachers 1 = 25%-59% of teachers 2 = 60%-89% of teachers 3 = 90% or more of teachers	Low = 0-1 Adequate = 2 High = 3
4	Grade-level meetings	Number of grade-level meetings held (across grades)	Attendance logs	0 = 0-14 meetings 1 = 15-22 meetings 2 = 23-28 meetings 3 = 29 or more meetings	Low = 0-1 Adequate = 2 High = 3
		Percentage of teachers who attended at least 5 grade-level meetings across school year	Attendance logs	0 = 0%-24% of teachers 1 = 25%-59% of teachers 2 = 60%-89% of teachers 3 = 90% or more of teachers	Low = 0-1 Adequate = 2 High = 3
5	Leadership academy	Number of 3-hour meetings attended by school principal or designee; ranges from 0 to 4	Attendance logs	0 1 2 3 4	Low = 0-2 Adequate = 3 High = 4

Note. A rating of "adequate" or "high" would be considered meeting the fidelity threshold.

Exhibit 8. Measuring Fidelity of Implementation for the Project: Innovations in Early Mathematics (Year 2)

Inputs		Indicators	Sources	Fidelity Scale	Criterion for Adequate/High Fidelity of Implementation
Project					
1	Learning labs	Percentage of schools who had at least adequate attendance (scored adequate or high)	Attendance logs	0 = 0-3 schools (0%-38%) 1 = 4-5 schools (50%-63%) 2 = 6-7 schools (75%-88%) 3 = 8 schools (100%)	Low = 0-1 Adequate = 2 High = 3
2	Summer institute	Percentage of schools who had at least adequate attendance (scored adequate or high)	Attendance logs	Canceled due to change in CPS calendar to make up days from CTU strike	N/A
3	Coaching	Percentage of schools who had at least adequate level of coaching (scored adequate or high)	Coaching logs	0 = 0-3 schools (0%-38%) 1 = 4-5 schools (50%-63%) 2 = 6-7 schools (75%-88%) 3 = 8 schools (100%)	Low = 0-1 Adequate = 2 High = 3
4	Grade-level meetings	Percentage of schools who had at least adequate level in both areas of grade-level meetings (scored adequate or high)	Attendance logs	0 = 0-3 schools (0%-38%) 1 = 4-5 schools (50%-63%) 2 = 6-7 schools (75%-88%) 3 = 8 schools (100%)	Low = 0-1 Adequate = 2 High = 3
5	Leadership academy	Percentage of schools who had at least adequate attendance (scored adequate or high)	Attendance logs	0 = 0-3 schools (0%-38%) 1 = 4-5 schools (50%-63%) 2 = 6-7 schools (75%-88%) 3 = 8 schools (100%)	Low = 0-1 Adequate = 2 High = 3

Note. A rating of “adequate” or “high” would be considered meeting the fidelity threshold.

Impact Study

As described above, the team at Erikson Institute implemented a 4-year professional development (PD) intervention in eight schools in the city of Chicago. The overarching goal of the initiative was to help students in prekindergarten to third grade (PK-3) reach or exceed state learning standards in mathematics by addressing the professional development and training needs of teachers. The project used a quasi-experimental design to examine the following research questions at the child, teacher, and school levels:

(1) What is the impact on teachers' knowledge, beliefs, and practices during each year of the 4-year PD intervention?

(2) What is the impact of the teacher professional development on participating children after teachers participated in 1 year of the intervention (fall 2011 to spring 2012) and 2 years of the intervention (fall 2011 to spring 2013)?[2]

(3) What is the impact of the PD intervention on school-level math achievement during the project?

As described above, the intervention provided four types of PD experiences to teachers (learning labs, summer institutes, coaching, and grade-level meetings), each aimed at increasing *teachers'* math *attitudes,* math classroom *practice,* and math content *knowledge.* Ultimately, these changes in teachers were hypothesized to lead to improved math achievement in children in PK-3 (i.e., improved performance on a standardized assessment of early math abilities as well as a new measure of children's numeracy and geometry). A fifth component (leadership academies) targeted administrators. The 1st year of the intervention occurred in the 2011–12 school year. The intervention lasted 4 years with teachers in the intervention schools receiving professional development throughout the 4-year project period (SY 2011–12 to SY 2014–15). However, the most intensive individual PD components for teachers occurred in the first 2 years; thus, the evaluation team measured impact of child outcomes during these 2 years expecting to see the greatest impact on child outcomes at the end of the 2nd year. During the 3rd and 4th years of the PD intervention, the focus changed to efforts to build capacity at the school level in creating a learning community around teaching and learning math. Exhibit 9 shows the timing of data collection for students in the quasi-experimental design.

[2] The implementation and evaluation teams decided jointly to focus the design of the implementation and impact studies on the first 2 years of the PD intervention because it was believed that the greatest impact on teachers and students would occur during the first 2 years. The last 2 years of the project focused on sustainability of the impacts on teachers' knowledge, beliefs, and practices.

Exhibit 9. Data Collection Schedule for Student Math Achievement Measures

Grade	Pretest (Fall 2011)	1-Year Posttest (Spring 2012)	2-Year Posttest (Spring 2013)
PreK	X	X	X
Kindergarten	X	X	X
First grade	X	X	X
Second grade	X	X	X
Third grade	X	X	- [a]

[a] Because the intervention was focused on teachers preK to third grade, children who were in third grade in the first year of the intervention were not followed into fourth grade for the 2-year posttest assessment.

Based on the logic model (Exhibit 1), it was hypothesized that students (PK-3) in the intervention schools would make greater gains in mathematics after 1 and 2 years relative to students in the comparison schools. Similarly, it was hypothesized that teachers in the intervention schools would demonstrate greater gains on measures of knowledge, beliefs, and practices related to early mathematics in each of the project's 4 years. Finally, it was hypothesized that schools in the intervention condition would demonstrate better school-wide math achievement on a state-administered achievement test in third grade.

Samples: School, Teachers, Children

To ensure that intervention and comparison schools had similar student populations and academic performance at baseline, propensity score matching techniques were used to find comparable schools for the intervention schools. The eight intervention schools were selected by a joint effort of the implementation team and school district administrators to identify schools that were likely to be able to implement the PD intervention and had administrative support to continue the project over the 4-year project period. Intervention schools were selected between January and June 2011. Comparison schools were then selected using propensity score analysis (Stuart, 2010) to find schools that matched on key student population demographics and math achievement test scores in third grade.

There was a pool of 65 nonintervention schools from which to select the eight comparison schools. To start, the evaluation team posited a logistic model to estimate what types of schools were likely to be the intervention schools, using school-level variables, including percentage of third-grade children who met math standards in 2009, percentage of third-grade children who exceeded math standards in 2009, percentage of children who were English language learners (ELL) (2010), percentage of children who were identified as minority (2010), percentage of children receiving free or reduced-price lunch (2010), and mobility rates (2009). Based on the estimated propensity model, the evaluation team calculated a propensity score (logit) of being an intervention school based on the above-mentioned school-level baseline variables. They next selected comparison schools that were closest to each intervention school on the propensity score using nearest neighbor matching. Two of the selected comparison schools dropped out after selection, but prior to pretest data collection. At that time, the organization of schools in the school district had changed such that the three district areas from which the intervention schools were selected were now represented by six networks. Two replacement comparison schools from schools in the six networks were then selected based on their characteristics and whether they were "good-enough" matches to the remaining intervention schools.[3]

SCHOOLS

Assignment to intervention versus comparison conditions occurred at the school level. Exhibit 10 shows the baseline demographic characteristics of the intervention and comparison schools for those variables used for matching the two groups of schools. None of the differences shown represented a greater than .25 standard deviation difference between the two groups of schools.

[3] The evaluators then selected the next best matches that were within the same networks of schools already identified as participants.

Exhibit 10. Baseline Demographic Characteristics and Achievement of Intervention and Comparison Schools

Percentage of students	Intervention	Comparison
Hispanic[a]	57.7	55.3
Black[a]	32.9	23.1
ELL[b]	17.8	32.3
FRPL[b]	91.6	90.9
Mobility[c]	17.3	17.3
Meet/exceed math standards in third grade[a]	70.1	72.4
Exceed math standards in third grade[a]	21.6	24.3
Meet/exceed reading standards in third grade[a]	53.0	54.1
Exceed reading standards in third grade[a]	10.4	11.2

Note. ELL = English language learner; FRPL = free or reduced-price lunch.
[a]Data were obtained from the 2009 State School Report Card.
[b]Data were obtained from the 2010 CPS Office of Performance website.
[c]Data were obtained from the 2009 CPS Office of Performance website.

TEACHERS

Assignment of condition occurred at the school level such that all teachers in the intervention schools participated in the intervention and all teachers in the comparison schools served as the comparison group. Full-time classroom teachers who provided the majority of mathematics instruction in PK-3 general education or integrated classrooms were required to attend learning labs, participate in coaching, and attend monthly grade-level meetings. Other teachers and assistants were encouraged by their principals to participate in learning labs and grade-level meetings but did not participate in coaching and were not included in the implementation or impact studies. A total of 220 teachers participated in baseline data collection (115 intervention and 105 comparison teachers, respectively). Analytic sample sizes vary by outcome and timepoint due to attrition and missing data.

CHILDREN

Assignment of condition occurred at the school level such that all children in the intervention schools received the intervention and all children in the comparison schools served in the comparison group. Parents in all participating teachers' classrooms were asked to provide signed consent for their children to participate in the study, and a random sample of consented children was selected to conduct assessments. Children were included in the study if they were 4 years old by September 1, 2011, were enrolled in the study schools and participating teachers' classrooms, and were able to complete the student assessments in either English or Spanish.[4] Children were excluded if they were unable to complete the assessment in English or Spanish, had an individualized education program (IEP) or a 504 accommodation plan, or were not present on the days the team visited the school at pretest. For children whose parents identified a home language other than English, the assessor asked the teacher to identify whether any of these children at pretest should not be assessed in English. If the child could not be tested in English but could be tested in Spanish, a bilingual assessor conducted the assessment in Spanish (using the Spanish version of the assessment). One-year and 2-year posttests were completed in the same language as pretest for consistency (unless the child insisted on changing the language of the assessment, which occurred a few times).

An estimated 6,000 children were enrolled in PK-3 across the 16 participating schools in 2011–12. Of those, 2,609 children were consented or 43% of total enrollment. The evaluation team aimed to assess between seven and 10 children per classroom teacher, resulting in 1,551 children assessed at pretest across 188 teachers. The research team followed the children over 2 years to collect outcome data in spring 2012 and then again in spring 2013. At the spring 2012 assessment wave, approximately 9% were lost to follow-up for a variety of reasons (e.g., family moved, child absences). At the spring 2013 assessment wave, 42% were lost to follow-up. Exhibit 11 shows attrition during the 1st year and 2nd year for each of the child math outcome measures separately (Woodcock-Johnson-III Applied Problems (WJ-AP) subtest, Tools for Early Assessment in Math–TEAM). Most children were lost to follow-up at 1-year posttest or 2-year posttest because the child had left the original school and transferred to another school in the district. The second most common reason for a child being lost to follow-up was because the child had an IEP or 504 plan and was not eligible to be tested.

[4] The target age range was 4 years to 9 years. However, some schools mistakenly consented 3-year-old children if they were enrolled in the preK classrooms, but this was not consistent across the 16 participating schools. These younger children were included in the data analyses presented below.

Exhibit 11. Sample of Participating Children and Attrition for Impact Study Child Math Measures

	Applied Problems			TEAM		
Number of children	Total	Intervention	Comparison	Total	Intervention	Comparison
Consented	2,609	1,298	1,311	2,609	1,296	1,311
Pretest (administered)	1,551	795	756	1,545	792	753
Pretest (useable)[a]	1,544	791	753	1,503	773	730
1-year posttest (administered)	1,411	725	686	1,411	725	686
1-year posttest (useable)	1,410	724	686	1,406	721	685
1-year posttest (ITT sample with both pretest and 1-year posttest)	1,404	720	684	1,367	704	663
2-year posttest administered	914	451	463	913	451	462
2-year posttest (useable)[a]	912	449	463	884	438	446
2-year posttest (ITT sample with both pretest and 2-year posttest)	907	446	461	855	427	428

Note. Applied Problems = Woodcock-Johnson-III Applied Problems (WJ-AP) subtest; TEAM = Tools for Early Assessment in Math–TEAM; ITT = intent-to-treat.

[a]Most of the Applied Problems assessments at each time point were useable; however, 42 TEAM assessments administered at pretest were completed incorrectly and/or missing basal/ceiling items. Additionally, five TEAM assessments at 1-year posttest and 27 TEAM assessments at 2-year posttest were not useable for the same reasons.

Measures

Below we describe the measures in detail, including how fidelity of implementation was measured, how the teacher outcome measures were collected and scored, and measures used to assess children's math abilities.

FIDELITY OF IMPLEMENTATION MEASURES

As described earlier, attendance at learning labs, leadership academy meetings, and the summer institute were tracked through the use of sign-in sheets or attendance logs at the events. Coaching cycles and grade-level meetings, which were held at the school, were tracked by project coaches. Coaches carried tablets, and via Wi-Fi, entered attendance and other data directly into a database as they delivered services (see Exhibit 12).

Exhibit 12. Data Collection Schedule and Measures for Fidelity of Implementation

Grade	Year 1 (2011–12)	Year 2 (2012–13)
Learning labs	Attendance logs	Attendance logs
Summer institutes	Attendance logs	N/A
Coaching cycles	Coaching logs	Coaching logs
Grade-level meetings	Coaching logs	Coaching logs
Leadership academy	Attendance logs	Attendance logs

TEACHER OUTCOME MEASURES

The three researcher-developed measures used for assessing teacher outcomes are described below. Given the scarcity of valid tools available, Erikson developed a survey of teachers' attitudes, beliefs, and confidence related to the teaching and learning of early mathematics (ABC-EM), a survey of teachers' pedagogical content knowledge (PCK-EM), and an observation tool of teachers' classroom math instruction practices (HIS-EM). For each measure, we describe how the data are collected and provide information relevant to scoring and a short description of the reliability and evidence of validity. Exhibit 13 shows the data collection schedule for teacher outcome measures.

Exhibit 13. Data Collection Schedule for Teacher Outcome Measures for Participating Teachers

Measures	Pretest (Fall 2011)	1-Year Posttest (Spring 2012)	2-Year Posttest (Spring 2013)	3-Year Posttest (Spring 2014)	4-Year Posttest (Spring 2015)
HIS-EM	X	X	X	X	X
ABC-EM	X	X	X	X	X
PCK-EM	X	X		X	X

Note. No data are available for PCK-EM in spring 2013 because coding procedures have not been developed for the PCK-EM videos used at that time point (also described below in the description of the PCK-EM measure). HIS-EM = High-Impact Strategies for Early Mathematics; ABC-EM = Attitudes, Beliefs, and Confidence for Early Mathematics; PCK-EM = Pedagogical Content Knowledge for Early Mathematics.

High-Impact Strategies for Early Mathematics (HIS-EM)

High-Impact Strategies for Early Mathematics (HIS-EM) is a lesson-based observation tool that is designed to be used in prekindergarten through third-grade classrooms in order to measure the quality of mathematics teaching. HIS-EM is comprised of three domains: *What*, *Who*, and *How*. The *What* domain refers to the degree to which observed practice incorporates a deep knowledge of foundational mathematics concepts. The *Who* domain refers to the degree to which observed practice demonstrates an understanding of young children's typical developmental growth in mathematics and an understanding of individual students' learning needs. The *How* domain refers to the degree to which observed practice includes the effective use of mathematics teaching strategies. The operational definition of each HIS-EM indicator is shown in Exhibit 14.

The three domains are further divided into nine dimensions representing teaching strategies that make a significant impact on students' mathematics learning. In the ***What*** domain, the three dimensions are *Learning Objectives*, *Mathematical Representations*, and *Concept Development*. For the ***Who*** domain, the dimensions are *Attention to Developmental Trajectories*, *Response to Students' Individual Needs*, and *Developmentally Appropriate Learning Formats*. The three dimensions of the ***How*** domain are *Planning*, *Student Engagement*, and *Establishment of a Mathematical Learning Community*. Finally, dimensions are described by various observable indicators. Each dimension consists of three to four indicators of high-impact instruction (Exhibit 14). It is important to note that these indicators may not always be present or salient for each lesson. In other words, the indicators under each dimension are not a checklist; rather, observers evaluate dimensions holistically.

Exhibit 14. Operational Definition of High-Impact Strategies for Early Mathematics (*HIS-EM) Indicators*

Domain	Dimension	Indicator	Operational Definition
What—Knowledge of Foundational Mathematics Concepts	Learning Objectives	Clarity	Learning objectives are clear.
		"Big Ideas"	Learning objectives reflect conceptual understanding and important learning.
		Integrates with prior knowledge	The teacher integrates the lesson with prior knowledge.
		Reorientation statements	The teacher effectively focuses students' attention toward the purpose of the lesson.
	Mathematical Representations	Words and Gestures	Mathematical words and gestures are used frequently and correctly to illustrate concepts.
		Tools	Mathematical tools enable students to investigate concepts and represent their ideas. Connections are made between tools and mathematical concepts.
		Models	Mathematical models are accurate, varied, and help students make connections between concepts.
	Concept Development	Accuracy	The teacher displays deep, connected content knowledge.
		Anticipates common student misconceptions	The teacher anticipates common student misconceptions and successfully clarifies concepts for students.
		Deeper understanding	The lesson leads students to a deeper understanding of the concept.
		Concluding statements	The teacher concludes the lesson by summarizing mathematical concepts and helping students generalize their understanding.
Who—Knowledge of Young Children	Attention to Developmental Trajectories	Typical mathematical development by topic	The teacher displays knowledge of the developmental trajectory for this mathematical topic.
		Scaffolding	The teacher consistently provides scaffolding that builds students' understanding within their mathematical zone of proximal development.
		Using student error	The teacher is consistently responsive to students who make errors and uses "wrong" answers to deepen students' understanding.
	Response to Students' Individual Needs	Differentiation	The teacher displays knowledge of all students' skills and conceptual understanding, including those with special needs. The lesson is differentiated to support all students.
		Monitors student work	The teacher consistently monitors student work and looks for evidence of learning.
		Flexibility	The teacher successfully adjusts the lesson in response to students' needs.
		Informal Assessment	Informal assessment is focused on conceptual understanding and process. There is evidence that the teacher has assessment criteria in mind that guides observation and/or documentation.
	Developmentally Appropriate Learning Formats	Grouping	The instructional grouping is appropriate and productive.
		Pace	Pacing of the lesson is appropriate for the students and productive.
		Variety of Modalities	The teacher uses a variety of modalities to effectively interest students and gain their active, hands-on participation.
		Connections	The teacher often helps students connect mathematics to their own experience, to the world around them, and to other disciplines.
How—Knowledge of Instructional Methods	Planning	Activity Selection	The activities of the lesson are focused on exploring mathematical concepts.
		Lesson Design	All components of the lesson are mathematically connected and coherent.
		Preparation	The teacher is fully prepared for the activities.
	Student Engagement	Problem Solving	The teacher provides many opportunities that excite students to participate to engage in problem solving.
		Questions	The teacher frequently asks open-ended questions with more than one possible solution/strategy.
		Explanation and Justification	The teacher often asks "what, how, why" questions or otherwise solicits students' explanations/justifications.
	Establishment of a Mathematical Learning Community	Attitude toward Mathematics	The teacher shows genuine enthusiasm for mathematics.
		Expectations	The teacher communicates high expectations for all students and consistently offers encouragement of students' efforts that increase their persistence.
		Regard for Student Perspectives	The teacher is flexible, incorporating students' ideas when appropriate and allowing choices based on students' interests.
		Mathematical Discussion	Mathematical discussion appears to foster a sense of community in which students feel free to express their mathematical ideas honestly and openly.

In this study, HIS-EM was used as a measure of the quality of teachers' instructional practices and strategies for teaching math. HIS-EM observations were completed by staff who had been certified in conducting the observations reliably (see below for certification procedures). Observations were scheduled during the designated math lesson times with the goal of being minimally disruptive to teachers and students and to ensure the observations captured a "typical" lesson. Teachers were asked to teach the math lesson they had planned for the day, with the exception of any student testing. Teachers were also encouraged to teach in Spanish or English as needed and could mark on the schedule sheet if a bilingual observer was needed. Teachers were invited to choose a time close to their usual lesson time and to be sure a full lesson could be observed from start to finish. Project staff explained no other special arrangements were necessary. The observer would simply be "a fly on the wall"—a quiet observer in their classroom with little interaction with the teacher or students. Teachers were reminded each year that their participation was voluntary, all information was confidential, and they would not receive feedback from the observations as they were used in the aggregate, for research purposes only. Teachers received two science-themed books for their participation at each data collection point.

Certified observers conducted observations in fall 2011 and spring 2012, 2013, 2014, and 2015 at each participating school. Each classroom was observed once per time point. Observers remained in each classroom for the duration of the mathematics lesson and took notes on each dimension. HIS-EM scoring was completed immediately following the end of the lesson. The observer could refer back to the manual and the notes taken during the lesson as much as he or she needed to arrive at a score. Observers assigned scores for each dimension and wrote notes of justification on the Observer Score Sheet. Observers rated each dimension using a 7-point scale (1 being the lowest and 7 being the highest score possible). The dimension descriptions provided explanations and examples of each scale at the low (1,2), middle (3,4,5) and high (6,7) ranges. The HIS-EM manual provided examples of typical low-range, mid-range, and high-range scores for mathematics teaching quality at the indicator level for each dimension. Domain scores for each observation were computed by averaging the appropriate dimension scores. An overall HIS-EM score was the sum across dimensions. An average HIS-EM score was the overall score divided by the nine dimensions, with low scores represented as 1 to 3 and high scores represented by 6 to 7.

At each time point of the data collection, six to eight certified HIS-EM observers completed HIS-EM observations at intervention and comparison schools. All observers had a bachelor's or master's degree in early childhood education, education, or other related field, prior experience in teaching preschool and/or elementary school, and/or in supporting, training, or evaluating teachers. Below we describe briefly the process through which HIS-EM observers were trained and certified.

HIS-EM master coding procedures. Previously videotaped mathematics lessons from the Collaborative's extensive video collection were used to create a master-coded video library for HIS-EM trainings. The lessons in the videos included a variety of early mathematics content (e.g., number and operations, measurement, geometry). In addition to the images of teachers' teaching throughout the lesson, some videos also zoomed in on students working independently, in small group discussions, or solving math problems on the board. Videos used for training and calibration were rated by master coders who developed the HIS-EM observation tool. Master coders had extensive training in child development and early education. Coders also were experienced in teaching math and providing professional development. After each member in the team watched and scored the assigned videos, the team got together to discuss their scores and justifications. Based on the group discussion, final scores (which all team members agreed upon) were assigned to each video and these codes are labeled as master scores.

HIS-EM certification procedures (Years 1–3)**.** Novice HIS-EM coders participated in a three-phase training program involving: initial training, reliability training, and drift training. The initial training consisted of 2-day in-person training that involved the following activities: reading the HIS-EM Manual and relevant literature in early mathematics (e.g., Common Core State Standards, Big Ideas, etc.), studying the HIS-EM coding guides, and viewing and scoring three videotaped classroom observations and HIS-EM anchor videos (a total of 18 low and high anchor videos) in preparation for the online reliability testing. During the last day of the training, observers practiced assigning scores for at least two videos of real mathematics lessons (lesson grade level tended to vary across trainings) and further developed their understanding of the HIS-EM through opportunities to ask questions and engage in discussion.

The reliability phase involved independent coding of eight video recorded lessons and meetings with project staff to review and discuss the observer's codes. During this phase, observers were first asked to watch and score a set of five videos. Observers who were reliable in at least three out of five videos were invited to the second reliability phase in which they were asked to watch and score a set of three videos. Observers who failed the initial reliability phase were no longer used as coders on the project and were not invited to the second phase. Observers who failed in second phase were asked to watch and score another set of three videos. If they still failed to achieve the required reliability, they were not retained as coders on the project. In order to be certified, observers were required to meet or surpass the following criteria: 80% of all observer-assigned scores had to be within 1 point of the master-coded scores on five out of eight videos. All the participating observers were certified at each time point except for fall 2011 and spring 2015. In fall 2011, 82% of the trained observers reached reliability and were certified. In spring 2015, 78% of the trained observers reached reliability and were certified.

For the drift phase, all certified observers independently coded one videotaped lesson to confirm their reliability in coding on a monthly basis during in-person meetings. The number of drift meetings ranged from one to three depending on the length of the data collection period. During the drift meeting, they discussed their codes and verified convergence among coders and with master coders. These meetings provided opportunities for observers to regularly ask questions and engage in HIS-EM-focused discussion.

For each posttest data collection, previously certified observers were asked to attend a 5-day recertification. They coded two master-coded and videotaped math lessons as a group and individually. The reliability phase involved coding three video recorded lessons and reviewing/discussing coding with project staff. In order to be recertified, observers were required to meet the following criteria: 80% of all observer-assigned scores had to be within 1 point of the master-coded scores on two of three videos. If they failed the first phase, they were given a second opportunity with another set of three videos. All observers that attempted recertification succeeded. In Year 4, additional training time and resources (i.e., Illinois Early Learning Project, 2013; Reinhart, 2000) were added to improve reliability and certification; certification was made more difficult (i.e., observers had to reach 80% agreement on four of five videos), and drift meetings occurred online.

HIS-EM Reliability and Evidence of Validity

The developers conducted analyses to examine reliability and gather evidence of validity using data collected during the 4-year project. Cronbach's alpha ranged from .92 to .98 across the data collection periods. Using a subset of the spring 2013 sample of teacher observations, interrater reliability coefficients ranged from 0.35 to 0.61 depending on the dimension. Using a subset of the spring 2014 sample of teacher observations, interrater reliability ranged from 0.47 to 0.85 depending on the dimension. To examine the level of disagreement between observers on HIS-EM dimensions in each of these samples, percent adjacent (equivalent or within 1 point) agreement was calculated. Percent adjacent agreement ranged from 71% to 83% (median = 75%) across dimensions (spring 2014 sample), and percent adjacent agreement ranged from 50% to 81% (median 62%) across dimensions (spring 2015 sample).

To assess evidence of convergent and discriminant validity of the HIS-EM, the developers examined how scores on the HIS-EM compared with scores on the CLASS (Classroom Assessment Scoring System) (Pianta, LsaParo, & Hamre, 2008). HIS-EM is a subject-specific measure of instructional quality in early mathematics teaching whereas the CLASS is a global measure of instructional quality. The developers used data from 27 preK and K intervention teacher observations (HIS-EM) in Year 1 of the current study and examined the HIS-EM and CLASS scores for convergent and discriminant validity. Essentially, a set of teachers were video recorded conducting a math lesson and these videos were then scored by certified HIS-EM observers and by certified CLASS observers.

The CLASS domains (emotional support, instructional support, and classroom organization) and HIS-EM average score were moderately correlated (rs = 0.44 to 0.58), with the strongest relationship between the HIS-EM overall score and the CLASS Instructional Support scores ($r = 0.58$). Bivariate linear regression analyses indicated that the HIS-EM average score was significantly associated with the CLASS average score (which was the average across the 10 dimensions of the CLASS), $\beta = 0.54$, $t(53) = 4.64$, $p < 0.001$, explaining a significant portion of variance in the CLASS average score, $R^2 = 0.29$, $F(1, 52) = 21.50$, $p < 0.001$.

The HIS-EM average score was significantly associated with CLASS Instructional Support score, $\beta = 0.58$, $t(53) = 5.13$, $p < 0.001$, explaining a significant portion of variance in the CLASS Instructional Support score, $R^2 = 0.34$, $F (1, 52) = 26.30$, $p < 0.001$. The HIS-EM average score was significantly associated with CLASS Emotional Support score, $\beta = 0.54$, $t(53) = 4.63$, $p < 0.001$, explaining a significant portion of variance in the CLASS Emotional Support score, $R^2 = 0.29$, $F (1, 52) = 21.45$, $p < 0.001$. The HIS-EM average score was significantly associated with CLASS Classroom Organization score, $\beta = 0.44$, $t(53) = 3.53$, $p < 0.001$, explaining a significant portion of variance in the CLASS Classroom Organization score, $R^2 = 0.19$, $F (1, 52) = 12.47$, $p < 0.001$.These analyses support convergent validity but do not support discriminant validity.

Attitudes, Beliefs, and Confidence in Early Mathematics (ABC-EM)

The *Attitudes, Beliefs, and Confidence in Early Mathematics (ABC-EM)* survey, developed by the Erikson team, consists of 51 statements that are designed to tap into three areas:
(1) teachers' *attitudes* toward math in general and their enjoyment in teaching math;
(2) teachers' *beliefs* about the appropriateness of early math for young children, effects of home environment on mathematics learning, English language learners and math learning, personal mathematics competence, adequacy of preparation, and current need for support; and
(3) teachers' *confidence* in understanding math content for teaching and performing specific math teaching tasks (see Appendix F for a list of the items used in the analyses in this report). The survey was presented in an online format. Teachers were asked to rate each statement on a 1 (*strongly disagree*) to 10 (*strongly agree*) scale. Data were collected in the fall of 2011 and the spring of 2012, 2013, 2014, and 2015.

ABC-EM Reliability and Validity

The developers conducted analyses on the reliability of the ABC-EM measure using data collected during the 4-year project.

Factor analyses completed by the developers revealed two factors:

- Factor 1, ***Confidence in Math Teaching***, was the strongest factor with several items with factor loadings > .60.
- Factor 2, ***Positive Math Attitudes***, was less strong, with no more than three factor loadings > .60 at any one time point.

The internal consistency of each factor was examined using Cronbach's alpha. Both subscales, at all time points, demonstrated an excellent level of internal consistency, ranging from .93 to .95 for Confidence in Math Teaching Subscale and ranging from .90 to .91 for Positive Math Attitudes Subscale.

Pedagogical Content Knowledge for Early Mathematics (PCK-EM)

The *Pedagogical Content Knowledge in Early Mathematics (PCK-EM)* survey developed by the PD team is a video-elicited, open-ended survey to capture educators' content knowledge for teaching mathematics from preK through third grade. Pedagogical content knowledge (PCK) represents the blending of content knowledge and pedagogy that is needed to effectively promote learning (Shulman, 1986). The current project utilized a three-dimension model of PCK for early mathematics: *What*—a deep understanding of mathematics topics necessary for teaching young children; *Who*—knowledge of learners' conceptions about specific mathematics content; and *How*—math specific pedagogical knowledge. Note these three domains match what was measured in the HIS-EM. Each dimension of PCK is further broken down into two subcomponents shown in Exhibit 15.

Exhibit 15. Conceptual Framework of the Pedagogical Content Knowledge in Early Mathematics *(PCK-EM) Survey*

Dimension	Subcomponent
What	*Depth*: Understanding of a specific *big idea or big ideas*, demonstrated by the capability of "deconstructing" a foundational math concept into its complex underlying ideas that young children need to learn.
	Breadth: Awareness of mathematical concepts related to a specific *big idea or big ideas*.
Who	*Prior Knowledge*: Understanding of young children's prior knowledge in learning a specific *big idea or big ideas*.
	Misunderstandings: Knowledge of students' likely misunderstandings and learning difficulties around a specific *big idea or big ideas*.
How	*Strategy:* Knowledge of pedagogical strategies (either from the video or for own teaching) that can facilitate, reinforce, and/or extend students' understanding of a specific *big idea or big ideas*.
	Representation: Knowledge of specific representations (illustrations, examples, models, demonstrations, and analogies) that can make clear a specific *big idea or big ideas to* facilitate, reinforce, and/or extend students' understanding.

As part of the survey, the teacher is asked to watch two videos of authentic teacher-led math lessons. In the teacher-led mathematic lesson, content domains relate directly to the common mathematical concepts taught in preschool, kindergarten, and early elementary levels. The content is intentionally selected and edited to focus on a math topic tied to a specific big idea of foundational math. A total of four videos were used to elicit responses on the PCK-EM (see Exhibit 16 for the sequence of videos used during PCK-EM data collection). Currently, the project has developed coding rubrics for the "Number" video and "Fraction" videos, but not the rubrics for "Map" video or "3D Shapes" video. Therefore, the data corresponding to Map video and 3D Shapes video will not be discussed or presented.

Exhibit 16. Delivery Design for Pedagogical Content Knowledge in Early Mathematics *(PCK-EM) Survey*

Time	Video 1	Video 2	Coded
Fall 2011	Number 7	Fraction	Yes
Spring 2012	Number 7	Fraction	Yes
Fall 2012	Map	3D Shapes	No
Spring 2013	Map	3D Shapes	No
Spring 2014	Number 7	Fraction	Yes
Spring 2015	Number 7	n/a	Yes

After watching each video, teachers were asked to answer nine open-ended questions aligned to the multiple facets of PCK regarding the central and relevant math concepts in the video, students' likely misunderstanding and prior knowledge of the particular topic, and instructional strategies to make the content accessible to students who are more advanced and those who are struggling. Below are the nine questions:

(1) What is the *central mathematical* concept of this activity? Please justify your answer.

(2) What are *other* important *mathematical* concepts you think are related to the central mathematical concept of this activity? Please justify your answer.

(3) What *prior mathematics* knowledge do children need to have in order to understand the central mathematical concept of this activity?

(4) Do the *children* appear to *understand* the central mathematical concept of this activity? Please provide evidence that supports your assessment.

(5) Based on your assessment, *what would you do next* to reinforce or extend children's understanding*?* Please justify your answer.

(6) What are some *common mathematical misunderstandings* children might have when learning this central mathematical concept?

(7) What has the teacher *done or said* to help the children understand the central mathematical concept? (Consider things like: materials, setting, lesson design, teacher language, pacing, interactions, etc.) Were those instructional choices *effective*? Please justify your answer.

(8) How could the teacher *change this activity* to meet the needs of a child who is *struggling*? (Consider things like: materials, setting, lesson design, teacher language, pacing, interactions, etc.) Please justify your answer.

(9) How could the teacher *change this activity* to meet the needs of a child who is *advanced?* (Consider things like: materials, setting, lesson design, teacher language, pacing, interactions, etc.) Please justify your answer.

Originally, responses to each prompted question were scored using a coding rubric; however, the developers found that teachers demonstrated their understanding of the content, students, and pedagogy throughout their responses. Therefore, the revised scoring system consisted of reviewing teachers' responses to all nine questions and then assigning codes to each of the six subcomponents using a 1 (*low*: more obvious, behavioral, or procedural) to 5 (*high*: sophisticated/concept) scale. The subcomponents in each dimension are averaged to produce a dimension score. The dimension scores are then averaged to produce an average PCK-EM score.

Procedures. The PCK-EM survey was administered online. Teachers who consented to participate in research received an email notice with an online link to the survey and a confidential ID number. The entire survey could be completed from any computer with Internet access, and in general took about an hour to finish. Technical support and assistance were available from the implementation project staff via phone or email (and the contact information was included in the same email).

Rater training procedures. After 2-week intensive discussions about the conceptualization of PCK, the big ideas covered in specific video stimuli and getting familiar with the coding framework and rubrics, coders independently coded 10 responses retrieved from the PCK-EM survey. After meeting a certain level of interrater reliability (2 adjacent scores from 1 to 5 were considered acceptable among four coders, and the overall agreement needed to be 80% or above), coders started coding about 20 responses per week.

Data coding procedures. Three trained coders coded teachers' PCK-EM responses within 6 weeks. Responses collected at multiple time points were coded during the same coding period, except for the data collected in spring 2015, which was coded separately. The coders were blind to whether the responses were from intervention or comparison teachers.

PCK-EM Reliability and Validity

Interrater reliability was examined by assigning the same set of responses (20%) to two coders; repeated within-coder reliability was checked through giving each coder a portion of their previously coded responses (4 weeks apart). The whole process was blind to the coders.

Coders were given feedback regarding inconsistencies; discrepancies were then discussed and resolved before moving to the next set of coding. The discrepancies were checked with two master coders' codes (a doctoral student and a post-doc researcher who worked closely on developing the coding rubrics). Interrater reliability was calculated using percent adjacent analysis and percentage of exact agreement and intra class correlations (ICCs). The percent adjacent ranged between 90% to 97% between coders, and 40% to 65% if using exact match criteria. Individual ICCs were between .48 and .84, and average ICCs were between .65 and .91. Intrarater reliability was calculated using percent adjacent analysis as well as percentage of exact agreement. Percent adjacent ranged from 87.5% to 100%, with exact match percentage ranged from 50% to 62.5%.

Predictive validity was also evaluated by conducting HLM models (that accounted for the fact that students were nested by classrooms/teachers) to examine associations between teachers' PCK-EM scores at the beginning of the year and students' performance on two measures of mathematical understanding at the end of the year: WJIII-Applied Problems subtest and TEAM. Teachers' knowledge demonstrated by the three dimensions of PCK-EM was significantly associated with students' math learning, but in different ways. For instance, scores on the *What* dimension of PCK-EM were significantly and reliably associated with students' end of the year math performance (β= 3.55, p < .05 for Applied Problems subtest score and β = 2.94, p < .05 for TEAM T-score). More specifically, a 1-point increase in teachers' *What* score was associated with a 3.55 point increase on the Applied Problems subtest (*es* = 0.20). In addition, there was an interaction between scores on the *Who* dimension of PCK-EM and Applied Problems pretest scores (β = .67, p < .001) when examining associations with Applied Problems posttest scores (β = -.12, p < .10). When teachers scored higher on the *Who* dimension of PCK-EM, students with limited mathematical understanding benefited more by the end of the school year. There was a positive significant interaction between scores on the *How* dimension of PCK-EM (β = .10, p < .10) and students' TEAM pretest scores (β = .47, p < .001) when examining associations with the TEAM posttest scores. That is, students with more advanced mathematical understanding benefited from teachers who scored higher on the *How* dimension of PCK.

CHILD OUTCOME MEASURES

The researchers hired and trained local staff to conduct two direct assessments of children's math abilities: the Woodcock-Johnson-III Applied Problems (WJ-AP) subtest (Woodcock, McGrew, & Mather, 2001) and the Tools for Early Assessment in Math–TEAM (D. H. Clements, Sarama, & Wolfe, 2011). The team of assessors, which had substantial experience conducting assessments of young children in early childhood settings, had previously been trained to conduct WJ-AP. As part of this project, they were trained to conduct the TEAM reliably as well. Assessors were deemed certified on the TEAM if they reached 85% reliability with master codes (these processes were overseen by the developers of the TEAM). Assessors averaged 92% reliability (or accuracy).

Both assessments were administered at pretest, 1-year posttest, and 2-year posttest. Pretest measures were collected between September and December of 2011; posttest measures were collected between April and June of 2012; and follow-up assessments between April and June 2013. Children were assessed as early as the end of September and as late as the end of November/beginning of December at pretest. Assessors were blind to children's experimental condition at the time of assessment.

Woodcock-Johnson III-Revised (WJIII-R) Applied Problems Subtest

Woodcock-Johnson-III Applied Problems (WJ-AP) subtest is an individually administered norm-referenced test that measures skills in analyzing and solving practical math problems with a total of 60 items. The tester verbally presents items involving counting, telling time or temperature, and problem solving. Testing is discontinued after six consecutive errors. The score is the number of correct items. One-year test-retest reliability is .92 for children ages 4 to 7 years. Internal alpha reliability estimates are .88 to .94 for children ages 4 to 7 years. Age-standardized scores were used in the analyses. In addition, children who were identified as ELL and required testing in Spanish were administered the Applied Problems subtest (Problemas aplicados) of the *Batería III Woodcock-Muñoz*, which is the Spanish adaptation of the Woodcock-Johnson III (Woodcock, Munoz-Sandoval, McGrew, & Mather, 2005). Internal alpha reliability estimates are .91 to .95 for children ages 4 to 9 years (Schrank et al., 2005).

Tools for Early Assessment in Math (TEAM)

The TEAM measures core mathematical abilities of young children across 19 learning trajectories. The assessment is administered individually to children ages 4 to 9 years of age with standardized protocol and scoring procedures. Trajectories are theoretically and empirically based developmental progressions and include verbal counting, object counting, subitizing, number comparison, number sequencing among others, and geometry progressions that include shape recognition, congruence, and spatial imagery, as well as measurement. The developers of the TEAM define mathematical competence as a latent trait using the Rasch model, yielding a score that locates children on a common ability scale with a consistent metric (Douglas H. Clements, Sarama, & Liu, 2008). All items are ordered by Rasch item difficulty; children stop the assessment after four consecutive errors. For the present study, we used the T-scores, which are derived from the Rasch scores computed for the total instrument following the developers' guidelines.

The TEAM does not have published psychometrics, but an earlier version of the assessment (REMA) has shown to be a reliable measure of the math abilities of young children ages 3 to 5 years (Douglas H. Clements et al., 2008). Coefficient alpha reliabilities ranged from .89 (number) to .71 (geometry). Interrater reliability was .98. Item reliability was .98 with the total test scores. Items were reviewed by a panel of experts to ensure content validity. Concurrent validity of the total test score was established with a .86 correlation with another measure of preschool math abilities (Starkey, Klein, & Wakeley, 2004). This version of the measure was further refined to yield the current TEAM (Tools for Early Assessment in Math), which has 158 items and can be used with children ages 4 to 9 years. The TEAM has been translated into Spanish, and all Spanish-speaking assessors were trained to reliability on both the Batería and administering the TEAM in Spanish.

Illinois Standards Achievement Test (ISAT) Scores

The Illinois Standards Achievement Test in Mathematics was a state assessment administered in the spring to students enrolled in Grades 3−8. Items on the Math ISAT were aligned with the Illinois State Learning Standards. Performance on the Math ISAT was expressed in scaled scores that ranged between 120 and 341 for third-grade students. Performance was also reported in terms of performance levels: Academic Warning—student demonstrates limited knowledge and skills; Below Standards—student demonstrates basic knowledge and skills; Meets Standards—student demonstrates proficient knowledge and skills; and Exceeds Standards—student demonstrates advanced knowledge and skills. ISAT scores were obtained from the school district for all student participants who were enrolled in third grade at any point between 2012 and 2014. Mean scaled scores and performance levels were also obtained for all 16 schools in the study from the Illinois State Board of Education website. In 2015, the ISAT was replaced by the Partnership for Assessment of Readiness for College and Careers (PARCC) assessment; therefore, 2015 ISAT scores were not available.

Analysis Approach

This study was designed to have sufficient power to detect a meaningful intervention impact of effect sizes of at least 0.30 on student outcome.[5] Given that the children and teachers in this study are nested within schools, the magnitude of the intervention effects on students was tested using multilevel analysis, also referred to as hierarchical linear modeling (HLM). This process adjusts standard errors to account for the dependence among students within schools, thus avoiding overestimation of statistical significance of the effect size. A series of multilevel analyses were conducted to examine the effect of the intervention on students' math achievement after 1 year of intervention on spring 2012 test scores and after 2 years of the intervention on spring 2013 test scores after controlling for students' test scores in fall 2011 and other background characteristics. In addition, a series of multilevel analyses were conducted to examine the effect of the intervention on teacher outcomes after 1 year of intervention (spring 2012), 2 years of intervention (spring 2013), 3 years of intervention (spring 2014), and 4 years of intervention (spring 2015), if available. Before conducting the impact analyses, we tested for baseline equivalence for all intent-to-treat (ITT) samples as described below in the findings section.

IMPACT MODELS FOR TEACHER OUTCOMES

Teacher demographic variables included years of teaching experience and grade level at baseline. Grade level at baseline variables were three dummy coded variables for kindergarten, first grade, second grade, and third grade. PreK was the reference group. We conducted multilevel ITT impact analysis comparing intervention and comparison teachers on HIS-EM, ABC-EM, and PCK-EM outcomes. Dependent variables were teacher outcomes on HIS-EM, ABC-EM attitude, ABC confidence, and PCK-EM. Independent variables included a constant, baseline scores (same as the outcome measure in 2011), demographic characteristics, and intervention indicator.

[5] SRI conducted a power analysis showing the minimum detectable effect (MDES) for HLM analysis using the methodology in Schochet (2008). We calculated power based on the eight treatment vs. eight control schools in the study using the following assumptions: A cluster randomized design with children (level 1) within school clusters (level 2); Treatment is at the school level (level-2); the between-school intra-class-correlation (ICC) is 0.055; The average number of children per school was 75 after 20% attrition; Proportion of variance explained by level-2 covariates is 0.49; 80% power; A two-sided 5% alpha test is used. Under these assumptions, the minimum detectable effect size (MDES) is 0.30 for the two-year effect analysis.

The two-level HLM model for intervention effects is as follows:

$Y_{ik} = \beta_{00} + \beta_{01}PD + \beta_{02}(COV_{ik} - COV...) + e_{ik} + \mu_{0k}$, where Y_{ik} is outcome of teacher i in school k at follow-up. *PD* indicates the initial assignment with 1 for intervention and 0 for comparison. The coefficient β_{01} associated with the intervention in the above HLM model indicates the average intervention effect in promoting improved teacher outcomes. COV_{ik} are the covariates (baseline scores and demographic characteristics) and they were centered by mean. β_{02} are coefficients associated with each covariate. e_{ik} is teacher random effect, and μ_{0k} is school random effects. For the six models that used propensity score weighting to establish baseline equivalence, weights were included in the HLM models. For the six models that did not use propensity score weighting, weights were not included in the HLM models.

Effect sizes are reported as Hedge's *g* (U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse, 2013) and calculated by dividing the intervention indicator coefficient by the pooled standard deviation of the intervention and comparison group. The improvement index (U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse, 2013) was also reported, which translated the effect size into an improvement in percentile rank. The improvement index indicates the expected change in percentile rank for the median comparison teacher if that teacher had received the PD intervention.

IMPACT MODELS FOR CHILD OUTCOMES

Demographic and other variables examined as covariates included gender, age in months, home language, and grade level (reported by parents and children's school). ITT is the average effect of the intervention based on the initial intervention assignment regardless of how many participants actually received the intervention. The ITT analyses present the impact of assignment to the PD instead of the impact of the PD on children with teachers receiving the PD. The ITT impact estimate is the expected effect of the PD when it was implemented in the real world, with less than perfect teacher implementation and student dosage. The evaluation team conducted multilevel analysis to take into account children nested in schools. Dependent variables were children's math achievement measures (WJ-AP and TEAM). Independent variables included a constant, pretest scores, demographic characteristics, and intervention indicator. The two-level HLM model for intervention effects is as follows:

$Y_{ik} = \beta_{00} + \beta_{01} PD + \beta_{02}(COV_{ik} - COV...) + e_{ik} + \mu_{0k}$, where Y_{ik} is outcome of student *i* in school *k* at follow-up. *PD* indicates the initial assignment with 1 for intervention and 0 for comparison. The coefficient β_{01} associated with the intervention in the above HLM model indicates the average intervention effect in promoting improved student outcomes. COV_{ik} are the covariates (pretest and demographic characteristics) and they were centered by mean. β_{02} are coefficients associated with each covariate. e_{ik} is student random effect, and μ_{0k} is school random effects.

Sensitivity analyses were conducted to check the robustness of the impact of the intervention across different specifications of the models. HLM analysis was conducted on the data based on three sets of models: (1) simple model—Model A (predictors are intervention and pretest); (2) full model—Model B (predictors are intervention, pretest, and a full set of demographic characteristics); and (3) interaction model—Model C (additional interaction terms were added to the full model). Effect sizes are reported as Hedge's *g* and the improvement index (U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse, 2013) are reported for each model.

EXPLORATORY ANALYSES

Conventional growth modeling assumes that all subjects in the analysis sample are drawn from a single growth population—all growth trajectories (e.g., changes in the outcome over time) in the sample can be characterized by a single set of parameters (e.g., mean intercept, slope, and variances). However, it is also possible that there are multiple growth populations contained within the sample, and that subgroups of subjects can be differentiated from one another on the basis of distinct growth trajectory *classes*. When the data do not meet the assumptions of conventional growth models, these more commonly utilized methods will not fully describe the variety of growth trajectories found in the sample; thus, researchers risk overlooking substantively important information in the data.

To protect against this prospect, for examined growth trajectories in teacher math knowledge, attitudes, and confidence, and instructional strategies, the evaluation team used a Growth Mixture Models strategy. By doing so, they were able to examine whether distinct *classes* of growth trajectories are contained within the teacher data. This method also allowed us to examine whether specific teacher characteristics (i.e., intervention vs. comparison teachers, teacher certification, etc.) were associated with trajectory class membership. Teacher characteristics included in the models were grade, teacher certification category (e.g., early childhood, elementary education, early childhood special education, special education), second language fluency, age, gender, ethnicity, and number of years teaching at baseline.

Teacher outcomes were the HIS-EM, ABC-EM, and PCK-EM. List-wise deletion was utilized for each analysis so that for a given model, included teachers needed to have both outcome data and teacher characteristic data. Though each outcome variable was not collected in every instance, outcome data that were used came from a set of 4 time points: (1) fall 2011, (2) spring 2012, (3) spring 2013, and (4) spring 2014. HIS-EM data were drawn from time points 1 through 4; ABC-EM from time points 1 through 4; and PCK-EM from time points 1, 2, and 4. Teachers were excluded from an analysis when data from the specific outcome at any time point were missing, or data were missing on teacher characteristics.

Findings

Below we describe the main findings relevant to the evaluation questions in the implementation and impact studies. First, were the intervention components implemented with fidelity as intended? Second, what was the impact of the intervention on teachers' classroom practices, attitudes and confidence, and content knowledge? Finally, did the intervention have an impact on students' math achievement and skills?

Fidelity of Implementation Findings

Fidelity of implementation data were analyzed both at the teacher- and school-level, based on the criteria established a priori and described previously in the Introduction section. Fidelity of implementation data were analyzed both for Year 1 and Year 2 of implementation. It is important to note, however, that the criteria established a priori for Year 1 differed slightly from the criteria established a priori for Year 2. These changes were based on lessons learned during Year 1, and most often the revisions to the criteria made the criteria more stringent to ensure that fidelity of implementation data captured variability in teacher and school participation and engagement. Finally, data were analyzed using both the sample that included all enrolled teachers (including teachers who dropped early or started late) and a subsample that included teachers who were available during the full school year.

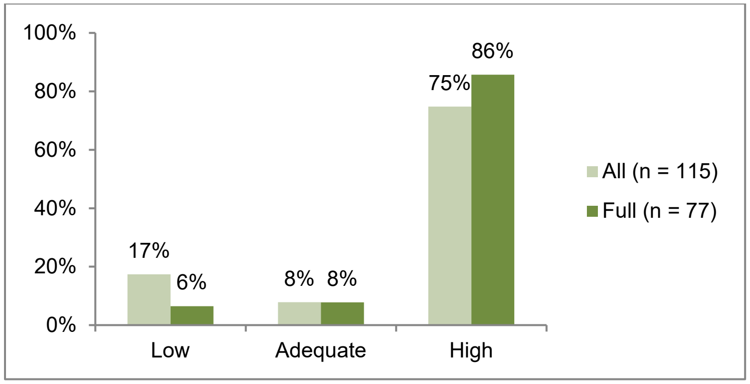
YEAR 1

Teacher Fidelity of Implementation

Four components of fidelity of implementation were analyzed at the teacher level in Year 1: hours attended learning labs; hours attended summer institute days; number of coaching cycles completed (cycle defined as planning, observation, and reflection); and number of grade-level meetings attended across the year, as shown in Exhibits 17 to 20. The criteria for establishing adequate or high fidelity are described below in each exhibit.

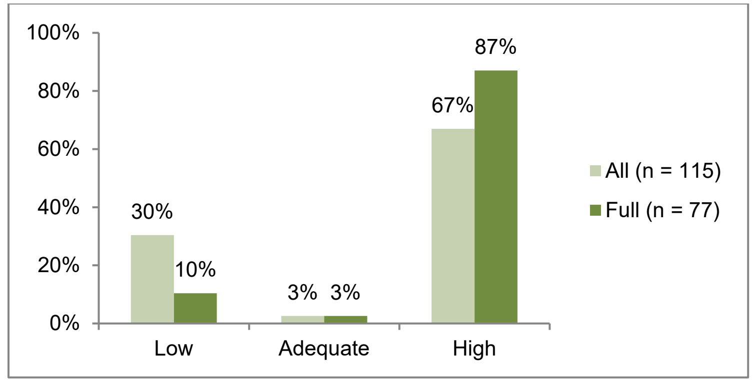
Overall, high fidelity of implementation was observed at the teacher level on all four components (see Exhibits 17 to 20). Findings from analyses using the two samples (all versus full) indicated a similar pattern of results. Slightly higher scores were observed across all four components in the full sample given this sample of teachers had more opportunities to reach fidelity. Specifically, about two-thirds of teachers met the high-fidelity criteria for grade-level meetings, learning labs, and summer institute meetings, and more than half of teachers met high-fidelity scores for completed coaching cycles.

Exhibit 17. Adequacy of Fidelity of Implementation: Learning Lab Attendance (Year 1)—Teacher Level



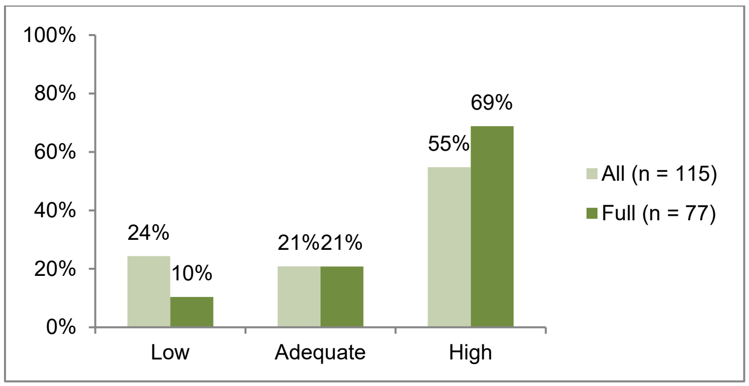
Note. Adequate fidelity was defined as teachers attending 15 to 18 hours of learning labs during the year; high fidelity was defined as teachers attending 19 to 22 hours of learning labs during the year.

Exhibit 18. Adequacy of Fidelity of Implementation: Summer Institute Attendance (Year 1)—Teacher Level



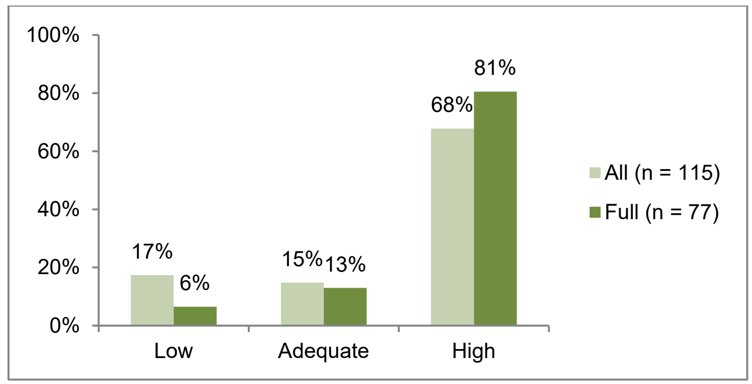
Note. Adequate fidelity was defined as teachers attending 10 to 14 hours of summer institute lab days; high fidelity was defined as teachers attending 15 to 20 hours of summer institute lab days.

Exhibit 19. Adequacy of Fidelity of Implementation: Coaching Participation (Year 1)—Teacher Level



Note. Adequate fidelity was defined as teachers completing 3 coaching cycles; high fidelity was defined as teachers completing 4 or more coaching cycles.

Exhibit 20. Adequacy of Fidelity of Implementation: Grade-Level Meetings Attendance (Year 1)—Teacher Level



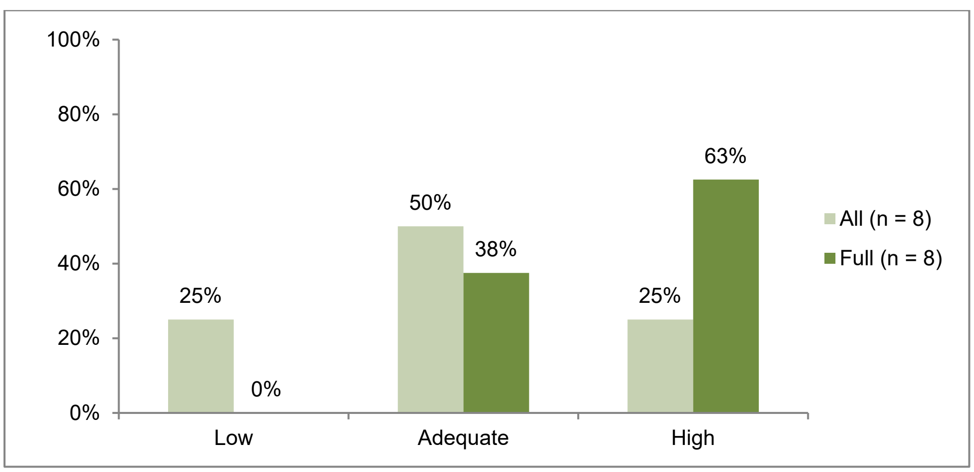
Note. Adequate fidelity was defined as teachers attending 4 to 5 meetings during the year; high fidelity was defined as teachers attending 6 meetings during the year.

School Fidelity of Implementation

Six components of fidelity of implementation at the school level were analyzed in Year 1: percentage of teachers who attended learning labs; number of grade-level meetings offered; percentage of teachers who attended grade-level meetings; percentage of teachers who attended summer institute meetings; percentage of teachers who participated in coaching; and number of leadership academy meetings offered.

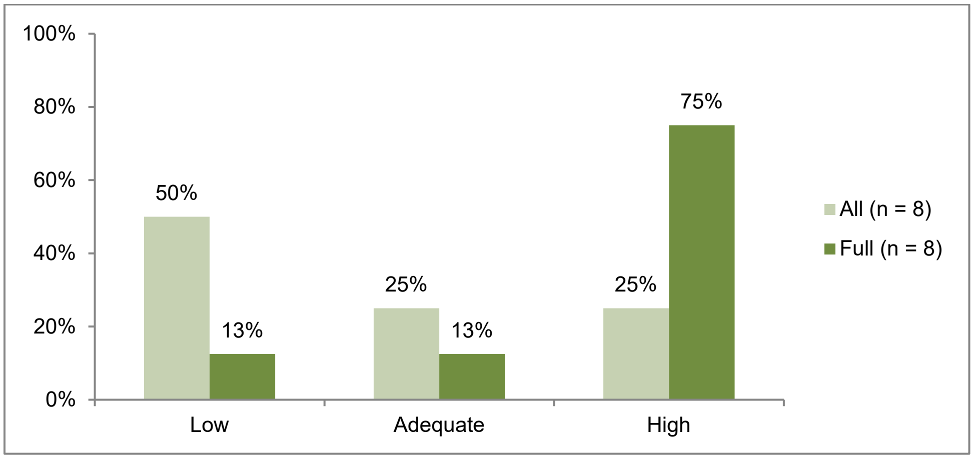
Findings for fidelity of implementation at the school level showed a mixed pattern, as shown in Exhibits 21 to 26. For instance, high fidelity was observed for some of the components (e.g., number of grade-level meetings offered) but not others (e.g., number of leadership academy meetings with adequate attendance). In addition, findings from analyses using the two samples (all versus full) indicated different patterns of results for some components and similar patterns of results for others. For instance, the percentage of schools that met the high fidelity for learning labs was 63% when only teachers who were available for the full school year were included, but it dropped to 25% when the sample included all teachers. A similar pattern was seen for coaching and summer institute components. These findings make sense but also demonstrate the challenge in reaching all teachers when schools are faced with high teacher turnover.

Exhibit 21. Adequacy of Fidelity of Implementation: Learning Labs (Year 1)—School Level



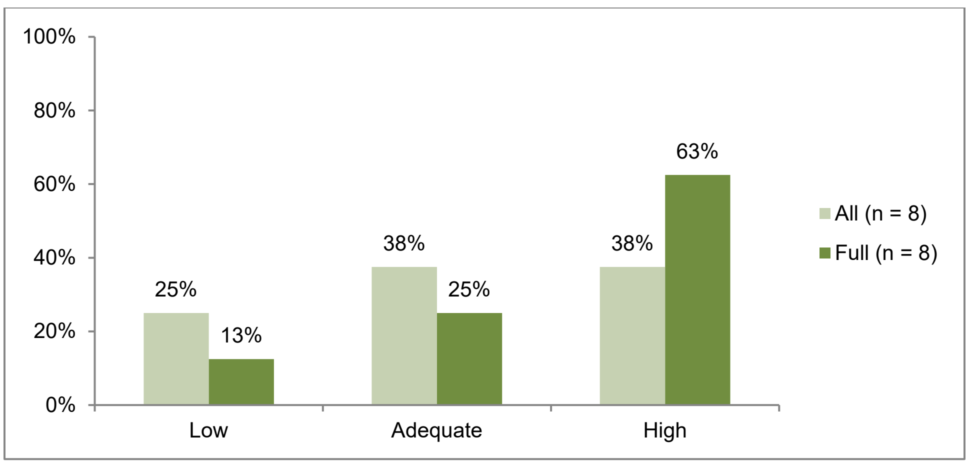
Note. Adequate fidelity for a school was defined as at least 70% of teachers with at least adequate attendance where adequate attendance was defined as attending at least 15 hours of learning labs during the year; high fidelity for a school was defined as at least 90% of teachers with at least adequate attendance.

Exhibit 22. Adequacy of Fidelity of Implementation: Summer Institute (Year 1)—School Level



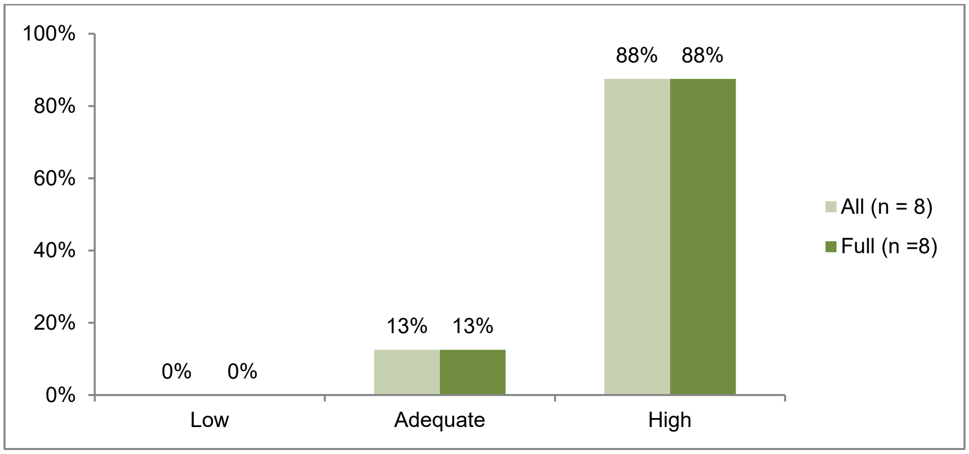
Note. Adequate fidelity for a school was defined as at least 70% of teachers with at least adequate attendance at summer institutes where adequate attendance was defined as attending at least 10 hours; high fidelity for a school was defined as at least 90% of teachers with at least adequate attendance in the summer institutes.

Exhibit 23. Adequacy of Fidelity of Implementation: Coaching Participation (Year 1)—School Level



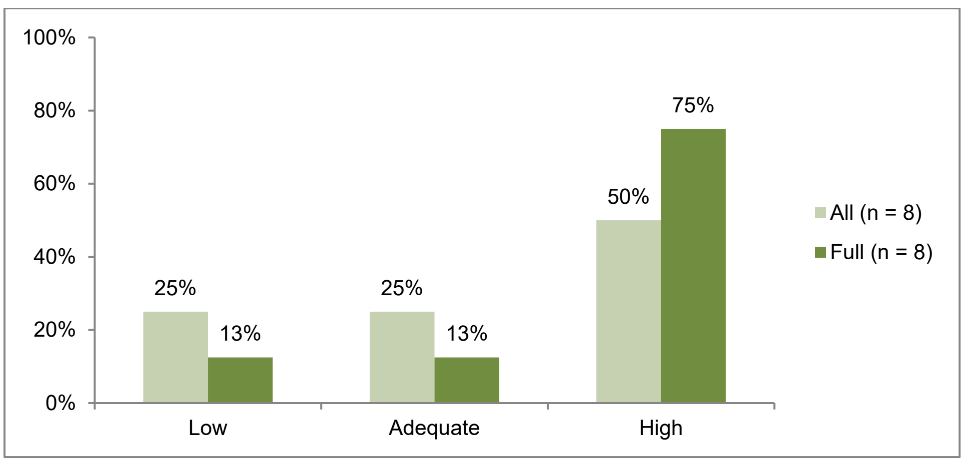
Note. Adequate fidelity for a school was defined as at least 60% of teachers with at least adequate coaching cycles completed where adequate coaching was defined as completing 3 or more coaching cycles; high fidelity for a school was defined as at least 90% of teachers with adequate coaching cycles completed.

Exhibit 24. Adequacy of Fidelity of Implementation: Grade-Level Meetings Held (Year 1)—School Level



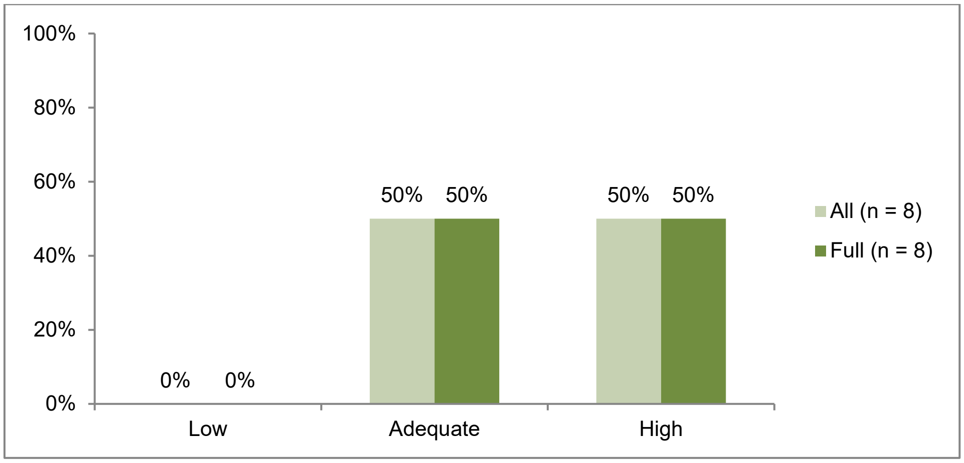
Note. Adequate fidelity for a school was defined as holding at least 18 grade-level meetings (across grades) and at least 60% of teachers attending at least 4 grade-level meetings across the school year; high fidelity for a school was defined as holding at least 25 grade-level meetings (across grades) and had at least 90% or more of teachers who attended at least 4 grade-level meetings across the school year.

Exhibit 25. Adequacy of Fidelity of Implementation: Grade-Level Meetings Attendance (Year 1)—School Level



Note. Adequate fidelity for a school was defined as holding at least 18 grade-level meetings (across grades) and at least 60% of teachers who attended at least 4 grade-level meetings across the school year; high fidelity for a school was defined as holding at least 25 grade-level meetings (across grades) and had at least 90% or more of teachers who attended at least 4 grade-level meetings across the school year.

Exhibit 26. Adequacy of Fidelity of Implementation: Leadership Academy Meetings (Year 1)—School Level



Note. Adequate fidelity for a school was defined as a principal or school designee attending at least three 3-hour leadership academy meetings; high fidelity was attending 4 meetings.

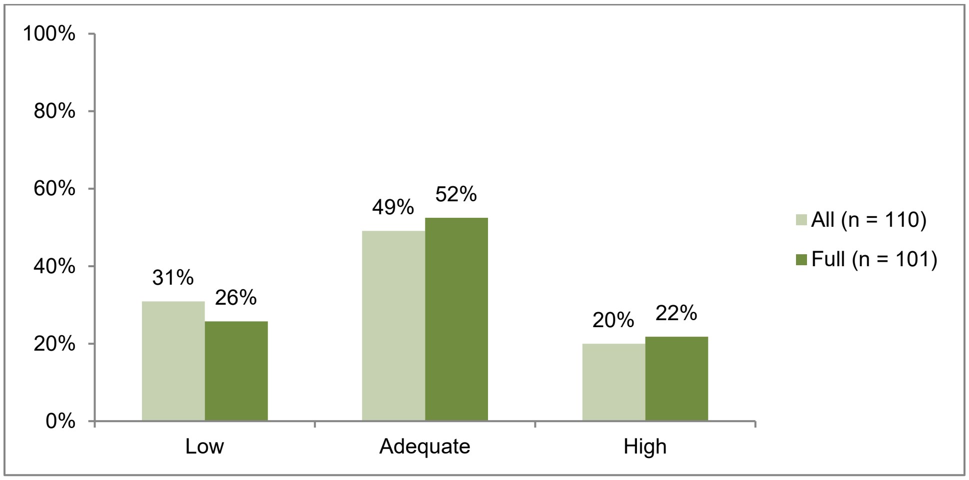
YEAR 2

Teacher Fidelity of Implementation

Three components of fidelity of implementation analyzed at the teacher level in Year 1 were also collected and analyzed in Year 2: coaching cycles implemented, grade-level meetings attended, and learning labs participation. Criteria for Year 2 are shown below in Exhibits 27 to 29. It is important to note that these criteria were different from that of Year 1; two of the components (coaching cycles implemented and grade-level meetings attended) were made more stringent and one (learning labs participation) was made less stringent. This was done to reflect the PD developers' intent that the focus of the work should shift toward the school locations in Year 2, and the belief that coaching and grade-level meetings in Year 2 would matter more to a school's success, while learning lab attendance would matter less.

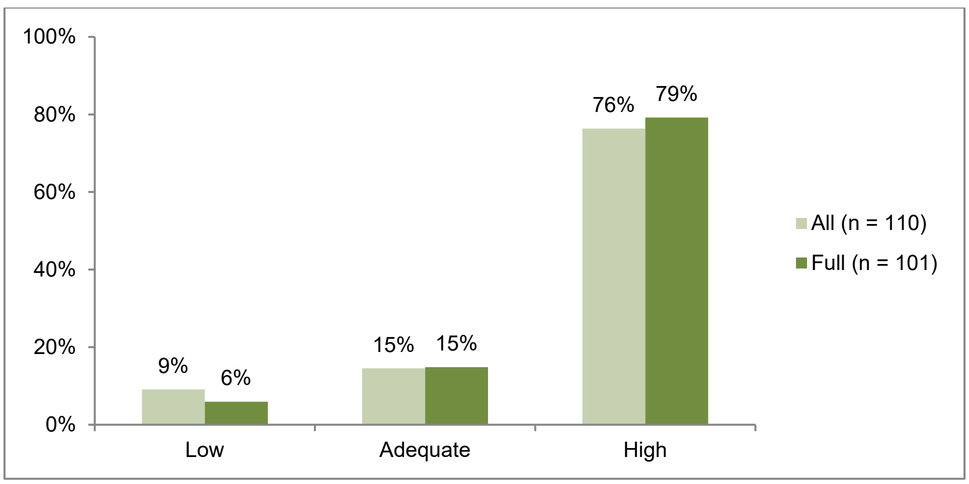
High fidelity of implementation was observed at the teacher level in Year 2 on two of the three components: grade-level meetings attended and learning labs participation. More specifically, roughly two-thirds of teachers had high-fidelity scores for grade-level meetings attended and more than half of teachers had high-fidelity scores for learning labs participation. Roughly half of the teachers had adequate fidelity for coaching cycles implemented. Findings from analyses using the two samples (all versus full) indicated a similar pattern of results. Slightly higher scores were observed across all three components in the full sample.

Exhibit 27. Adequacy of Fidelity of Implementation: Coaching (Year 2)—Teacher Level



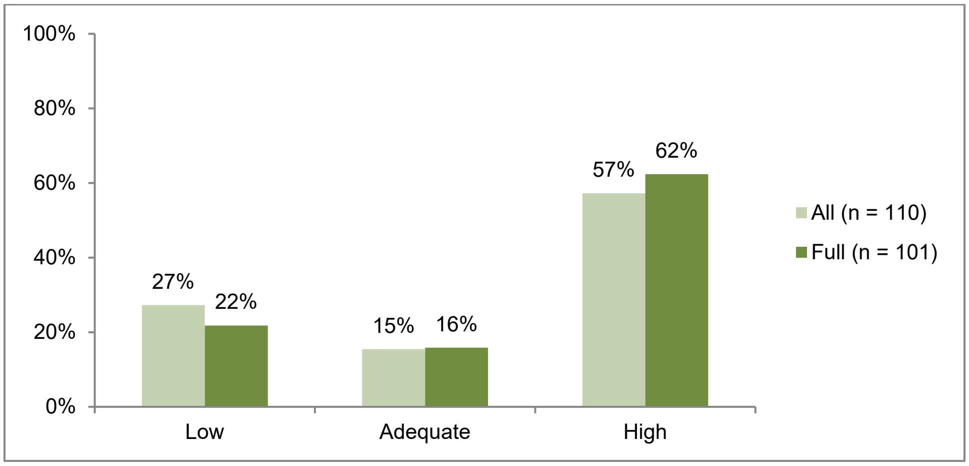
Note. Adequate fidelity was defined as teachers completing 4 coaching cycles; high fidelity was defined as teachers completing 5 or more coaching cycles.

Exhibit 28. Adequacy of Fidelity of Implementation: Grade-Level Meetings (Year 2)—Teacher Level



Note. Adequate fidelity was defined as teachers attending 5 to 6 meetings during the year; high fidelity was defined as teachers attending 7 meetings during the year.

Exhibit 29. Adequacy of Fidelity of Implementation: Learning Labs (Year 2)—Teacher Level



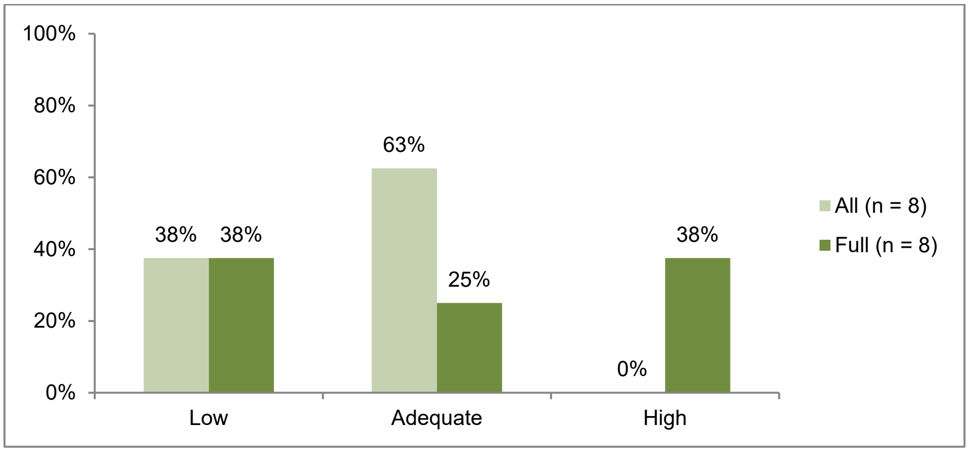
Note. Adequate fidelity was defined as teachers attending 13 to 16 hours of learning labs during the year; high fidelity was defined as teachers attending 17 to 20 hours of learning labs during the year.

School Fidelity of Implementation

Five components of fidelity of implementation analyzed at the school level in Year 1 were also collected and analyzed in Year 2: percentage of teachers who attended learning labs; number of grade-level meetings offered; percentage of teachers who attended grade-level meetings; percentage of teachers who participated in coaching; and number of leadership academy meetings offered.

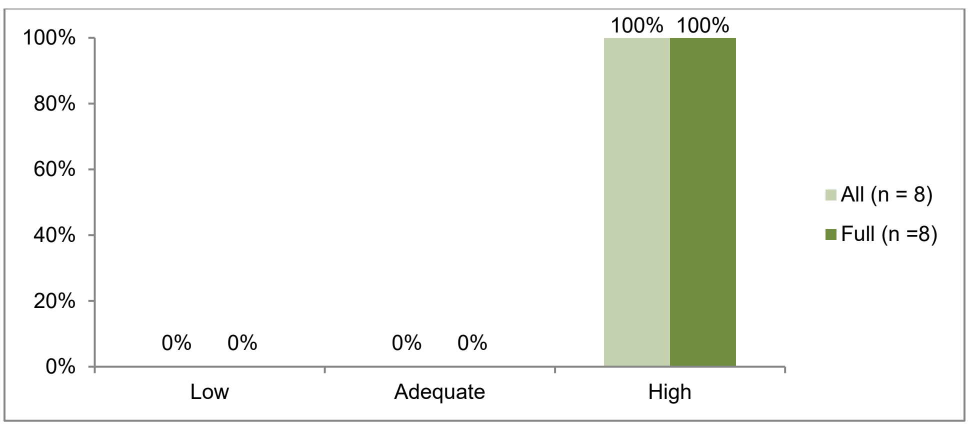
As in Year 1, findings for fidelity of implementation at the school level were mixed, as shown in Exhibits 30 to 34. For instance, high fidelity was observed for some of the components (e.g., number of grade-level meetings offered) but not others (e.g., number of leadership academy meetings offered). In addition, findings from analyses using the two samples (all versus full) showed similar patterns of results for some components but different patterns for others. For instance, findings were exactly the same across samples for number of grade-level meetings offered, percentage of teachers who engaged in coaching, and number of leadership academy meetings offered. However, for attendance at learning labs, 38% of schools had high attendance when including the sample of teachers who were available during the full school year but no schools met the high attendance criteria when including the sample that included all teachers. The percentage of teachers who attended grade-level meetings was high (75%) in the sample of teachers who were available during the full school year, but only half (50%) in the sample that included all teachers (Exhibit 32).

Exhibit 30. Adequacy of Fidelity of Implementation: Learning Labs (Year 2)—School Level



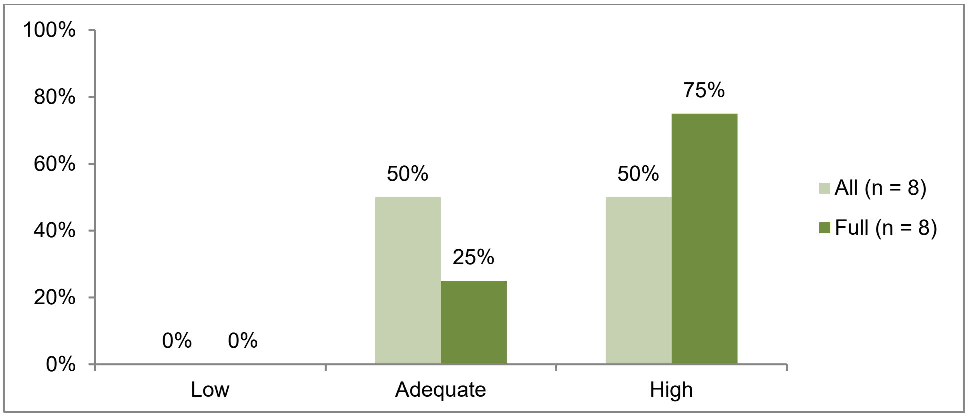
Note. Adequate fidelity for a school was defined as at least 70% of teachers with at least adequate attendance where adequate attendance was defined as attending at least 13 hours of learning labs during the year; high fidelity for a school was defined as at least 90% of teachers with at least adequate attendance.

Exhibit 31. Adequacy of Fidelity of Implementation: Grade-Level Meetings (Year 2)—School Level



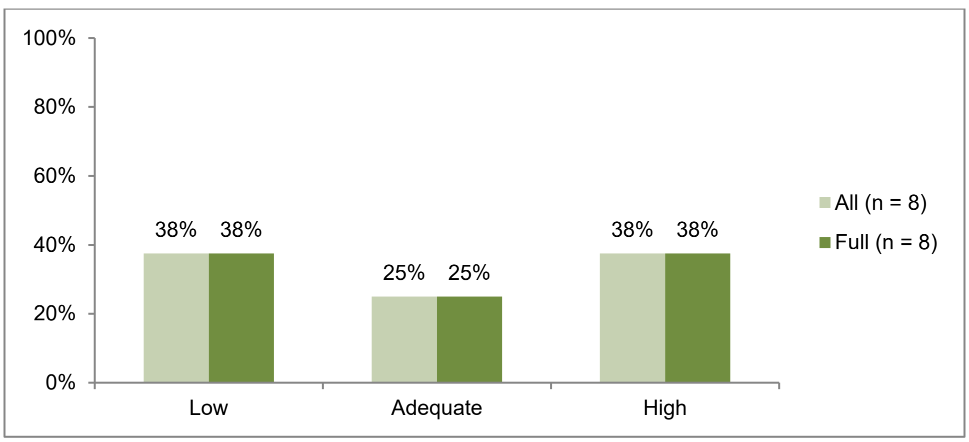
Note. Adequate fidelity for a school was defined as holding at least 23 grade-level meetings (across grades) and at least 60% of teachers who attended at least 5 grade-level meetings across the school year; high fidelity for a school was defined as holding at least 29 grade-level meetings (across grades) and had at least 90% or more of teachers who attended at least 5 grade-level meetings across the school year.

Exhibit 32. Adequacy of Fidelity of Implementation: Grade-Level Meetings Attendance (Year 2)—School Level



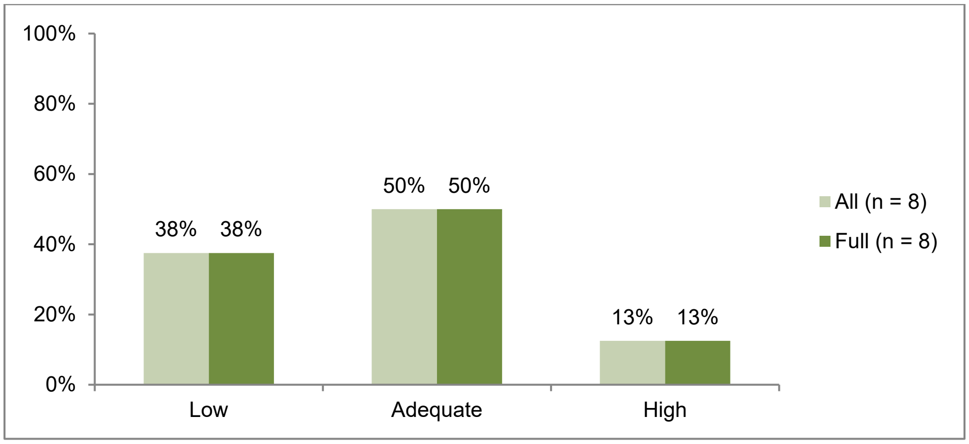
Note. Adequate fidelity for a school was defined as holding at least 23 grade-level meetings (across grades) and at least 60% of teachers who attended at least 5 grade-level meetings across the school year; high fidelity for a school was defined as holding at least 29 grade-level meetings (across grades) and had at least 90% or more of teachers who attended at least 5 grade-level meetings across the school year.

Exhibit 33. Adequacy of Fidelity of Implementation: Coaching (Year 2)—School Level



Note. Adequate fidelity for a school was defined as at least 60% of teachers with at least adequate coaching cycles completed where adequate coaching was defined as completing 4 or more coaching cycles; high fidelity for a school was defined as at least 90% teachers with at least adequate coaching cycles completed where adequate coaching was defined as completing 4 or more coaching cycles.

Exhibit 34. Adequacy of Fidelity of Implementation: Leadership Academy Meetings (Year 2)—School Level



Note. Adequate fidelity for a school was defined as a principal or school designee attending at least three 3-hour leadership academy meetings; high fidelity was attending 4 meetings.

Impact Findings

TEACHER OUTCOMES

A total of 183 out of 220 teachers provided information about their background characteristics at baseline. Exhibit 35 displays the grades teachers taught at baseline. A comparative descriptive analysis of selected demographic characteristics of teachers was conducted (see Exhibit 36). The intervention and comparison groups differed statistically on several variables. The intervention group had significantly more teachers with certifications or endorsements in Special Education (PK-21 and K-12). The comparison group had significantly more teachers who had a Bilingual/ESL endorsement. There were no significant differences between groups when it came to years of teaching experience [(intervention: $M = 12.24$, $SD = 9.19$; comparison: $M = 13.80$, $SD = 9.69$), $F(1, 1991) = 1.32$, $p = 0.25$]. While not statistically significant, the intervention group had more teachers who spoke English only, whereas more comparison teachers spoke Spanish. Comparisons between the groups on gender and race/ethnicity were not conducted due to the degree of missing data on each variable.

Exhibit 35. Teacher Grade-level Taught at Baseline

Grade Taught	Intervention (*n* = 115)		Comparison (*n* = 105)	
	n	%	*n*	%
Preschool	25	21.7	16	15.2
Kindergarten	22	19.1	26	24.8
Kindergarten-First Grade Split	2	1.7	1	1.0
First Grade	19	16.5	24	22.9
First-Second Grade Split	5	4.3	1	1.0
Second Grade	17	14.8	20	19.0
Second-Third Grade Split	3	2.6	0	0
Third Grade	22	19.1	17	16.2

Exhibit 36. Demographic Characteristics of Baseline Teacher Sample

Characteristic	Intervention (*n* = 115)		Comparison (*n* = 105)			
	n	%	*n*	%	$\chi2$	*p*
Gender						
Female	74	64.3	58	55.2	---	---
Missing	41	35.7	42	40.0	---	---
Race/Ethnicity						
Black	11	9.6	4	3.8	---	---
Asian	5	4.3	5	4.8	---	---
White	33	28.7	29	27.6	---	---
Hispanic	22	19.1	24	22.9	---	---
Other	2	1.7	0	0	---	---
Missing	42	36.5	42	40.0	---	---
Language Spoken[a]						
English Only	58	50.4	40	38.1	3.48	0.06
Spanish	36	31.3	44	41.9	3.18	0.08
Other	12	10.4	13	12.4	---	---
Education[a]						
Master's Degree	72	62.6	62	59.0	0.23	0.63
Certification/Endorsement[a]						
Early Childhood	36	31.3	27	25.7	0.78	0.38
Elementary	66	57.4	70	66.7	3.11	0.08
Early Childhood Special Education	13	11.3	2	1.9	7.57	<0.01
K-12 Special Education	16	13.9	4	3.8	6.72	0.01
Bilingual/ESL	33	28.7	46	43.8	6.32	0.01
Other	21	18.3	30	28.6	---	---
Years of Teaching Experience[a]						
1-5	32	27.8	20	19.0	---	---
6-10	19	16.5	22	21.0	---	---
11-20	33	28.7	30	28.6	---	---
21+	17	14.8	19	18.1	---	---

Note. Fourteen teachers from each condition did not provide this information.

Teacher outcome findings from each measure are described below. For each measure, descriptive data on the scores in the two groups over time are presented first. The evaluation team then examined baseline equivalence according to What Works Clearinghouse's (WWC) guidelines as described below. Then they conducted impact analyses examining changes from pretest to each posttest. Finally, they used Growth Mixture Modeling (GMM) to examine whether there were patterns of teacher change and whether these patterns were associated with the intervention and/or teacher characteristics.

Baseline equivalence test. In addition to a descriptive analysis conducted for intervention and comparison teachers in the analytic sample, the evaluation team followed the standards recommended by the Institute of Education Science's What Works Clearinghouse (U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse, 2013) to estimate baseline equivalence between the intervention and comparison group on the outcome measures. They calculated the standardized intervention-comparison difference (the difference in means for the intervention and comparison groups, divided by a pooled standard deviation) on all baseline teacher measures for each model. For six out of the 12 models, they found that the differences between intervention and comparison groups were smaller than 0.25, thereby demonstrating baseline equivalence between the intervention and the comparison groups. For the other six models where the differences between intervention and comparison groups exceeded 0.25, propensity score weighting was used to reestablish baseline equivalence. They controlled for pretest scores in the analyses.

Propensity score weighting. The propensity score is the predicted probability of participating in an intervention based on a set of potentially confounding covariates (i.e., teacher demographic characteristics and baseline teacher measures) using logistic regression. Propensity scoring attempts to equalize the mean values of potentially confounding observed covariates in the intervention and comparison groups, assuring that differences in outcomes between the intervention and covariate effect are not the result of differences in mean values of those covariates. Our analysis to estimate the average intervention effect on the treated (ATT) for the six models that showed intervention-comparison difference larger than 0.25 were adjusted for confounding using inverse propensity score estimators, as recommended by Curtis et al. (2007); Hirano et al. (2003); and Rosenbaum and Rubin (1983). Specifically, the weight for treated teachers was 1.0; the weight for comparison teachers was equal to pi/(1-pi), where pi is the propensity score for the i-th teacher. Weighting was selected over other approaches such as matching because of good performance in this dataset, flexibility with the distribution of the data, an ability to manage time-dependent covariates and censored data, and because it retains all sample members in the analysis. After propensity score weighting for comparison teachers, the evaluation team examined the standardized intervention-comparison score difference to assure that they were less than 0.25, thereby assuring baseline balance (U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse, 2013).

Impact on Teachers' Instructional Practices (HIS-EM)

Below we provide descriptive data on the *High-Impact Strategies for Early Mathematics* (HIS-EM) scores in the two groups over time, including information on the content and format of math instruction that was observed to provide context for the HIS-EM scores at each time point. We then describe the findings on the ABC-EM and PCK-EM measures for baseline equivalence and the impact for each of the posttest outcome time points. Finally, we present analyses that examined the trajectories of teachers over time using GMMs.

Descriptive Findings

Content and format of math instruction. Mathematics instruction in both conditions included the use of a number of different published curricula and teacher-developed materials. The most commonly used curriculum was Everyday Mathematics.[6] The instructional focus and format varied, with most teachers focusing more on number and operations concepts, and others focusing on particular aspects of geometry, data analysis, algebraic thinking, and measurement. Instruction was delivered through a variety of different instructional formats, including whole group, small group activities, partner work, and individual work.

Lesson duration. Duration of the observed math lessons ranged from 9 to 118 minutes across research conditions and time points. Exhibit 37 shows the minimum, maximum, mean, and standard deviations with regard to the duration of observed lessons from fall 2011 to spring 2015.

[6] One school used Trailblazers.

Exhibit 37. Math Lesson Duration in Minutes by Condition across Time Points

	Intervention				Comparison			
	n	Range	*M*	*SD*	*n*	Range	*M*	*SD*
Fall 2011	108	9–95	47.1	17.47	102	9–75	45.4	15.51
Spring 2012	96	13–118	46.5	19.50	96	15–116	47.6	17.75
Spring 2013	69	10–100	44.0	18.14	62	15–95	47.2	17.71
Spring 2014	57	15–90	45.9	19.87	47	14–85	47.3	15.25
Spring 2015	46	15–85	45.02	19.64	39	10– 5	47.1	17.28

Note. n = sample size; *M* = mean; *SD* = standard deviation.

Grade level. Observed math lessons were either at the preschool, kindergarten, first grade, 1-2 split, second grade, 2-3 split, or third-grade level. Exhibit 38 shows the grade-level distribution in percentages by condition across time points.

Exhibit 38. Percentage of Classrooms by Grade and Condition across Time Points

	Fall 2011		Spring 2012		Spring 2013		Spring 2014		Spring 2015	
	%		%		%		%		%	
Grade	I (*n* = 108)	C (*n* = 102)	I (*n* = 96)	C (*n* = 96)	I (*n* = 69)	C (*n* = 62)	I (*n* = 57)	C (*n* = 47)	I (*n* = 46)	C (*n* = 39)
Preschool	21.3	15.7	20.8	16.7	18.8	19.4	24.6	14.9	23.9	25.6
Kindergarten	19.4	24.5	20.8	22.9	26.1	19.4	19.3	27.7	19.6	28.2
First	16.7	23.5	20.8	21.9	15.9	24.2	26.3	27.7	26.1	23.1
Second	14.8	17.6	14.6	18.8	20.3	21	12.3	23.4	15.2	15.4
Third	19.4	16.7	19.8	17.7	13	12.9	15.8	6.4	15.2	7.7

Note. I = intervention group; C = comparison group; a small number of classrooms (between 1% and 5%) were K-1, 1-2, or 2-3 split classrooms. These classrooms are not included in the table.

Lessons of varying lengths by grade level and condition. Length of observed math lessons showed variations across time points by grade level and condition. On average, preschool and kindergarten lessons were shorter in both intervention and comparison groups across each time point. The longest lessons were observed at second and third grade in the intervention group (Exhibit 39).

Exhibit 39. Lessons of Varying Lengths by Grade Level and Condition across Time Points

	Fall 2011		Spring 2012		Spring 2013		Spring 2014		Spring 2015	
	M		*M*		*M*		*M*		*M*	
	I (*n* = 108)	C (*n* = 102)	I (*n* = 96)	C (*n* = 96)	I (*n* = 69)	C (*n* = 62)	I (*n* = 57)	C (*n* = 47)	I (*n* = 46)	C (*n* = 39)
Preschool	30.2	30.6	25.1	31.9	23.0	27.2	27.8	35.9	29	28
Kindergarten	34.5	41.6	36.4	47.0	43.1	44.6	41.5	46.4	34.4	55.3
First	47.5	45.2	46.5	49.9	46.5	50.4	48.5	49.5	47.9	46.9
Second	58.5	59.0	65.5	60.2	50.8	62.6	70.8	51.1	62.1	57.2
Third	61.3	55.0	63.0	53.2	43.5	57.6	58.8	53.7	61.7	61.7

Note. M = mean; I = intervention group; C = comparison group; *n* = sample size; a small number of classrooms (between 1% and 5%) were K-1, 1-2, or 2-3 split classrooms. These classrooms are not included in the table.

Content strand. The "major" content strand that was most frequently observed in all lessons was *Number and Operations* at each data collection period and across both conditions (Exhibit 40).

Exhibit 40. Content Strand Distribution by Condition across Time Points

	Fall 2011		Spring 2012		Spring 2013		Spring 2014		Spring 2015	
	%		%		%		%		%	
	I (*n* = 108)	C (*n* = 102)	I (*n* = 96)	C (*n* = 96)	I (*n* = 69)	C (*n* = 62)	I (*n* = 57)	C (*n* = 47)	I (*n* = 46)	C (*n* = 39)
Number and Operations	57.4	70.6	70.8	57.3	78.3	56.5	80.7	76.6	82.6	66.7
Measurement	15.7	11.8	14.6	18.8	5.8	21	10.5	6.4	-	10.3
Geometry	9.3	3.9	6.3	10.4	4.3	6.5	7	8.5	17.4	10.3
Data Analysis	7.4	7.8	4.2	7.3	1.4	3.2	-	6.4	-	10.3
Algebraic Thinking	10.2	5.9	4.2	6.3	10.1	12.9	1.8	2.1	-	2.6

Note. I = intervention group; C = comparison group; *n* = sample size.

Instructional grouping. The most frequently observed major instructional grouping was "Whole Group" at each data collection period and across both conditions (Exhibit 41).

Exhibit 41. "Major" Instructional Grouping of the Lesson by Condition across Time Points

	Fall 2011		Spring 2012		Spring 2013		Spring 2014		Spring 2015	
	%		%		%		%		%	
	I (*n* = 108)	C (*n* = 102)	I (*n* = 96)	C (*n* = 96)	I (*n* = 69)	C (*n* = 62)	I (*n* = 57)	C (*n* = 47)	I (*n* = 46)	C (*n* = 39)
Whole Group	63.9	80.4	64.6	77.1	58	77.4	56.1	63.8	58.7	69.2
Small Group	25	8.8	30.2	7.3	33.3	9.7	33.3	17	30.4	5.1
Individual	6.5	6.9	4.2	11.5	1.4	11.3	3.5	14.9	8.7	17.9
Partner	4.6	3.9	1	4.2	7.2	1.6	7	4.3	2.2	7.7

Note. I = intervention group; C = comparison group; *n* = sample size.

Language of instruction. In both comparison and intervention schools, most of the observed lessons were instructed in English at each time point (Exhibit 42).

Exhibit 42. Language of Instruction by Condition and across Time Points

	Fall 2011		Spring 2012		Spring 2013		Spring 2014		Spring 2015	
	%		%		%		%		%	
	I (*n* = 108)	C (*n* = 102)	I (*n* = 96)	C (*n* = 96)	I (*n* = 69)	C (*n* = 62)	I (*n* = 57)	C (*n* = 47)	I (*n* = 46)	C (*n* = 39)
English	88	85.3	92.7	83.3	85.5	82.3	87.7	76.6	76.1	74.4
Spanish	7.4	11.8	6.3	11.5	8.7	11.3	8.8	19.1	15.2	15.4
Bilingual	4.6	2.9	1	5.2	5.8	6.5	3.5	4.3	8.7	10.3

Note. I = intervention group; C = comparison group; *n* = sample size.

A total of 210 teachers out of 220 (95%) were observed at baseline in fall 2011 (108 teachers in the intervention group and 102 teachers in the comparison group). Across the 4 time points, a total of 783 observations were conducted. Of these observations, 210, 207, 205, 161, and 130 were completed in fall 2011, spring 2012, spring 2013, spring 2014, and spring 2015, respectively. With regard to teacher participants, 210 were observed at baseline, and 192, 131, 104, and 85 had data in spring 2012, spring 2013, spring 2014, and spring 2015, respectively. Only 85 teachers (39% of the original 220 teachers) participated in the entire study and were observed at all 4 time points.

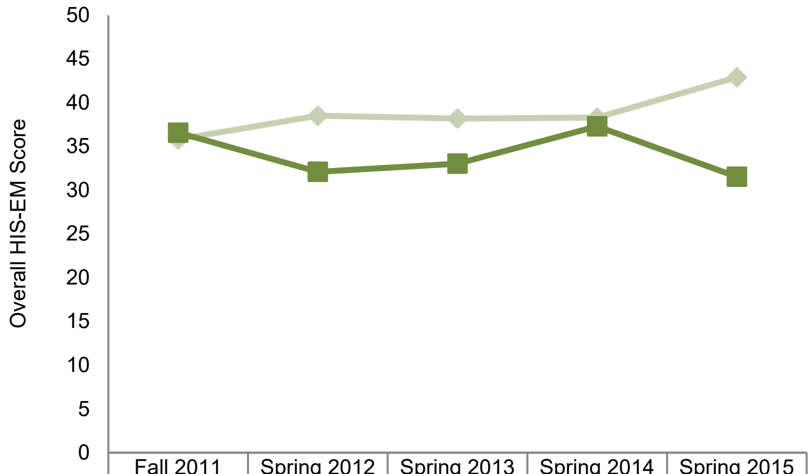
Exhibit 43 presents the mean and standard deviations of the overall scores for *What*, *Who*, and *How* domains, dimensions listed under each domain, and overall HIS-EM scores (sum of all scores received in each dimension) by condition (i.e., intervention and comparison) across time points. The number of points possible for dimensions, domains, and overall HIS-EM scores was 7, 21, and 63, respectively. Exhibit 44 shows the trends in HIS-EM scores for the intervention and comparison groups over the 4 years of the project.

Exhibit 43. HIS-EM Domain, Dimension, and Overall Mean Score (Standard Deviation) by Condition across Time Points

	Fall 2011 (*N* = 210)		Spring 2012 (*N* = 192)		Spring 2013 (*N* = 131)		Spring 2014 (*N* = 104)		Spring 2015 (*N* = 85)	
	I (*n* = 108)	C (*n* = 102)	I (*n* = 96)	C (*n* = 96)	I (*n* = 69)	C (*n* = 62)	I (*n* = 57)	C (*n* = 47)	I (*n* = 46)	C (*n* = 39)
WHAT	11.9 (4.3)	12.3 (3.6)	12.7 (3.3)	11.0 (3.4)	12.3 (4.0)	11.2 (3.8)	12.3 (3.5)	12.3 (3.7)	14.0 (4.1)	10.4 (4.4)
Learning Objectives	4.1 (1.5)	4.3 (1.3)	4.3 (1.2)	3.9 (1.2)	4.2 (1.4)	4.0 (1.4)	4.4 (1.3)	4.1 (1.3)	4.8 (1.4)	3.7 (1.5)
Math Representations	4.1 (1.5)	4.2 (1.4)	4.2 (1.2)	3.6 (1.2)	4.2 (1.5)	3.8 (1.4)	4.3 (1.3)	4.1 (1.1)	4.8 (1.4)	3.3 (1.5)
Concept Development	3.8 (1.6)	3.9 (1.3)	4.1 (1.2)	3.5 (1.2)	3.9 (1.4)	3.5 (1.4)	3.9 (1.4)	3.9 (1.2)	4.5 (1.5)	3.3 (1.6)
WHO	11.7 (4.5)	12.1 (3.7)	12.7 (3.5)	10.4 (3.2)	12.3 (3.9)	11.0 (4.2)	12.6 (3.7)	12.3 (4.1)	14.2 (4.5)	10.0 (4.2)
Att. to Dev. Trajectories	4.0 (1.5)	4.2 (1.2)	4.4 (1.2)	3.6 (1.3)	4.1 (1.4)	3.7 (1.4)	4.3 (1.5)	4.2 (1.4)	4.8 (1.6)	3.4 (1.5)
Response to Students' Ind. Needs	3.9 (1.6)	3.9 (1.4)	4.3 (1.3)	3.5 (1.2)	4.1 (1.5)	3.7 (1.5)	4.2 (1.4)	4.2 (1.3)	4.6 (1.6)	3.3 (1.5)
Dev. App. Learning Formats	3.8 (1.6)	3.9 (1.4)	4.1 (1.3)	3.3 (1.2)	4.1 (1.4)	3.6 (1.5)	4.2 (1.3)	4.0 (1.2)	4.8 (1.5)	3.3 (1.4)
HOW	12.1 (4.9)	12.2 (4.3)	13.0 (3.8)	10.6 (3.7)	12.7 (4.2)	11.3 (4.2)	13.0 (3.6)	12.5 (4.4)	14.7 (4.2)	11.2 (5.0)
Planning	3.9 (1.6)	4.1 (1.4)	4.2 (1.3)	3.5 (1.3)	4.2 (1.5)	3.8 (1.5)	4.2 (1.4)	4.4 (1.2)	4.9 (1.5)	4.0 (1.7)
Student Engagement	4.1 (1.7)	3.9 (1.5)	4.4 (1.4)	3.4 (1.3)	4.3 (1.5)	3.9 (1.4)	4.2 (1.5)	4.2 (1.3)	5.0 (1.5)	3.7 (1.8)
Establishing Math Learning Community	4.1 (1.7)	4.1 (1.5)	4.5 (1.4)	3.7 (1.5)	4.2 (1.6)	3.7 (1.6)	4.2 (1.6)	4.2 (1.4)	4.9 (1.5)	3.5 (1.8)
OVERALL HIS-EM	35.8 (13.4)	36.6 (11.2)	38.5 (10.1)	32.1 (10.0)	37.4 (11.6)	33.5 (11.7)	38.0 (11.9)	37.2 (9.9)	42.9 (12.5)	31.5 (13.1)

Note. I = intervention group; C = comparison group; *n* = sample size. HIS-EM = High-Impact Strategies for Early Mathematics.

Exhibit 44. Trends in HIS-EM Scores for Intervention and Comparison Groups across 4 Years



	Fall 2011	Spring 2012	Spring 2013	Spring 2014	Spring 2015
Intervention	35.81	38.52	38.19	38.3	42.91
Comparison	36.58	32.11	33.04	37.3	31.54

Baseline Equivalence

Exhibit 45 shows the baseline equivalence of the four samples of teachers (1-year impact, 2-year impact, 3-year impact, and 4-year impact). The analyses showed that the two samples were equivalent at baseline for all samples.

Impact Models (Fall 2011 to Spring 2015)

Exhibit 45 shows the impact estimates for HIS-EM over time. Estimates of the intervention impacts were derived from the ITT analyses. After controlling for the baseline teacher measure in fall 2011 and teacher demographic characteristics, the analysis showed that intervention teachers had higher posttest scores than comparison teachers after teachers participated for 1 year and 4 years of the intervention (1-year effect size = 0.65; 4-year effect size = 1.01).

Exhibit 45. HIS-EM Baseline Equivalence and Impact Estimates

Models		Sample Size				Pretest Equivalence			Posttest		HLM Model Results					
Pretest	Posttest	I Clusters (*N*)	I Teachers (*N*)	C Clusters (*N*)	I Teachers (*N*)	I Mean (*SD*)	C Mean (*SD*)	Standardized T-C Difference	I Mean (*SD*)	C Mean (*SD*)	Impact Estimate	Standardized Effect Size	Impact S.E.	*p* value	Degrees of Freedom	Improvement Index
HIS-EM 11	HIS-EM 12	8	97	8	96	4.0 (1.5)	4.1 (1.3)	-0.09	4.3 (1.1)	3.6 (1.1)	0.73	0.65	0.20	0.003	14	25
HIS-EM 11	HIS-EM 13	8	69	8	63	4.0 (1.5)	4.2 (1.2)	-0.21	4.2 (1.3)	3.8 (1.3)	0.52	0.40	0.25	0.057	14	15
HIS-EM 11	HIS-EM 14	8	56	7	51	3.9 (1.5)	4.2 (1.2)	-0.18	4.2 (1.2)	4.2 (1.3)	0.07	0.06	0.25	0.787	13	2
HIS-EM 11	HIS-EM 15	8	46	6	44	4.0 (1.5)	4.1 (1.2)	-0.16	4.8 (1.4)	3.5 (1.5)	1.45	1.01	0.30	0.0004	12	34

Note. Propensity score weighting was not needed to establish baseline equivalence. I = intervention group; C = comparison group; *n* = sample size. HIS-EM = High-Impact Strategies for Early Mathematics.

HIS-EM Teacher Trajectories

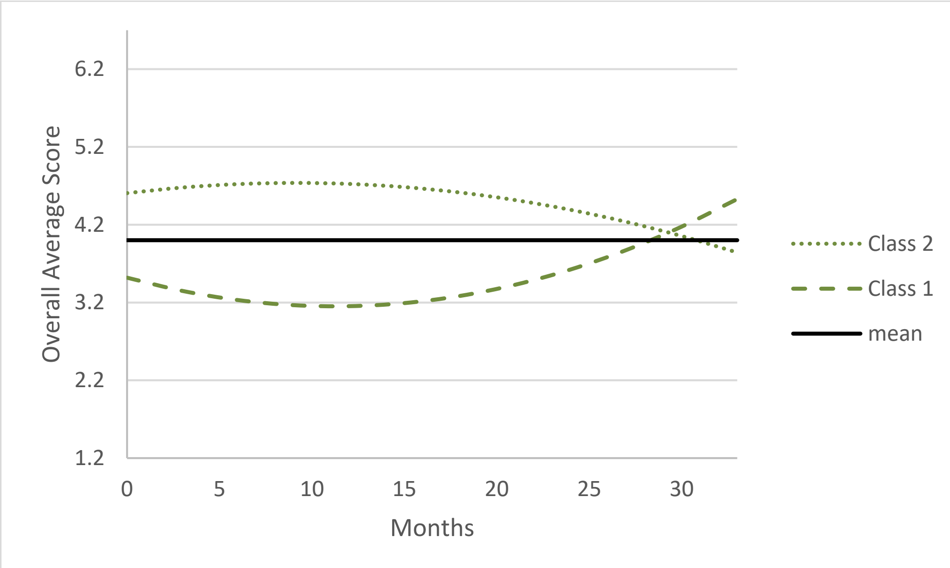
Growth Mixture Models (GMM) analyses were conducted to examine whether there were classes of teachers that emerged from fitting the HIS-EM data. Again, this approach allowed us to examine whether different teachers belonged to different subgroups or subpopulations. To conduct these analyses, teacher data collected at 4 time points were analyzed. Across intervention and comparison groups, 128 teachers had data at 3 time points: fall 2011 or baseline; spring 2012 or 1-year posttest; and spring 2013 or 2-year posttest. Approximately half of the sample was from each condition. Approximately 80% of the teachers in this analytic sample (50 in each condition) had data at 4 time points (fall 2011, spring 2012, spring 2013, and spring 2014). Data from all 4 time points were used in the GMM analyses. We did not use the 5th time point (because the sample was too small to analyze.)

There was a fair amount of variability in teachers' HIS-EM measured strategies at baseline as well as in the rate of change over time. It is important to note that the intervention and comparison teachers' scores on the HIS-EM were not statistically significantly different at baseline.

The model fit statistics suggest that a model that assumes two classes best fits the data; one class included teachers that start with high scores and decrease over time (Class 1) and another class included teachers that start with low scores and increase over time (Class 2; Exhibit 46). The intervention teachers were significantly more likely to be in the latter class ($p < .01$).

The evaluation team then examined whether specific teacher characteristics were associated with teacher trajectories. Teacher characteristics included in the model were grade, teacher certification category (e.g., elementary education, special education), second language fluency, age, gender, ethnicity, and number of years teaching at baseline. Only number of years teaching was significantly associated with teacher trajectory; more specifically, more years teaching was significantly associated with being in the class trajectory of starting with low scores and increasing over time ($p < .01$).

Exhibit 46. Growth Mixture Model Analysis: Classes of Teachers Based on HIS-EM Scores



Impact on Teacher Attitudes, Beliefs, and Confidence (ABC-EM)

Below we describe the findings on the ABC-EM measure at baseline and across the 4 years of the intervention. We also present analyses that examined the trajectories of teachers over time using GMMs.

Descriptive Findings

Exhibit 47 shows the means and standard deviations for each subscale by condition and time point.

Exhibit 47. Attitudes, Beliefs, and Confidence in Early Mathematics (ABC-EM) Subscale Scores by Condition across Time Points

	Intervention					Comparison				
Subscales	Fall 2011	Spring 2012	Spring 2013	Spring 2014	Spring 2015	Fall 2011	Spring 2012	Spring 2013	Spring 2014	Spring 2015
Confidence in Math Teaching										
Mean	6.9	7.6	7.8	8.0	8.5	7.5	7.8	7.8	8.0	8.0
Standard Deviation	1.6	1.3	1.2	1.0	1.0	1.7	1.5	1.5	1.4	1.3
N	94	93	105	77	59	83	78	94	69	66
Positive Math Attitudes										
Mean	6.2	6.5	6.5	6.0	7.1	6.8	6.7	6.8	6.1	6.9
Standard Deviation	1.9	1.8	1.8	1.5	1.7	1.9	2.0	1.8	1.6	1.9
N	94	93	105	77	59	83	78	94	69	66

Baseline Equivalence

At baseline, 177 out of 220 teachers (80%) completed the ABC-EM. Exhibit 48 shows the baseline equivalence of the three samples for each subscale score (1-year impact, 2-year impact, 3-year impact, and 4-year impact) after applying propensity score weighting for all comparisons except 1-year and 2-year impacts on the Attitudes subscale.

Impact Models (Fall 2011 to Spring 2015)

Exhibit 48 shows the impact estimates for ABC-EM for the 1-year, 2-year, 3-year, and 4-year impact samples. Estimates of the intervention impacts were derived from the ITT analyses. After controlling for the baseline teacher measure in fall 2011 and teacher demographic characteristics, the analysis shows that intervention teachers had higher confidence than comparison teachers after the intervention teachers participated in 2 years of the intervention (effect size = 0.51). One-year and 3-year impacts were not significant. The evaluation team also found a large intervention effect on confidence scores for the 4-year impact model with an effect size favoring intervention group of 0.71.

Exhibit 48. ABC-EM Baseline Equivalence and Impact Estimates

Models		Sample Size				Pretest Equivalence			Posttest		HLM Model Results					
Pretest	Posttest	I Clusters (*N*)	I Teachers (*N*)	C Clusters (*N*)	I Teachers (*N*)	I Mean (*SD*)	C Mean (*SD*)	Standardized T-C Difference	I Mean (*SD*)	C Mean (*SD*)	Impact Estimate	Standardized Effect Size	Impact S.E.	*p* value	Degrees of Freedom	Improvement Index
ABC Att 11	ABC Att 12	8	83	8	67	6.4 (1.9)	6.7 (1.9)	-0.20	6.4 (1.8)	6.9 (1.9)	-0.21	-0.11	0.18	0.263	14	-8
ABC Att 11	ABC Att 13	8	65	8	49	6.3 (1.8)	6.8 (2.0)	-0.23	6.5 (1.8)	6.9 (1.9)	-0.09	-0.05	0.20	0.683	14	-9
ABC Att 11[a]	ABC Att 14	8	51	6	33	6.3 (1.7)	6.2 (2.3)	0.04	5.9 (1.5)	6.2 (1.7)	0.07	0.04	0.19	0.738	12	2
ABC Att 11[a]	ABC Att 15	7	43	6	33	6.3 (1.6)	6.1 (2.3)	0.09	7.1 (1.6)	6.2 (2.2)	0.64	0.34	0.32	0.069	11	13
ABC Conf 11[a]	ABC Conf 12	8	83	8	67	6.9 (1.6)	6.8 (2.0)	0.13	7.6 (1.3)	7.4 (1.7)	0.12	0.08	0.18	0.535	14	3
ABC Conf 11[a]	ABC Conf 13	8	65	8	49	7.0 (1.6)	6.7 (1.8)	0.24	8.1 (1.2)	7.2 (1.8)	0.73	0.51	0.29	0.026	14	20
ABC Conf 11[a]	ABC Conf 14	8	51	6	33	7.0 (1.6)	6.8 (1.9)	0.12	7.9 (0.9)	7.6 (1.7)	0.23	0.23	0.21	0.289	12	7
ABC Conf 11[a]	ABC Conf 15	7	43	6	33	6.9 (1.6)	6.7 (1.9)	0.09	8.4 (1.0)	7.5 (1.4)	0.85	0.71	0.32	0.023	11	26

Note. I = intervention group; C = comparison group; *n* = sample size; ABC-EM = Attitudes, Beliefs, and Confidence for Early Mathematics; ABC Att = Attitudes; ABC Conf = Confidence.

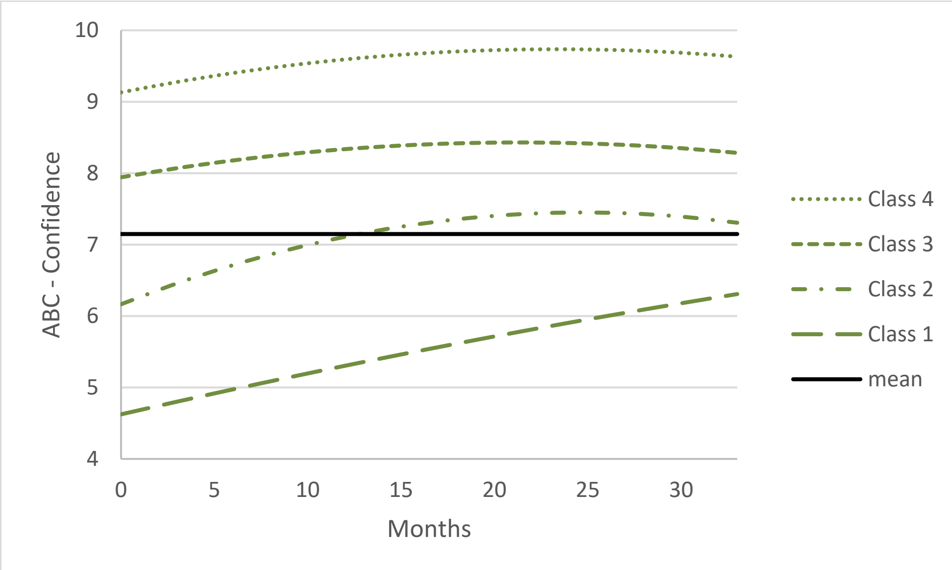
[a]Propensity score weighting was used to establish baseline equivalence.

ABC-EM Teacher Trajectories

GMM analyses were conducted to examine whether there were classes of teachers that emerged from fitting the ABC-EM data [note: we fit models for two subscores: the ABC Confidence in Math Teaching (Confidence) score and the ABC Positive Math Attitudes (Attitude) score]. This approach allowed us to examine whether different teachers belonged to different subgroups or subpopulations. To conduct these analyses we analyzed teacher data collected at 4 time points. Across intervention and comparison groups, 128 teachers had data at the following 3 time points: fall 2011 or baseline; spring 2012 or 1-year posttest; and spring 2013 or 2-year posttest. Approximately half of the sample was from each condition. Approximately 80% of the teachers in this analytic sample (50 in each condition) had data at 4 time points (fall 2011, spring 2012, spring 2013, and spring 2014). Data from all 4 time points were used in the GMM analyses. The evaluation team did not use the 5th time point as the sample was too small to analyze.

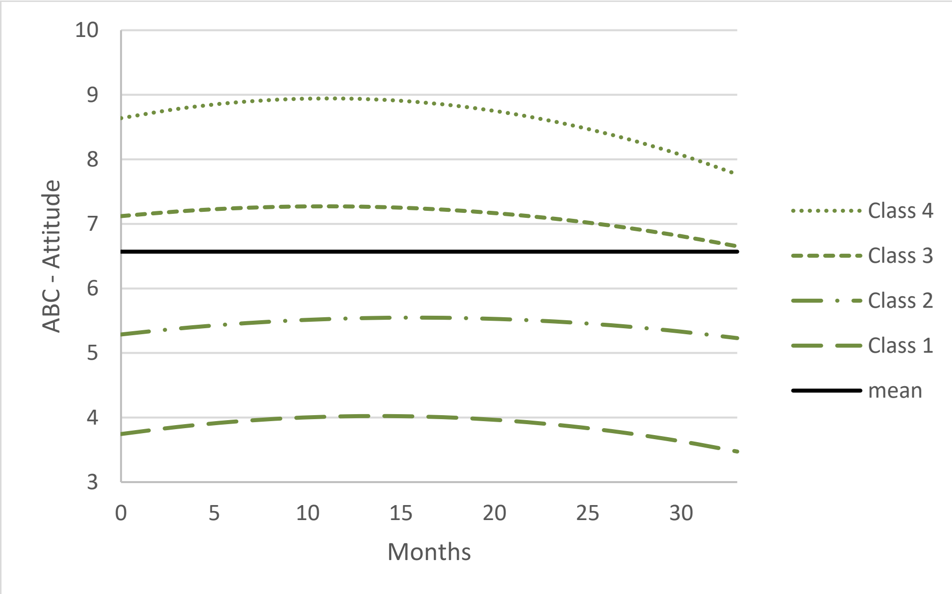
The model fit statistics for ABC confidence scores suggest a model that assumes four classes best fits the data; all four classes exhibited some growth over time, but initiated their trajectories at different levels (Exhibit 49). A large majority (~75%) of the sample was nearly equally split across Classes 2 and 3, while about 25% of the teachers were split between Classes 1 and 4. Relative to teachers without early childhood certification, teachers who are certified in early childhood education were less likely to be in Class 4 and more likely to be in Class 1 ($p < .05$). Additionally, relative to teachers who speak a second language, teachers who only speak English were less likely to be in Class 3 and more likely to be in Class 1 ($p < .05$). Finally, intervention teachers were more likely to be in Class 2 ($p < .05$).

Exhibit 49. Growth Mixture Model Analysis: Classes of Teachers Based on ABC-EM Confidence Scores



When examining the ABC attitudes scores, model fit statistics again suggested that a model that assumes four classes best fits the data. For ABC attitudes, however, classes did not show much improvement over time (see Exhibit 50). About 15% of the sample belonged to the class with lowest attitudes (i.e., Class 1). The remainder of the sample is split almost equally among the remaining three classes. Teachers with more years of experience were more likely to be in Class 3 and less likely to be in Class 1. Additionally, relative to teachers without early childhood certification, teachers who are certified in early childhood education were less likely to be in Classes 3 and 4, and more likely to be in Class 1 ($p < .01$; $p < .05$). Intervention teachers were no more or less likely than expected to belong to any of the trajectory classes.

Exhibit 50. Growth Mixture Model Analysis: Classes of Teachers Based on ABC-EM Attitudes Scores



Impact on Teacher Knowledge (PCK-EM)

Below we describe the findings on the *Pedagogical Content Knowledge in Early Mathematics* (PCK-EM) measure for baseline equivalence and the impact for each of the posttest outcome time points. We also present analyses that examined the trajectories of teachers over time using GMMs.

Descriptive Findings

Exhibit 51 shows descriptive statistics for PCK-EM collected at baseline.

Exhibit 51. Baseline Descriptive Statistics of Pedagogical Content Knowledge in Early Mathematics *(PCK-EM) Scores*

	What (Knowledge of foundational math)	Who (Knowledge of young children learning math)	How (Knowledge of math-specific pedagogy)
Mean	2.5	2.3	2.3
SD	0.65	0.7	0.8
Minimum	1	1	1
Maximum	5	4	4.5
Percentiles			
25%	2.0	2.0	1.9
50%	2.5	2.5	2.5
75%	3.0	2.5	3.0

Baseline Equivalence

At baseline, 154 out of 220 teachers (70%) completed the PCK-EM. Exhibit 49 shows descriptive statistics for PCK-EM collected at baseline. Exhibit 52 shows the baseline equivalence of the two samples (i.e., the 1-year impact sample and 3-year impact sample).

Impact Models (Fall 2011 to Spring 2014)

Exhibit 52 shows the impact estimates for PCK-EM Number 7 and Fraction scores for the 1-year impact and 3-year impact. Estimates of the intervention impacts were derived from the ITT analyses. After controlling for the baseline teacher measure in fall 2011 and teacher demographic characteristics, the analysis showed that intervention teachers had higher *Number 7* posttest scores than comparison teachers after teachers participated in 3 years of the intervention (effect size = 0.48). The 1-year impact was not significant for *Number 7* scores, and neither the 1-year nor the 3-year impacts were significant for *Fraction* scores.

Exhibit 52. Baseline Equivalence and Impact Estimates for PCK-EM Number 7 and Fraction Scores

Models		Sample Size				Pretest Equivalence			Posttest		HLM Model Results					
Pretest	Posttest	I Clusters (*N*)	I Teachers (*N*)	C Clusters (*N*)	C Teachers (*N*)	I Mean (*SD*)	C Mean (*SD*)	Standardized T-C Difference	I Mean (*SD*)	C Mean (*SD*)	Impact Estimate	Standardized Effect Size	Impact S.E.	*p* value	Degrees of Freedom	Improvement Index
PCK-EM 11 Number 7[a]	PCK-EM 12 Number 7	8	83	8	71	2.5 (0.6)	2.5 (0.6)	-0.02	2.5 (0.6)	2.4 (0.6)	0.12	0.19	0.08	0.157	14	8
PCK-EM 11 Number 7[a]	PCK-EM 14 Number 7	8	51	5	25	2.5 (0.6)	2.4 (0.6)	0.11	2.9 (0.8)	2.5 (0.8)	0.37	0.48	0.15	0.031	11	18
PCK-EM 11 fraction[a]	PCK-EM 12 fraction	8	83	8	67	2.8 (0.7)	2.9 (0.9)	-0.06	2.7 (0.6)	2.7 (0.7)	0.02	0.17	0.11	0.878	14	7
PCK-EM 11 fraction[a]	PCK-EM 14 fraction	8	51	5	33	2.7 (0.7)	2.7 (0.8)	0.09	2.9 (0.7)	2.7 (0.6)	0.23	0.35	0.13	0.093	12	14

Note. I = intervention group; C = comparison group; *n* = sample size; PCK-EM = Pedagogical Content Knowledge in Early Mathematics.

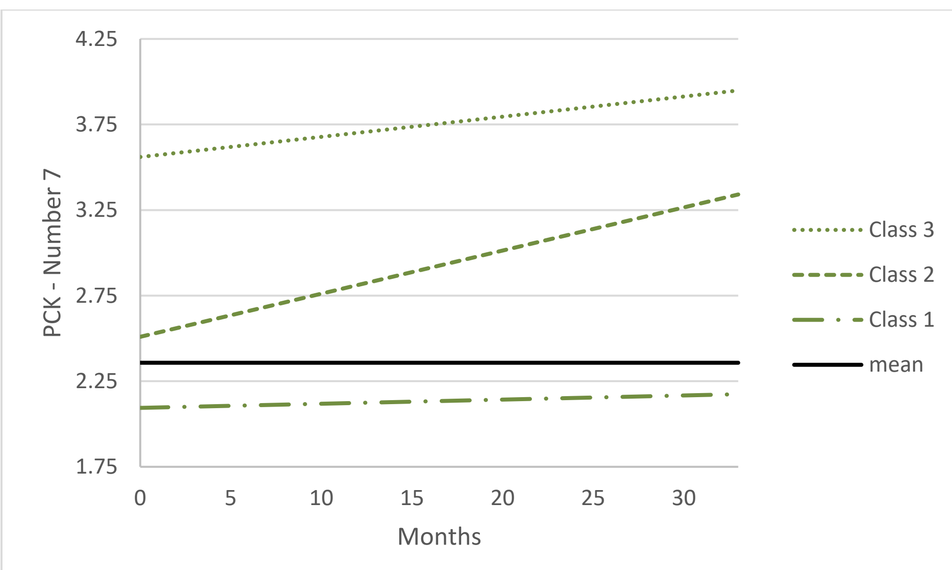
[a]Propensity score weighting was used to establish baseline equivalence.

PCK-EM Teacher Trajectories

GMM analyses were conducted to examine whether there were classes of teachers that emerged from fitting the PCK-EM data (note: we fit models for two subscores: the PCK number 7 score and the PCK fraction score). This approach allowed us to examine whether different teachers belonged to different subgroups or subpopulations. To conduct these analyses, teacher data collected at 3 time points were analyzed. Approximately 80% of the teachers in this analytic sample (50 in each condition) had data at 3 time points (fall 2011 or baseline, spring 2012 or 1-year posttest, and spring 2014 or 3-year posttest). The evaluation team did not use the 4th time point (i.e., spring 2015) as the sample was too small to analyze.

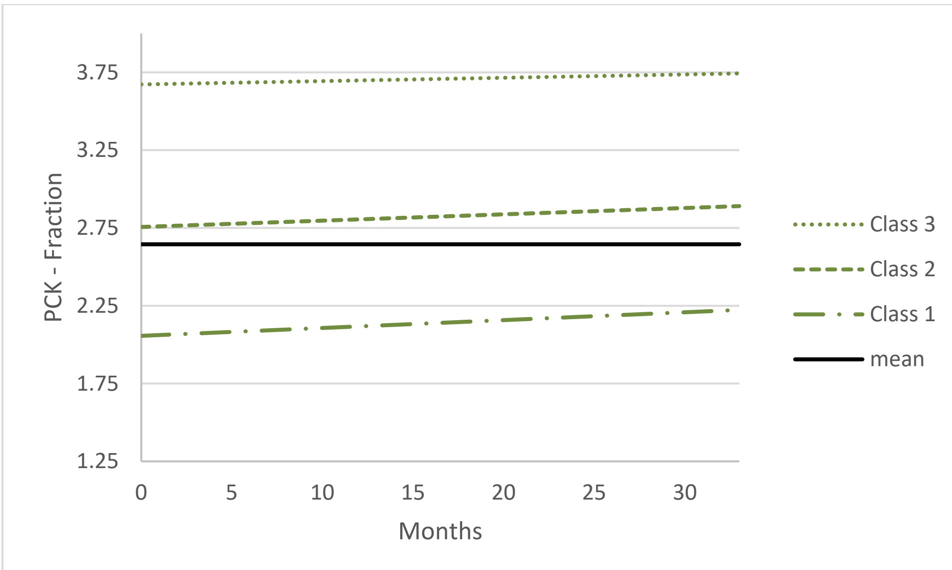
When examining the PCK *Number 7* outcome, model fit statistics suggested that three classes of trajectories exist in the sample (Exhibit 53). Class 1 begins with below average knowledge scores and remains nearly flat over time. Class 2 begins with average knowledge scores and increases over time. Class 3 begins with higher than average knowledge scores and increases over time. About 60% of the teachers in the sample belong to Class 1, 35% to Class 2, and only 5% to 6% to Class 3. No teacher characteristics were associated with trajectory class. Intervention teachers were more likely to be in Class 3 ($p < .05$).

Exhibit 53. Growth Mixture Model Analysis: Classes of Teachers Based on PCK-EM Number 7 Scores



Finally, when examining the PCK *Fraction* outcome, model fit statistics again suggested that three classes of trajectories exist in the sample. All three classes showed very small improvement over time (Exhibit 54). Class 1 begins with below average knowledge scores, Class 2 begins with average knowledge scores, and Class 3 begins with higher than average knowledge scores. About 35% of the teachers in the sample belonged to Class 1, 56% to Class 2, and only 9% to 10% to Class 3. When examining how teacher characteristics are related to trajectory class, teachers holding master's degrees were more likely to belong to Class 3 and less likely to belong to Class 1 ($p < .05$). Again, intervention teachers were more likely to be in Class 3 ($p < .01$).

Exhibit 54. Growth Mixture Model Analysis: Classes of Teachers Based on PCK-EM Fraction Scores



DESCRIPTIVE ANALYSIS AND BASELINE EQUIVALENCE

Exhibit 55 shows descriptive analysis and tests of baseline equivalence of the child participants in the intervention and comparison groups for the 1-year analytic sample (those who had both fall 2011 and spring 2012 scores). Exhibit 56 shows baseline equivalence test results of the participants in the intervention and comparison groups for the 2-year analytic sample (those who had both fall 2011 and spring 2013 scores). A series of HLM models using SAS (version 9.3 for Windows) were conducted to examine baseline equivalence. Results from these analyses indicated that children in the intervention group were not significantly different from children in the comparison group on demographics variables or pretest mathematics scores for either the 1-year (Exhibit 57) or 2-year impact analyses (Exhibit 58).

Exhibit 55. Baseline Equivalence Test Results for Children in the Intervention and Comparison Groups for Demographic Characteristics, WJ III Applied Problems Pretest Standard Scores, and TEAM Pretest T-Score for the 1-Year Analytic Sample

Variable	Intervention		Comparison		*T*	*p*
	% or *M* (*SD*)	*N*	% or *M* (*SD*)	*N*		
Gender						
Male	49.4%	357	49.9%	341	-0.03	0.978
Female	50.6%	366	50.2%	343	0.03	0.978
Grade Level						
PreK	20.1%	145	20.5%	140	0.27	0.789
K	22.1%	159	26.5%	181	-0.52	0.611
1	19.3%	139	22.8%	156	-0.94	0.363
2	18.1%	130	13.9%	95	0.74	0.472
3	20.4%	147	16.4%	112	0.27	0.788
Age	79.3 (17.9)	720	77.5 (17.2)	684	0.49	0.838
Home Language – English	86.9%	626	86.4%	591	0.00	1.00
WJ III Applied Problems Standard Score at Pretest	94.7 (12.8)	720	94.9 (13.4)	684	-0.69	0.647
TEAM T-score at Pretest	28.1 (20.5)	704	27.5 (19.4)	663	0.22	0.831

Note. Standard deviations for continuous variables are in parentheses. Demographic variable descriptive analysis used the WJ III Applied Problems analytic sample.

Exhibit 56. Baseline Equivalence Test Results for Intervention and Comparison Children for Demographic Characteristics, WJ III Applied Problems Pretest Standard Scores, and TEAM Pretest T-Scores for the 2-Year Analytic Sample

Variable	Intervention		Comparison		*β*	*s.e.*	*p*
	% or *M* (*SD*)	*N*	% or *M* (*SD*)	*N*			
Gender							
Boys	49.4	219	49.8	229	-0.01	0.04	0.762
Girls	50.6	224	50.2	231	0.01	0.04	0.762
Grade level at pretest							
PreK	23.7	105	21.7	100	0.02	0.06	0.702
K	26.4	117	34.1	157	-0.05	0.06	0.348
1	24.8	110	28.0	129	-0.04	0.05	0.464
2	25.1	111	16.1	74	0.07	0.06	0.245
Age at pretest	73.6 (14.2)	443	72.2 (12.9)	460	0.85	2.05	0.685
Home language – English	84.9	376	84.8	390	-0.01	0.09	0.923
WJ III Applied Problems standard score at pretest	95.7 (12.4)	443	96.0 (13.1)	460	-1.10	1.72	0.533
TEAM T-score at pretest	23.4 (18.8)	424	23.7 (17.2)	426	-1.06	3.13	0.739

Note. Standard deviations for continuous variables are in parentheses. Demographic variable descriptive analysis used the WJ III Applied Problems analytic sample.

IMPACT ON MATH ACHIEVEMENT

One-Year Child Outcomes Impact (Fall 2011 to Spring 2012)

Primary estimates of the intervention impacts were derived from the ITT analyses. Regardless of fidelity of implementation, these analyses compared all children in intervention schools to all those who were in comparison schools. The 1-year impact analysis results indicated that there was no significant effect of the PD intervention on children's WJ-AP and TEAM scores (Exhibit 57), relative to children in the comparison condition. The intervention effect was consistent regardless of whether or not baseline demographics were included in the model (Model A vs. Model B).

Exhibit 57. Results for 1-Year Intent-to-Treat (ITT) Impact Analysis for WJ Applied Problems Standard Scores and TEAM T-Scores

Outcomes	Intervention				Comparison				Estimated Impact (s.e.)	Effect Size (Hedge's *g*)	Improvement Index	*p*
	M	Model-Adjusted *M*	*SD*	*N*	*M*	Model-Adjusted *M*	*SD*	*N*				
WJ Applied Problems Standard Score												
Model A	97.0	96.8	13.5	720	96.5	96.1	13.4	684	0.56 (1.12)	0.04	2	0.626
Model B	97.0	96.7	13.5	720	96.5	96.2	13.4	684	0.47 (1.10)	0.03	1	0.677
Model C	97.0	97.4	13.5	720	96.5	96.2	13.4	684	1.14 (2.65)	0.08	3	0.673
TEAM T-Score												
Model A	35.2	34.6	18.8	704	33.9	34.2	18.2	663	0.44 (1.23)	0.02	1	0.728
Model B	35.2	34.4	18.8	704	33.9	34.2	18.2	663	0.14 (1.13)	0.01	0	0.906
Model C	35.2	38.6	18.8	704	33.9	34.2	18.2	663	4.44 (3.93)	0.24	9	0.278

Note. Two-level HLM models were used where children were nested in schools. For the 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. Effect size measures the change (measured in standard deviations) in an average student's outcome that can be expected if the student is exposed to 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 exposed to the intervention. Improvement Index is a way to translate the effect size into a meaningful metric in educational research. WWC 2.0 recommends translating the effect size into "improvement in percentile rank," which is supposed to indicate the expected change in percentile rank for the median comparison students if that student had received the early math intervention.

Model-adjustment mean for comparison group = the estimated mean of comparison group, which is the intercept of the HLM model.

Model-adjustment mean for intervention group = the estimated mean of the intervention group, which is the sum of intercept and coefficient for the intervention indicator variable.

Estimated impact and standard errors are the coefficient and standard errors associated with intervention variable from the 2-level HLM model (children nested in school model).

Effect size = Estimated impact/pooled standard deviations of the intervention and comparison group.

Model A for Applied Problem = HLM impact models controlling for pretest scores.

Model B for Applied Problem = HLM impact model controlling for pretest, gender, grade, and language spoken at home.

Model C for Applied Problem = HLM impact model controlling for pretest, gender, age, grade, language spoken at home, school-level mean score at baseline, an interaction term between intervention and pretest, and an interaction term between intervention and age.

Model A for TEAM T-Scores = HLM impact models controlling for pretest scores.

Model B for TEAM T-Scores = HLM impact model controlling for pretest, gender, age, grade, and language spoken at home.

Model C for TEAM T-Scores = HLM impact model controlling for pretest, gender, age, grade, language spoken at home, school-level mean score at baseline, an interaction term between intervention and pretest, and an interaction term between intervention and age.

All the predictors except the intervention indicator variable were centered in the HLM model.

Two-Year Child Outcomes Impact (Fall 2011 to Spring 2013)

The 2-year impact analyses include results for the child outcomes at the 2-year posttest (spring 2013) when the children had been enrolled in the intervention schools for 2 years and moved from one grade to the next.[7] Of the 1,551 children who had pretest assessments, the evaluation team was able to find and assess 903 (or 58% of the original pretest sample). Generally speaking, children who were boys, older, had higher pretest scores, or whose home language was Spanish scored higher on Applied Problems at posttest. Similarly, older children scored higher on TEAM than their younger peers. Note that for the 2-year impact analyses, the results are shown separately for Applied Problems (Exhibit 58) and TEAM (Exhibit 59).

Although the 2-year impact analysis did not find an effect of the intervention regardless of whether or not child demographic characteristics were controlled for, results from interaction analysis (Model C in Exhibit 58) indicated there was a significant intervention by age interaction effect for the Applied Problems subtest, suggesting that intervention had a differential effect for children in different age groups. To further understand the intervention by age interaction effect, we ran additional HLM models for children with child age centered at 48, 60, 72, 84, or 96 months at the start of the intervention (Exhibits 60 and 61). For example, by centering children's age at 48 months, the intercept represents the estimated average score for the comparison group at 48 months old, and the coefficient associated with the intervention indicator indicates the difference between intervention and comparison group at 48 months old. For models where age was centered at 48 months old, the evaluation team detected that children who received the intervention for 2 years had significantly higher Applied Problems scores ($p < 0.05$, effect size = 0.30) (Exhibit 60). Exhibits 62 and 63 visually display the relationship between age and child outcome test scores for intervention and comparison group for WJ-AP and TEAM scores, respectively.

[7] Exhibits 57 to 59 present the HLM coefficient and standard errors estimate of the impact of treatment, each child covariates, and treatment by age interaction on child outcome. Exhibit 56 concisely displays the coefficient and standard errors of treatment indicator variable.

Exhibit 58. Results for 2-Year Intent-to-Treat Impact Analysis for WJ Applied Problems Standard Scores

Predictors	Model A			Model B			Model C		
	β	*s.e.*	*p*	*B*	*s.e.*	*p*	*β*	*s.e.*	*p*
Intercept	96.08	0.86	<.0001	96.61	0.78	<.0001	96.98	0.83	<.0001
Intervention	0.92	1.21	0.459	0.82	1.08	0.460	0.18	1.16	0.881
Pretest	0.55	0.03	<.0001	0.60	0.03	<.0001	0.60	0.03	<.0001
Boy				2.42	0.74	0.001	2.35	0.74	0.002
Age at pretest				0.11	0.03	0.0001	0.19	0.04	<.0001
Home language – English				-4.40	1.13	<.0001	-4.53	1.13	<.0001
Intervention by age interaction							-0.13	0.06	0.022
Intervention effect size	0.07			0.06			0.01		
Intervention improvement index	3			2			0		

Note. Two-level HLM models were used where children were nested in schools. Estimated impact (*β*) and standard errors (*s.e.*) are the coefficient and standard errors associated with intervention variable from the 2-level HLM model (children nested in school model). For the 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. Effect size measures the change (measured in standard deviations) in an average student's outcome that can be expected if the student is exposed to 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 exposed to the intervention. Improvement Index is a way to translate the effect size into a meaningful metric in educational research. WWC (2008) recommends translating the effect size into "improvement in percentile rank," which is supposed to indicate the expected change in percentile rank for the median comparison children if that student had received the early math intervention.

Effect size = Estimated impact/pooled standard deviations of the intervention and comparison group.

Model A for Applied Problem = HLM impact models controlling for pretest scores.

Model B for Applied Problem = HLM impact model controlling for pretest, gender, age, and language spoken at home.

Model C for Applied Problem = HLM impact model controlling for pretest, gender, age, language spoken at home, and an interaction term between intervention and age.

All the predictors except the intervention indicator variable were centered in the HLM model.

Exhibit 59. Results for 2-Year Intent-to-Treat Impact Analysis for TEAM T-Scores

Predictors	Model A			Model B			Model C		
	β	*s.e.*	*p*	*β*	*s.e.*	*p*	*β*	*s.e.*	*p*
Intercept	45.65	0.55	<.0001	44.29	0.80	<.0001	44.33	0.81	<.0001
Intervention	-0.41	0.77	0.606	-0.95	0.75	0.223	-1.03	0.79	0.210
Pretest	0.60	0.02	<.0001	0.38	0.03	<.0001	0.38	0.03	<.0001
Boy				0.93	0.57	0.105	0.92	0.58	0.108
Grade K at pretest				5.98	1.18	<.0001	5.98	1.18	<.0001
Grade 1 at pretest				5.94	1.85	0.001	5.93	1.85	0.001
Grade 2 at pretest				11.71	2.54	<.0001	11.71	2.54	<.0001
Age at pretest				0.09	0.06	0.164	0.10	0.07	0.153
Home language – English				0.03	0.85	0.975	0.01	0.85	0.991
Intervention by age interaction							-0.01	0.04	0.731
Intervention effect size	-0.03			-0.07			-0.07		
Intervention improvement index	-1			-3			-3		

Note. Two-level HLM models were used where children were nested in schools. Estimated impact (*β*) and standard errors (*s.e.*) are the coefficient and standard errors associated with intervention variable from the 2-level HLM model (children nested in school model). For the 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. Effect size measures the change (measured in standard deviations) in an average student's outcome that can be expected if the student is exposed to 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 exposed to the intervention. Improvement Index is a way to translate the effect size into a meaningful metric in educational research. WWC (2008) recommends translating the effect size into "improvement in percentile rank," which is supposed to indicate the expected change in percentile rank for the median comparison children if that student had received the early math intervention.

Effect size = Estimated impact/pooled standard deviations of the intervention and comparison group.

Model A for TEAM T-Scores = HLM impact models controlling for pretest scores.

Model B for TEAM T-Scores = HLM impact model controlling for pretest, gender, age, and language spoken at home.

Model C for TEAM T-Scores = HLM impact model controlling for pretest, gender, age, language spoken at home, and an interaction term between intervention and age.

All the predictors except the intervention indicator variable were centered in the HLM model.

Exhibit 60. Results for 2-Year Intent-to-Treat Impact Analysis for WJ Applied Problems Standard Scores by Age Group

Predictors	Age group at pretest																	
	37 months			48 months			60 months			72 months			84 months			96 months		
	β	*s.e.*	*p*	*β*	*s.e.*	*p*	*β*	*s.e.*	*p*	*β*	*s.e.*	*p*	*β*	*s.e.*	*p*	*β*	*s.e.*	*p*
Intercept	92.25	3.09	<.0001	93.33	2.14	<.0001	94.50	1.22	<.0001	95.67	0.87	<.0001	96.84	1.56	<.0001	98.01	2.55	<.0001
Intervention	5.51	2.34	0.034	4.07	1.82	0.043	2.49	1.36	0.089	0.92	1.15	0.438	-0.66	1.31	0.625	-2.23	1.75	0.223
Pretest	0.57	0.03	<.0001	0.57	0.03	<.0001	0.57	0.03	<.0001	0.57	0.03	<.0001	0.57	0.03	<.0001	0.57	0.03	<.0001
Boy	2.39	0.74	0.001	2.39	0.74	0.001	2.39	0.74	0.001	2.39	0.74	0.003	2.39	0.74	0.001	2.39	0.74	0.001
Grade K at pretest	1.73	1.50	0.249	1.73	1.50	0.249	1.73	1.50	0.249	1.73	1.50	0.249	1.73	1.50	0.249	1.73	1.50	0.249
Grade 1 at pretest	1.05	2.34	0.655	1.05	2.34	0.655	1.05	2.34	0.655	1.05	2.34	0.655	1.05	2.34	0.655	1.05	2.34	0.655
Grade 2 at pretest	4.84	3.28	0.141	4.84	3.28	0.141	4.84	3.28	0.141	4.84	3.28	0.141	4.84	3.28	0.141	4.84	3.28	0.141
Age at pretest	0.10	0.09	0.285	0.10	0.09	0.285	0.10	0.09	0.285	0.10	0.09	0.285	0.10	0.09	0.285	0.10	0.09	0.285
Home language–English	-4.54	1.13	<.0001	-4.54	1.13	<.0001	-4.54	1.13	<.0001	-4.54	1.13	<.0001	-4.54	1.13	<.0001	-4.54	1.13	<.0001
Intervention by age interaction	-0.13	0.06	0.022	-0.13	0.06	0.022	-0.13	0.06	0.022	-0.13	0.06	0.022	-0.13	0.06	0.022	-0.13	0.06	0.022
Intervention effect size	0.41			0.30			0.18			0.07			-0.05			-0.17		
Intervention improvement index	16			12			7			3			-2			-7		

Note. Two-level HLM models were used where children were nested in schools. The HLM in each column was centered by the age group indicated by the top of the column. Estimated impact (*β*) and standard errors (*s.e.*) are the coefficient and standard errors associated with intervention variable from the 2-level HLM model (children nested in school model). For the 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. Effect size measures the change (measured in standard deviations) in an average student's outcome that can be expected if the student is exposed to 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 exposed to the intervention. Improvement Index is a way to translate the effect size into a meaningful metric in educational research. WWC (2008) recommends translating the effect size into "improvement in percentile rank," which is supposed to indicate the expected change in percentile rank for the median comparison children if that student had received the early math intervention.

Effect size = Estimated impact/pooled standard deviations of the intervention and comparison group.

Model A for TEAM T-Scores = HLM impact models controlling for pretest scores.

Model B for TEAM T-Scores = HLM impact model controlling for pretest, gender, age, and language spoken at home.

Model C for TEAM T-Scores = HLM impact model controlling for pretest, gender, age, language spoken at home, and an interaction term between intervention and age.

All the predictors except the intervention indicator variable were centered in the HLM model.

Exhibit 61. Results for 2-Year Intent-to-Treat Impact Analysis for TEAM T-Scores by Age Group

Predictors	Age group at pretest																	
	37 months			48 months			60 months			72 months			84 months			96 months		
	β	*s.e.*	*p*	*β*	*s.e.*	*p*	*β*	*s.e.*	*p*	*β*	*s.e.*	*p*	*β*	*s.e.*	*p*	*β*	*s.e.*	*p*
Intercept	40.36	2.30	<.0001	41.42	1.60	<.0001	42.58	0.90	<.0001	43.74	0.61	<.0001	44.89	1.12	<.0001	46.05	1.85	<.0001
Intervention	-0.42	1.73	0.8128	-0.58	1.31	0.6646	-0.76	0.94	0.4286	-0.94	0.75	0.2307	-1.12	0.90	0.2306	-1.30	1.26	0.3183
Pretest	0.38	0.03	<.0001	0.38	0.03	<.0001	0.38	0.03	<.0001	0.38	0.03	<.0001	0.38	0.03	<.0001	0.38	0.03	<.0001
Boy	0.92	0.58	0.1084	0.92	0.58	0.1084	0.92	0.58	0.1084	0.92	0.58	0.1084	0.92	0.58	0.1084	0.92	0.58	0.1084
Grade K at pretest	5.98	1.18	<.0001	5.98	1.18	<.0001	5.98	1.18	<.0001	5.98	1.18	<.0001	5.98	1.18	<.0001	5.98	1.18	<.0001
Grade 1 at pretest	5.93	1.85	0.0014	5.93	1.85	0.0014	5.93	1.85	0.0014	5.93	1.85	0.0014	5.93	1.85	0.0014	5.93	1.85	0.0014
Grade 2 at pretest	11.71	2.54	<.0001	11.71	2.54	<.0001	11.71	2.54	<.0001	11.71	2.54	<.0001	11.71	2.54	<.0001	11.71	2.54	<.0001
Age at pretest	0.10	0.07	0.153	0.10	0.07	0.153	0.10	0.07	0.153	0.10	0.07	0.153	0.10	0.07	0.153	0.10	0.07	0.153
Home language – English	0.01	0.85	0.9906	0.01	0.85	0.9906	0.01	0.85	0.9906	0.01	0.85	0.9906	0.01	0.85	0.9906	0.01	0.85	0.9906
Intervention by age interaction	-0.01	0.04	0.7305	-0.01	0.04	0.7305	-0.01	0.04	0.7305	-0.01	0.04	0.7305	-0.01	0.04	0.7305	-0.01	0.04	0.7305
Intervention effect size	-0.03			-0.04			-0.05			-0.07			-0.08			-0.09		
Intervention improvement index	-1			-2			-2			-3			-3			-4		

Note. Two-level HLM models were used where children were nested in schools. The HLM in each column was centered by the age group indicated by the top of the column. Estimated impact (*β*) and standard errors (*s.e.*) are the coefficient and standard errors associated with intervention variable from the 2-level HLM model (children nested in school model). For the 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. Effect size measures the change (measured in standard deviations) in an average student's outcome that can be expected if the student is exposed to 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 exposed to the intervention. Improvement Index is a way to translate the effect size into a meaningful metric in educational research. WWC (2008) recommends translating the effect size into "improvement in percentile rank," which is supposed to indicate the expected change in percentile rank for the median comparison children if that student had received the early math intervention.

Effect size = Estimated impact/pooled standard deviations of the intervention and comparison group.

Model A for TEAM T-Scores = HLM impact models controlling for pretest scores.

Model B for TEAM T-Scores = HLM impact model controlling for pretest, gender, age, and language spoken at home.

Model C for TEAM T-Scores = HLM impact model controlling for pretest, gender, age, language spoken at home, and an interaction term between intervention and age. All the predictors except the intervention indicator variable were centered in the HLM model.

Exhibit 62. Estimated Impact of Early Math Intervention on WJ-AP Standard Scores for Each Age Group

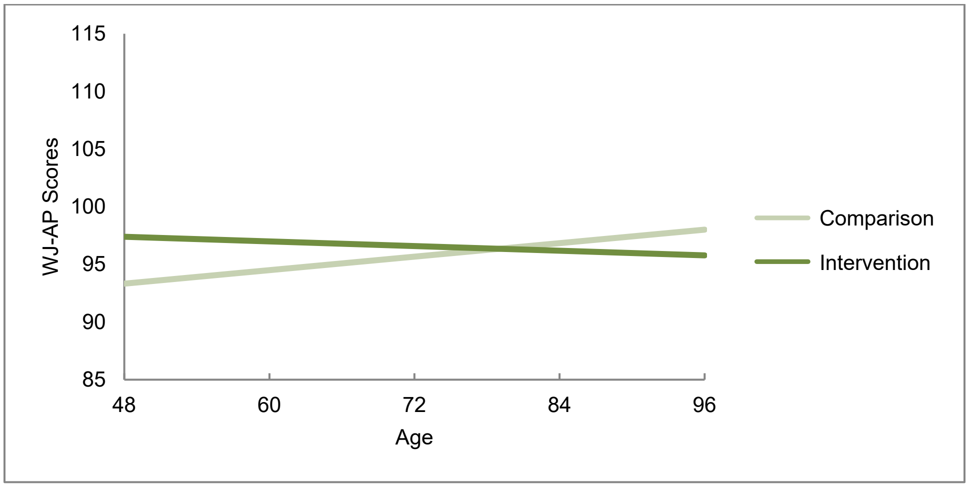
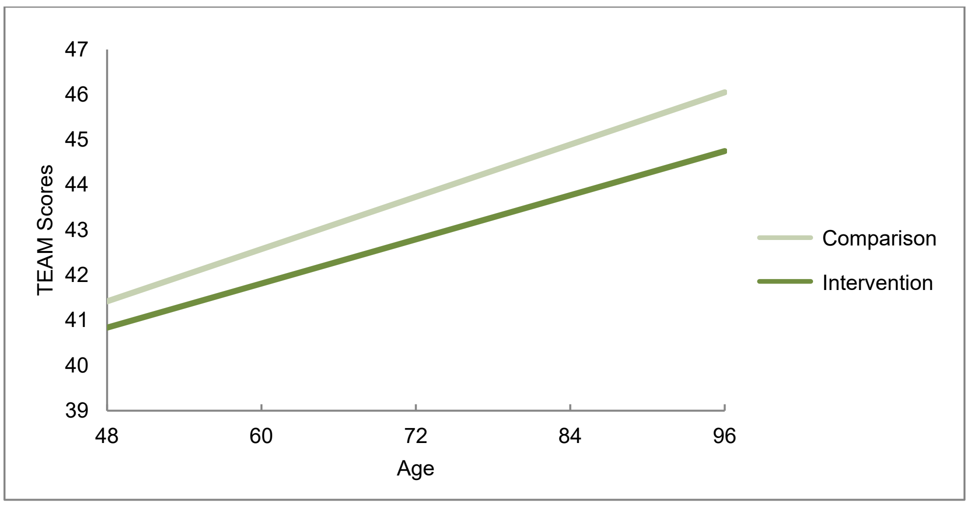


Exhibit 63. Estimated Impact of Early Math Intervention on TEAM T-Scores for Each Age Group



ILLINOIS STATE ACHIEVEMENT TEST (ISAT) OUTCOMES

ISAT scores were requested for 1,195 students with parental consent enrolled in third grade in 2012, 2013, and 2014. Scores were not available in 2015, the last year of the project, because the ISAT was replaced by the Partnership for Assessment of Readiness for College and Careers (PARCC) as the state achievement assessment. Third-grade Math ISAT scores were available for 1,107 students. Of these students, 130 were identified as being eligible for special education services in third grade and were excluded from the analyses per the inclusion criteria. An additional 106 students were no longer enrolled at a school that was involved in the Innovations project and were excluded from the analyses. This resulted in a total of 871 students with third-grade Math ISAT scores. Of these students, 50.5% were male; 66.5% were

Hispanic, 15.4% were Black, 8.7% were White, 8.2% were Asian, and 1.3% were classified as other; and 95.1% qualified for free or reduced-price lunch. Data were available for 334 (38.3%) students who were enrolled in third grade, 274 (31.5%) enrolled in second grade, and 263 (30.2%) enrolled in first grade, when the project started.

Of the 871 students, 512 of students were in the intervention condition and 359 were in the comparison condition. Exhibit 64 reports background characteristics of the sample at each time point. Examining the overall sample across years, initial analyses indicated significant differences between the intervention and comparison group in terms of percentage of Hispanic, $\chi^2 = 12.61$, $p < 0.01$, Asian, $\chi^2 = 12.88$, $p < 0.01$, and White students $\chi^2 = 6.34$, $p = 0.04$. Specifically, there was a significantly higher proportion of White students in the intervention condition than in the comparison condition in 2013. There were significantly higher proportions of Hispanic and Asian students in the intervention condition in 2013 compared to the comparison condition and a significantly higher proportion of Asian students in the intervention condition in 2014. There were also significant group differences in terms of years involved in the project (represented by year of assessment). That is, a higher proportion of intervention students were involved in the project for 1 year (2012) and a higher proportion of comparison students were involved in the project for 2 years (2013).

Exhibit 64. Background Characteristics of Intervention and Comparison Students by Year of ISAT Assessment

Year of Assessment	Intervention		Comparison			
	n	%	*n*	%	$\chi2$	*p*
2012	217	42.4	117	32.6	**8.56**	**0.04**
Female	110	50.7	56	47.9	0.24	0.65
Hispanic	130	59.9	82	70.1	3.40	0.07
Black	38	17.5	24	20.5	0.45	0.56
Asian	21	9.7	6	5.1	2.12	0.21
White	26	12.0	5	4.3	**5.36**	**0.03**
F/RL Eligible	205	94.5	114	97.4	1.56	0.28
2013	143	27.9	131	36.5	**7.17**	**0.01**
Female	70	49.0	66	50.4	0.06	0.90
Hispanic	86	60.1	98	74.8	**6.67**	**0.01**
Black	21	14.7	16	12.2	0.36	0.60
Asian	23	16.1	7	5.3	**8.09**	**0.01**
White	10	7.0	10	7.6	0.41	1.00
F/RL Eligible	131	91.6	125	95.4	0.44	0.60
2014	152	29.7	111	30.9	0.15	0.71
Female	78	51.3	54	45.9	0.74	0.45
Hispanic	100	65.8	130	59.9	2.44	0.14
Black	17	11.2	38	17.5	1.41	0.27
Asian	12	7.9	21	9.7	**4.73**	**0.05**
White	19	12.5	6	5.4	3.75	0.06
F/RL Eligible	144	94.7	107	96.4	0.41	0.57

Mean Math ISAT scores are reported by year in Exhibit 65. Because the data were cross-sectional and no pretest scores were available, the evaluation team limited the analyses to determining whether the individual Math ISAT scores of students enrolled in Innovations schools significantly differed from those of their comparison school peers at each available time point. The results of Hierarchical Linear Modeling analyses indicated no such effects were discernable at any time point.

Exhibit 65. Estimated Impact of Innovations on Math ISAT Scores by Year of Assessment

Year of Assessment	Intervention			Comparison					
	n	*M.*	*SD*	*n*	*M.*	*SD*	*β*	*s.e.*	*p*
2012	217	210.71	25.85	117	213.05	27.04	3.18	4.00	0.43
2013	143	211.92	27.94	131	212.69	27.90	-0.43	5.11	0.93
2014	152	214.82	27.43	111	212.54	24.13	3.03	4.90	0.54

For each of the 16 schools in our sample, information was collected about the percentage of students who performed at or above grade expectations on the Math ISAT from 2007 to 2013. Grade-level expectations were changed in 2014, making that year's data ineligible for comparison. Math ISAT performance trends for the intervention and comparison group are displayed in Exhibit 66. An analysis of covariance, with percent of students performing at or above grade expectations in 2011 as the covariate, did not reveal a significant effect of the intervention on the percent of students performing at or above grade expectations in 2013, $F(1,15) = 1.85$, $p = 0.20$, most likely due to the small sample size (eight schools per group). However, a Hedge's *g* effect size of 0.58 was calculated on group mean differences in 2013 scores adjusted for 2011 performance. After 2 years of intervention, the long-standing trend of schools in the intervention group having a lower percentage, on average, of students performing at or below grade expectations than the comparison group on the Math ISAT was reversed. It is notable that average scores in intervention schools in 2014, while not suited for the analysis of covariance, continued an upward trend.

Exhibit 66. Average Percent of Students Performing At or Above Grade Expectations on the Math ISAT by Condition

Mean percent of third graders who performed at or above grade expectations on the Illinois State Achievement Test (ISAT) by experimental condition. The vertical dashed line indicates the start of Innovations implementation in the intervention schools.

Conclusions

From 2011 to 2015, the Innovations in Early Mathematics project reached administrators and teachers in eight elementary schools and 106 classrooms serving an estimated 3,000 students each year with its unique approach to providing professional development in mathematics instruction to prekindergarten through third-grade teachers and school administrators. Below, we summarize: 1) implementation fidelity; 2) impacts on teachers' attitudes/confidence, practice, and pedagogical content knowledge; 3) impacts on children's math learning across 2 years; and 4) evidence for sustainability of effects on teachers during the last 2 years.

Implementation Fidelity

Fidelity of implementation in Year 1. Fidelity to all intervention components was high during Year 1 at the teacher level and adequate at the school level, suggesting the program "got off to a good start." School level fidelity scores were particularly susceptible to the inclusion of all teachers present at any point during the school year in the sample, as opposed to only including those teachers who were available to receive the many different components of the intervention delivered from fall to spring. Teacher turnover means many teachers did not experience the full intervention as intended, and that impacted fidelity at the school level.

Fidelity of implementation in Year 2. Overall, fidelity of implementation declined in the 2nd year of the project, with coaching in particular being difficult to successfully deliver. Grade-level meetings and learning labs were still fairly well-attended, and school level scores showed the same vulnerability to teacher turnover that was found in Year 1 fidelity scores.

Summary of Fidelity

Overall fidelity, defined as attendance or participation in PD, was high in Year 1 and adequate in Year 2. School level fidelity findings indicated a mixed pattern for Years 1 and 2, both in terms of fidelity across components and fidelity within components when comparing the sample of teachers available for the entire year versus the full sample. During both years, there was high fidelity of implementation for grade-level meetings, but not for other components.

Impacts on Teacher Outcomes

Shifts among teachers in Year 1. At the end of Year 1, teachers showed positive and significant shifts in teaching practice (HIS-EM) compared to baseline scores (effect size = 0.65, $p < .003$). There were no discernible effects, however, on measures of attitudes toward math, confidence in mathematics teaching (ABC-EM), or mathematical pedagogical content knowledge (PCK-EM) after 1 year of implementation.

Shifts among teachers in Year 2. Intervention teachers made significant improvements in confidence in early math teaching compared to comparison teachers (effect size = 0.51, $p = 0.026$) in Year 2 of implementation. Differences in teaching practice (HIS-EM) that were significant in Year 1 became only marginally significant in Year 2 ($p < 0.057$). However, it is worth noting that the overall mean score of intervention teachers remained essentially at its Year 1 level. Measures of teachers' knowledge (PM-PCK) were not available at the Year 2 posttest.

Summary of Teacher Impacts

There was an overall trend of positive growth in intervention teachers' math *teaching confidence,* their *practices*, and their *knowledge.* The evidence was not monolithic, but the fact that it was found in all three aspects of teacher development that were measured suggests a certain robustness. One can characterize the findings as: between-group differences in teaching *practice* appeared in Year 1; there was stability in these *practice* scores and new differences appeared between groups in *teaching confidence* in Year 2; and while both *practice* and *teaching confidence* scores in Year 3 remained stable, there were new between-group differences in at least some aspects of teachers' *pedagogical content knowledge*. The fact that *practice* and *confidence* showed an additional and sizable level of between-groups difference in Year 4, despite teacher attrition, suggests a real trend toward improved early math teaching in the intervention group.

Impacts on Child Outcomes

Children's learning outcomes in Year 1. Schools in the intervention did not show any significant impact on overall child outcomes for preK–3 students on the WJ-AP, TEAM, or ISAT scores at the end of Year 1.

Children's learning outcomes in Year 2. As in Year 1, the intervention did not have an overall significant impact on children's mathematics achievement in Grades preK–3 after 2 years of implementation. However, there was evidence of a differential impact of the intervention on students' mathematics achievement based on age. That is, students who were preschoolers in intervention schools at the beginning of the study significantly improved their scores on the WJ-AP by the end of Year 2 ($\beta = .30$, $p = .043$), whereas preschool-aged students in the comparison condition did not. A similar result at the kindergarten level was marginally significant ($p < 0.08$)

with an effect size of .18. No such differences were observed on the TEAM. And while there were no significant differences in ISAT scores, most likely due to the small sample size (eight schools per group), the evaluation team found a large effect (g = 0.58) on group mean differences in 2013 scores adjusted for 2011 performance. For the previous 6 years, from 2007 to 2012, intervention schools had had fewer students scoring at or above grade level in math on the ISAT than the comparison schools, but in 2013, this trend was reversed. It was unfortunate no further ISAT data were available, so the team could be more confident that this was a meaningful shift.

Summary of Child Impacts

It is likely that the lack of significant between-group differences in child outcomes for the entire PK-3 sample at the school level was due, in part, to small sample sizes at the school level. There were eight intervention and eight comparison schools in our study. With a total of 16 schools and assignment at the school level, this study was powered to detect an impact of 0.30. However, the effect size of the 1-year impact was 0.04 and 0.02 on AP and TEAM for the overall sample, respectively. This study was under-powered to detect an effect size of 0.04 or 0.02.

It is also not surprising that the effect of the PD on children at the youngest grades might be larger, more powerful, and therefore more readily detected by the analysis. Recall that the ideas on which this PD was based—the Big Ideas—were originally developed for preK and kindergarten teachers and had been tested and honed with them for several years. For Innovations, the Collaborative extended those Big Ideas upward, to include Grades 1, 2, and 3. While there was some limited piloting before implementation, PD methods particular to the upper grades were not as well-vetted as those that had been previously developed for preK and kindergarten.

The PD designers also chose to bring all preK to third-grade teachers together for learning labs, which were a central learning experience for delivery of content. These grade bands represent a very broad developmental range, and so bringing preK to third-grade teachers together to study the same mathematics as a group means that some topics pertain more to one grade-level teacher than another. To address this concern, learning lab content was designed to be broad and allow opportunities for tailoring to grade-specific concerns, and coaching and grade-level meetings were consciously used as opportunities to differentiate content, but it still may be that the PD could not meet everyone's needs equally and was stronger for teachers of children in younger grades.

It is unclear why significant between-group differences would be found (in preK) on one child outcome measure and not the other. While it is true they are fairly different measures—the WJ-AP has a stronger emphasis on number and the TEAM involves a contextualizing storyline and the use of manipulatives—each offers assessment items that align well with the topics covered by the Innovations PD. The Collaborative team previously found significant impacts of its PD (also based on the *Big Ideas*) on child outcomes at the preK and kindergarten levels using the WJ-AP, so it was sensible that those results would be essentially repeated here. This was the first time that the TEAM was used to attempt to detect differences in children's learning associated with the developers' work.

Finally, it is worth noting that the shift in ISAT scores found in Year 2 of implementation was based on the testing of third-grade students who were entering second grade when the intervention began, suggesting the possibility of an impact not just at preK and kindergarten but also at these upper grades. It is possible that more years of ISAT data might show the effects of the intervention "creeping up" the grade levels, since it makes sense that improved foundational knowledge can impact later learning. That is, it would be good to see whether this between-groups difference would continue and/or increase as students who were in first grade, kindergarten, and preK during the 1st year of the PD, and therefore had 3, 4, or 5 years of math teaching affected by Innovations prior to assessment, take the ISAT during their third-grade year.

Evidence for Sustainability

Teacher outcome data were collected for 2 more years (Year 3 and Year 4 of the project). This allowed the project to assess whether positive changes in teachers' *practice* and *confidence* could be sustained, and whether any changes in *pedagogical content knowledge* or *attitudes* might be detected over a longer period of time. Teacher *confidence* did not show a significant positive difference in Year 3, though it remained at essentially the same level as in Year 2. In Year 4, however, it shot up, significantly outpacing results for the comparison group (effect size = 0.71, p =0.023). A similar pattern was seen for teacher *practice*, which was essentially stable in Year 3 and increased during the last year of data collection, showing significant positive differences (effect size = 1.01, p = 0.0004).

Teachers' *attitudes* about mathematics proved very difficult to shift, though there was a marginally significant between-groups difference in Year 4 (effect size = 0.32, p = 0.069). Teachers did show a positive between-groups difference in Year 3 in *pedagogical content knowledge* on one of the two video samples (Number 7: effect size = 0.48, p = 0.031), and differences on the other video sample were marginally significant (Fraction: effect size = 0.13, p =0.093) at that point. No data on teacher knowledge outcomes were available for Year 4.

Study Limitations

IMPLEMENTATION STUDY LIMITATIONS

This study was not immune to the kinds of implementation challenges that educational interventions commonly face. Teacher turnover was high; high enough, in fact, that there was a differential impact of the intervention when all teachers who were in the sample at any given point during a school year were compared with those who were there for the majority of the school year and received more of the intervention. Note, too, that this fact did not account for additional changes in the teaching sample between years. Although the developers offered a brief induction to "new" teachers each fall, the research sample was essentially halved over the course of the 4 years as teachers left their schools for other positions. By Year 4, only about half the teacher population had received the entire intervention as intended.

In addition, the measures of fidelity of implementation and sustainability were simple attendance and participation counts, aggregated to the school level, per the requirement of the i3 grant. More specific fidelity information may have helped to inform if the materials from the professional development components were implemented in the classrooms and if so, how they were integrated with other curricula and math activities and lessons.

There were also some external events that may have negatively impacted our intervention schools. In the spring of 2013 (at the end of the 2nd year of intervention), three of our intervention schools and one of our comparison schools were placed on a list of possible school closings by the district administration. That spring was a chaotic time at these schools, with parents protesting and attending divisive meetings. Eventually, our intervention schools were not closed, but there is no doubt that learning at these schools was affected.

Between the 3rd and 4th years of our implementation, three principals at our intervention schools left their positions. One of these declined to support continued participation of their school in Innovations due to other priorities designated by the LEA, the Chicago Public Schools, but the other two remained involved in part of the work. The intervention team worked hard to work with the new administrators and capitalize on existing relationships at each of these schools, but the shift in leadership may have diluted the effects of the Innovations work.

IMPACT STUDY LIMITATIONS

There were also particular circumstances, beyond the actual delivery of the intervention, that may have impeded our ability to detect an overall impact on student outcomes and consistent impacts on teacher outcomes. As described above, due to capacity constraints, eight was the maximum number of schools that could be served by the developers and coaching team. This meant the sample size at the school level was 8, and that in turn produced a Minimum Detectable Effect (MDE) of 0.30 for student outcomes. Note that the significant effect size found by centering the analysis at 48 months was a healthy .30, meaning that the effect on 4-year-olds was sizable enough to be detected by this analysis. It is possible that more moderate effects on children at older grades were simply not big enough to be observed. Effect sizes above .20 in educational interventions are rare, so the sample and design may not have been adequate to detect the kinds of smaller effects that might have been actually achieved.

Another limitation was following children for only 2 years. First, students were also very mobile. Although the study began with 1,551 children at pretest, only 903 of these children, or 58%, were still enrolled in these schools at the end of the 2nd year of intervention. As with teachers, this means just over half the population in the school at the end of Year 2 had the opportunity to receive 2 years of the intervention. Second, given the way that mathematical knowledge builds from grade to grade, it may be the case that effective intervention in preschool set a better, stronger foundation for later learning. If we had another year of data, we might see an even larger effect for those students who were preschoolers when Innovations began—an effect that would be detected at the end of their first grade year. Longitudinal work with children in schools is difficult, of course, due to attrition within the school, so it is not clear how this could have been achieved.

Finally, while our findings about teacher development paint a compelling picture of learning that begins with shifts in practice, followed by an increase in teaching confidence and then demonstrated growth in pedagogical content knowledge, it is important to note that the tools used to assess these changes are new and relatively untested.

Innovations in Early Mathematics, while only partially achieving its intended goals, produced many useful outcomes. New tools and processes for supporting teachers and instructional leaders were developed throughout the work and remain in use to this day, having been shared with other PD providers and teachers across the nation. The group of coaches and PD facilitators that worked closely on Innovations went on to produce more substantive results when they combined the “whole school” PD approach with their strengths in preschool mathematics. The new intervention they created, Collaborative Math, is the subject of an almost-completed National Science Foundation grant-funded study that shows significant shifts in both teacher and child outcomes and is designed to help early childhood programs become centers of excellence in early math. The tools used to assess changes in teachers’ attitudes, practices, and knowledge have been refined and revised, and led to the development of a new teacher observation tool, the EQUIP-M, which is the subject of an Institute of Education Sciences measurement grant, testing its reliability and validity in a broad scale study.

References

Blank, R. K., de las Alas, N., & Smith, C. (2008). *Does teacher professional development have effects on teaching and learning? Analysis of evaluation findings from programs for mathematics and science teachers in 14 states*. Washington, DC: Council of Chief State School Officers.

Chen, J. Q., & Chang, C. (2006). Testing the whole teacher approach to professional development: A study of enhancing early childhood teachers' technology proficiency. *Early Childhood Research & Practice, 8*(1), 1-18.

Chicago Program Evaluation Project. (2008). *Final report*. Chicago, IL: Erikson Institute.

Claessens, A., Duncan, G., & Engel, M. (2009). Kindergarten skills and fifth-grade achievement: Evidence from the ECLS-K. *Economics of Education Review, 28*, 415-427.

Clements, D. H., Sarama, J., & Wolfe, C. B. (2011). *TEAM—Tools for early assessment in mathematics*. Columbus, OH: McGraw-Hill Education.

Clements, D. H., Sarama, J. H., & Liu, X. H. (2008). Development of a measure of early mathematics achievement using the Rasch model: The Research-Based Early Math Assessment. *Educational Psychology, 28*(4), 457–482.

Curtis, L. H., Hammill, B. G., Eisenstein, E. L., Kramer, J. M., & Anstrom, K. J. (2007). Using inverse probability-weighted estimators in comparative effectiveness. Analyses with observational databases. *Medical Care, 42*(2), 158-165.

Darling-Hammond, L., & Bransford, J. (Eds.). (2005). *Preparing teachers for a changing world: What teachers should learn and be able to do*. San Francisco, CA: Jossey-Bass.

Flanagan, K. D., McPhee, C., & Mulligan, G. (2009). *The children born in 2001 at kindergarten entry: First findings from the kindergarten data collections of the early childhood longitudinal study, birth cohort (ECLS-B)*.

Ginsburg, H. P., Lee, J. S., & Boyd, J. S. (2008). Mathematics education for young children: What it is and how to promote it. *Social Policy Report, 22*(1), 3-23.

Heckman, J. J., & Masterov, D. V. (2004). *The productivity argument for investing in young children*. Washington, DC: Invest in Kids Working Group, Committee for Economic Development.

Hirano, K., Imbens, G. W., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica, 71*(4), 1161-1189. doi:10.1023/a:1020371312283

Illinois State Board of Education. (2002). Illinois learning standards. Retrieved from http://www.isbe.net/ils/math/standards.htm

Kerachsky, S. (2009). *Achievement gaps: How black and white students in public schools perform in mathematics and reading on the National Assessment of Educational Progress*. Retrieved from nces.ed.gov/whatsnew/commissioner/remarks2009/7_14_2009.asp

McCray, J., & Chen, J. Q. (2012). Pedagogical content knowledge for preschool mathematics: Construct validity of a new teacher interview. *Journal of Early Childhood Research, 26*, 291-307.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics (PSSM)*. Reston, VA: Author.

National Research Council. (2009). *Mathematics learning in early childhood: Paths toward excellence and equity*. Committee on Early Childhood Mathematics, Christopher T. Cross, Taniesha A. Woods, and Heidi Schweingruber, Editors. Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Pianta, R. C., LsaParo, K. M., & Hamre, B. (2008). *Classroom Assessment Scoring System (CLASS), manual. Pre-K*. Baltimore, MD: Brookes.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41-55.

Sarama, J., & DiBiase, A.-M. (2004). The professional development challenge in preschool mathematics. In D. H. Clements & J. Sarama (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education*. Mahwah, NJ: Lawrence Erlbaum Associates.

Schrank, F. A., McGrew, K. S., Ruef, M. L., Alvarado, C. G., Muñoz-Sandoval, A. F., & Woodcock, R. W. (2005). *Overview and technical supplement* Rolling Meadows, IL: Riverside Publishing. Retrieved from http://www.riverpub.com/products/bateriaIII/pdf/BateriaIIIASB1.pdf

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14.

Starkey, P., Klein, A., & Wakeley, A. (2004). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. *Early Childhood Research Quarterly, 19*(1), 99-120.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25*, 1-21.

U.S. Department of Education. (2009). Investing in Innovation Fund (i3). Retrieved from http://www2.ed.gov/programs/innovation/index.html?utm_source=rssutm_medium=rssutm_campaign=the-u-s-department-of-education-announced-the-start-of-the-134-million-2014-investing-in-innovation-i3-grant-competition

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013). *What Works Clearinghouse: Procedures and standards handbook (Version 3.0)*. Washington, DC: U.S. Department of Education. Retrieved from http://whatworks.ed.gov

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Itasca, IL: Riverside Publishing.

Woodcock, R. W., Munoz-Sandoval, A. F., McGrew, K. S., & Mather, N. (2005). *Bateria III Woodcock-Munoz*. Itasca, IL: Riverside Publishing.

Appendix A

Sample Big Idea Page

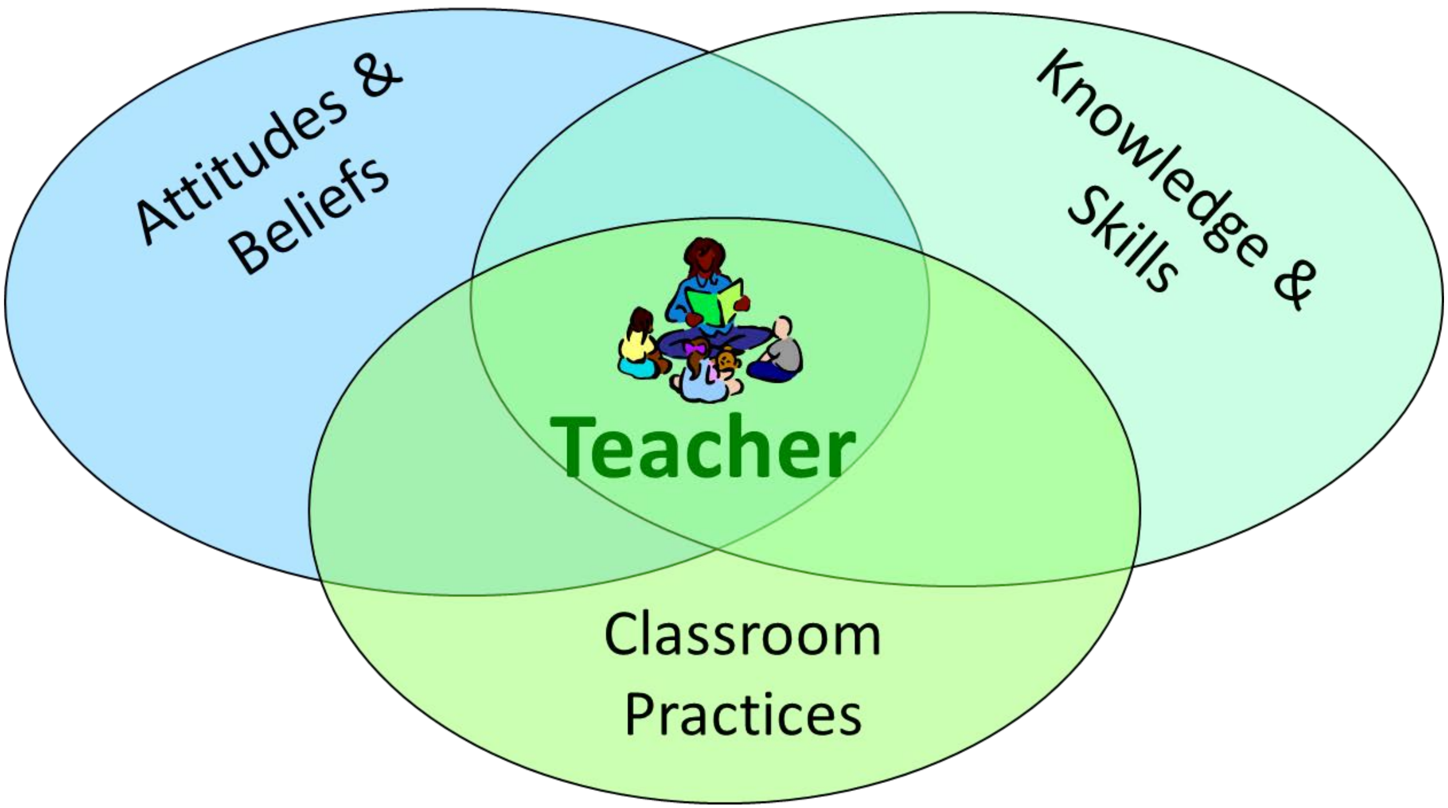
Big Ideas of Number Sense

Topic	Big Ideas	Examples
Uses of Number  5th	• Numbers are used many ways, some more mathematical than others.	• *Tommy has 5 books.* (cardinal) • *Ava is fifth in line today.* (ordinal) • Numbers on basketball jerseys, home addresses, telephone numbers (nominal) • *Let's meet at 5 pm on December 5.* (referential)
Numerosity  = 5	• Quantity is an attribute of a set of objects and we use numbers to name specific quantities. • The quantity of a small collection can be intuitively perceived without counting.	• 5 mice and 5 elephants are alike in quantity, though different in other ways. • Children just "see" three objects and know it's 3.

erikson
early math collaborative

Appendix B

Whole Teacher Approach Venn Diagram



Appendix C

Looking at Student Work Protocols

Looking at Student Work

Phase 1 Guidelines

Phase 1 Guidelines structure conversations among teachers to select and plan for a common assessment task. The goal is to examine what a task offers teachers in terms of understanding student thinking and to foster common understanding about the development of student strategies.

Each teacher brings one task to propose to the group. Tasks should have high-cognitive demand with multiple pathways to a solution.

1. **Getting Started** (5 min.)
   - Choose a facilitator to guide the group through the process and keep track of time.
   - Make sure everyone has a chance to review all proposed tasks.
     - *What kind of thinking does the task require? What big idea(s) are addressed? Will all of your students be able to engage in the task?*
2. **Selecting a Common Task** (10 min.)
   - Facilitator asks each teacher to comment on the proposed tasks. Keep the focus on what the task will reveal about student understanding. Everyone has an opportunity to talk.
     - *What are we hoping to learn about our students? Which task is best suited to our formative assessment goal?*
   - Come to a consensus about which task to implement.
3. **Anticipating Student Responses** (5 min.)
   - All teachers use sticky notes to brainstorm possible student solution strategies (one strategy per sticky note). This is quiet, individual think-time.
4. **Sorting and Sequencing Student Responses** (10 min.)
   - Moving from "Less Sophisticated" to "More Sophisticated," work together to create a developmental sequence of student strategies, grouping ones that are similar.
   - Consult the *Landscape of Learning* (© Catherine Twomey Fosnot. Heinemann, Portsmouth, NH) for language to label strategies and to connect them to big ideas.
5. **Planning** (10 min.)
   - Decide when to implement the task based on your formative assessment goal (i.e., where within the unit of study?)
   - Plan how to introduce the task. Discuss what scaffolding to offer.
     - *What access considerations are there? What math materials will be available? How can we offer support without lowering the cognitive demand of the task?*
   - Use the sequence of strategies you created to establish categories for selecting 3-4 pieces of student work to share at the next meeting.

Total Time: 35 minutes

Developed by the Early Math Collaborative at Erikson Institute with support from U.S. Department of Education Investing in Innovations Fund (i3) and CME Group Foundation.

Looking at Student Work:

A Collaborative Approach to Formative Assessment

Phase 2 Guidelines

Phase 2 Guidelines structure conversations among teachers to analyze and learn from student work. The goal is to create a supportive, safe space for thinking together about implications for instruction.

Each teacher brings a representative sample of student work (3 to 4 pieces) that reflects a range of abilities and may include incorrect or partially correct responses.

1. Getting Started (5 min.)

- Choose a facilitator to keep the group focused and on time.
- One teacher volunteers to present student work and displays it where everyone can see it. The presenting teacher says nothing about the work or the students until Step 3.
- Review work in silence, taking notes for the discussion

2. Discussing the Work in Three Rounds (10 min.)

Teachers take turns speaking around the table. Individuals are free to pass. There is no cross-dialogue (yet!).

- Round 1 DESCRIBE: ***What do you notice about the work?*** Describe what you see.
- Round 2 INTERPRET: ***What do the students understand?*** State evidence.
- Round 3 QUESTION: ***What questions do you have about the work?***
  - Comments need to be non-judgmental and focus on what the students ***can*** do.
  - Keep it brief! If you say "*and,*" you are probably saying too much.
  - Presenting teacher listens and takes notes.

3. Reflections from Presenting Teacher (5 min.)

Facilitator asks presenting teacher to share his or her reflections, react to observations, and answer questions raised. Facilitator may choose to insert probing questions, such as

- *What did someone say that made you think differently about a student's work? Did anything surprise you?*
- *How did the discussion deepen your mathematical understanding?*

4. Suggestions for Teaching and Learning (10 min.)

Facilitator invites everyone to relate key ideas raised in the discussion to suggestions for teaching and ways to support students' learning.

- *Based on the discussion of the students' performance what might you suggest doing next with the class?*
- *How well did the selected task give students an opportunity to demonstrate what they know?* (Consider the Mathematical Task Analysis guide)

5. Debriefing (5 min.)

Facilitator asks the group reflects on their experience.

- *Did this process give us the information we were looking for about our students?*
- *How can this process be improved?*

Total Time: 35 minutes

Adapted from the *Milwaukee Mathematics Partnership*. Developed by the Early Math Collaborative at Erikson Institute with support from U.S. Department of Education Investing in Innovations Fund (i3) and CME Group Foundation.

Appendix D

Learning Lab Summaries

Learning Lab 1: October 20 or 22, 2011 (3 hours)
Content Focus: Numerosity & Number Sense
Strategy Focus: Turn and Talk

Greeting Activity. The learning lab began with *Count Around*, a greeting activity designed to introduce the content focus of numerosity and number sense. The group formed a circle and selected a number. Teachers counted one by one until the number was reached. Then, that teacher introduced themself and sat down. Starting back at zero, this process was repeated until everyone was seated. Teachers then *turned and talked* to the person sitting next to them about what could be learned from this activity. This modeled the turn and talk strategy that they could use with their students.

Investigation. The main investigation of the learning lab then took place. Learning lab participants were given the first twenty numerals of a counting system traditionally used by shepherds in various locations across the British Isles.
Teachers again used the turn and talk strategy to share ideas on how to interpret this unfamiliar number system. Then they formed larger groups at tables to continue exploration. They used cubes to represent various quantities (*bumpit* or *yan-a-pimp*, for example), drew pictures depicting *tan-a-figgit* sheep, and worked together to create written symbols to represent the number system. Then, they made posters explaining how this number system functions.

The groups then did a *gallery walk*, analyzing each poster and attaching sticky notes with their questions or comments. After reading the comments on their posters, the whole group discussed such ideas as how numerosity is distinct from number systems, how number systems frequently have perceivable patterns, and how number names and symbols are arbitrary. They also discussed the learning process they experienced: how struggle can lead to discovery; how expert support and peer collaboration can enrich thinking; and how deep learning takes time. To conclude the investigation, learning lab participants did the "Count Around" activity once again, this time using the shepherd's counting system.

Big Ideas. After the investigation, a presentation on the Big Ideas of numerosity and number sense was given. The importance of developing abilities to subitize and identify landmark numbers was stressed. These ideas were reinforced by video clips from our *Focus on the Child* series, which show one-on-one interviews with children intended to elicit evidence of their mathematical thinking.

Strategies & Procedures. As this was the first learning lab, the rationale behind some of the procedures of Innovations, particularly the coaching process, was explained. The overarching goal of coaching is to turn conscious acts of effective teaching into unconscious, automatic schemas, and to turn unconscious, ineffective teaching habits into conscious, avoidable behaviors.

Learning Lab 2: November 18, 2011 (6 hours)
Content Focus: Counting, Cardinality & Number Sense
Strategy Focus: Learners reiterate/rephrase/articulate other learners' thinking

Greeting Activity. Upon entry, participants were given a card with 1, 2, or 3 dots pictured on it. A number—7 or 11, for example—was announced, and participants formed groups by finding others whose cards, when combined with their own, would add up to that number.

Investigation. Now in groups, were given a large quantity of red and white beads. Together, they determined how many *counting racks* they could create with the beads, as well as how many beads would be left over. Counting racks can have 20 beads (10 red, 10 white) or 10 beads (5 red, 5 white). They created a poster detailing their process.

After a *gallery walk*, in which groups studied each other's posters and left sticky notes with their questions or comments, a full-group discussion commenced.
Some main ideas that emerged were: counting is "procedure-heavy," which makes it complex; making groups or units that can be counted is a means to create efficiency; and unitizing is critical to the understanding of base-10.

Big Ideas. Teachers were presented with several principles of counting. They were instructed to *turn and talk* to discuss their understanding of each, as well as how the principle was illustrated in the investigation. When used together, these principles form the Big Idea that *quantity is an attribute of a set of objects, and numbers are used to name specific quantities*. These concepts were reinforced with a video of a K/1 class taking inventory of various objects in their classroom by making "packs" of ten and charting how many packs they had as well as how many objects they had left over.

Strategies & Procedures. The group discussed effective strategies they witnessed in the classroom from the video. Then, they discussed their own challenges, surprises, and successes in the past month using the turn and talk strategy discussed in the previous learning lab. This month's strategy focus, *learners reiterating/rephrasing/articulating other learners' thinking* was introduced. Facilitators pointed out how it could be built off of the turn and talk strategy. The general conclusion was reached that conversation between teachers and students about their mathematical thinking and problem solving not only builds mathematical understanding but also strengthens their ability to explain their thinking with precision in many contexts.

Grade-Level Lesson Analysis. Teachers broke into groups based on grade-level and looked at lessons that focus on the content discussed earlier in the morning. The groups discussed whether these lessons take the developmental characteristics of the grade into account. Then they turned the lens on to their own classrooms and brainstormed ideas on how to be more reflective and intentional in this sense.

Learning Lab 3: January 19 or 21, 2012 (3 hours)
Content Focus: Number Composition
Strategy Focus: Learners rephrase other learners' thinking

Greeting Activity. Participants were distributed a *ten frame* (a grid of two rows of five). The ten frames had any number of grid spaces between zero and ten filled in. Participants were instructed to find a person whose filled-in spaces, when added to theirs, would equal ten.

Classroom Resources. Participants learned about the benefits of using the *rekenrek counting rack* in their classrooms. The rekenrek is visually similar to an abacus. It has one or two rows of five red and five white beads, which can be used to count by sliding from one side to the other. Using the rekenrek, various number structures become easily recognizable: doubles; doubles plus one; five as a landmark; ten as a landmark.

These basic procedures and strengths were demonstrated to the group. Then, participants made their own mini-rekenreks out of pipe cleaners, beads, and cardstock. They were prompted to depict several different numbers on their rekenreks and discussed the different combinations and number structures represented at their tables. Video was then shown demonstrating further applications of the rekenrek, including adding and subtracting (done by sliding the beads in accordance with the problem) and using the rekenrek as a tool during daily attendance.

Big Ideas. It was then emphasized how these rekenrek activities can both strengthen and accelerate children's natural understanding of number composition. The rekenrek gives a clear visual representation of how quantities can be broken apart (decomposed), and put back together (composed) to form the whole. It also shows that for any given quantity (whole), there are different ways to compose and decompose it. Expanding upon these ideas, it was discussed that as numbers get larger, using doubles or other landmarks (fives, tens, hundreds, and so on) allow for more efficient composition and decomposition of numbers.

Common Core. The various ideas and activities presented in the Learning Lab were then related to the Common Core State Standards. Participants discussed how the videos and activities they viewed and took part in represented these standards.

Learning Lab 4: February 3, 2012 (6 hours)
Content Focus: Number Operations
Strategy Focus: Sharing multiple solutions or strategies without comment

Greeting Activity. As participants entered, they wrote their name on a card and placed it in a *rekenrek*-styled attendance board. Once everyone was accounted for, participants were asked to count how many people were present. They discussed the various ways they could count the cards—by ones, by landmarks (5s or 10s), etc. They also determined that moving the cards could help to create new landmark numbers.

Investigation. The following problems were written one-by-one, horizontally on the board:

- **1st string**: 70 + 45 ; 72 + 43 ; 67 + 48 ; 76 + 39
  *Encourages solvers to think about **compensating** to get a **constant sum**.*
- **2nd string**: 70 – 25 ; 71 – 26 ; 69 – 24 ; 72 – 27 ; 74 – 29
  *Encourages solvers to think about **compensating** to get a **constant difference**.*

After each problem, participants were invited to explain their solution. Throughout the discussion, each solution strategy was written on the board as they were described. Then, other participants were invited to explain how their peer solved the problem, whether they could rephrase that solution strategy, and whether they had a different way of solving the same problem.

Video Analysis. Videos were shown to demonstrate that math problems always have stories associated with them. Successful classrooms will initially focus on the problem situation rather than the numbers, as children need to understand what is happening before they can think about finding a solution.

Grade-Level Lesson Analysis. The teachers then split up based on grade level (PreK-K and 1-3) to study student work in depth. Each group was given a problem along with four student responses. Their task was to rank the student responses in order of complexity. After doing so, they discussed the evidence of student thinking that led them to their conclusions, what Big Ideas they could see in the problem, and how they could practice the day's strategy focus in their own classroom with this problem.

Learning Lab 5: March 22 or 24, 2012 (3 hours)
Content Focus: Grouping Situations
Strategy Focus: Teachers model students' thinking

Greeting Activity. Participants formed groups based on their schools. They determined how many elbows and how many toes their groups have.

Investigation. Participants worked together in their groups to solve the following problem:

> *Think about arranging square tables for Euclid Catering Company. This company has some odd rules: they always create rectangles with their tables and they always seat one person per exposed table edge. How could they arrange their tables to seat 28 people?*

When every group came up with at least one solution, they made posters diagramming and labeling their table arrangements, as well as organized lists with the number of tables used and the dimensions of each table arrangement. They briefly discussed how using arrays and rectangles is a powerful model for multiplication. Then, they were given part two of the problem:

> *This month, Euclid Catering Company has 79 events, each serving 28 guests. How many napkins will they need this month?*

After solving the problem, teachers were given 28 x 79 grids and were asked to make smaller rectangles within the large rectangle to show the steps they took.

Strategy Focus. Using actual student explanations of multiplication, facilitators modeled the process of modeling. Then, pairs of teachers were given the opportunity to try this themselves, with one role-playing as the student and the other as the teacher. A video was then shown of a student who could effectively model her thinking in one setting but did not exhibit full understanding of the concept in another setting. It was concluded that modeling is an extremely useful tool to develop and represent children's understandings of mathematical ideas but should not be the singular instruction strategy.

Learning Lab 6: April 26 or 28, 2012 (3 hours)
Content Focus: Using Number Sense & Algorithms to Solve Problems
Strategy Focus: Students explain and model their own thinking

Investigation. Participants split into six groups and were assigned one of six different algorithms for subtraction. Each algorithm has different ways of dealing with the concept of "borrowing." Some were simpler and more familiar: borrowing from the next largest place value, as is taught in most schools in the United States. The most unfamiliar and arguably most complicated mimicked how computers perform subtraction in the binary counting system.

Each group made a poster and explained how their assigned method works. While some groups had more familiar systems than others, explaining them clearly still required some creative thinking and wording. In discussing the posters, it was emphasized that these different methods should yield the same results, and that if an algorithm works every time, it is no more or less correct than another algorithm. More generally, it was affirmed that struggling to understand a concept at first could ultimately help one understand it better.

Video Analysis. Participants looked at two videos in which children use algorithms (which, by definition, are guaranteed to lead to a correct answer) yet do not respond correctly. In one case, a child chooses an appropriate algorithm but plugs in the wrong numbers. In the other video, a child combines two algorithms while attempting to solve a single problem. The importance of finding a balance between using algorithms and developing broader number sense was discussed. While algorithms can reliably result in a correct answer, children can still run into issues if their basic understanding of the overriding concept is lacking.

Grade-Level Lesson Analysis. Participants read over grade-level appropriate lessons. In groups, they discussed actions teachers could take throughout these lessons that would encourage children to show and explain their own thinking. Instructors reminded participants that last month's strategy focus, *teachers model students' thinking*, could scaffold students who are still learning to explain and illustrate their thinking clearly.

Innovations Summer Institute: Day 1 – June 18, 2012 (6 hours)

Investigation. Participants worked individually and then in groups on a problem about splitting pizza. They made posters describing their problem-solving process and debriefed as a whole group. In this case, the after eating 2/6 of one pizza and 3/8 of another pizza, 5/14 of the total pizza was eaten, which was not the expected answer; this problem was revisited throughout the duration of the Summer Institute.

Common Core State Standards. Participants were given a copy of the Common Core State Standards. First individually, then as table groups, they highlighted words or concepts that were unclear or needed some unpacking. Together, they came up with "friendly" language or classroom examples that could explain these passages. Then they shared their ideas with the whole group.

Big Ideas. The whole group briefly discussed the Big Ideas of how kids come to learn about *numbers and operations*. The instructor stressed that development of mathematical thinking is not linear nor is it identical from child to child. The instructor then introduced and distributed copies of the *Landscape of Learning*—a guide displaying the various strategies children might use, fall back on, and progress towards when thinking about number operations.

Games. Participants applied their number sense and strategic thinking to play *Mancala* and *21*.

Landscape of Learning. Participants studied problems from different grade levels. For each problem, they located strategies on the *Landscape* that children might use and discussed them as a whole group. They also discussed the logic behind the design of some of the problems. For instance, one of the problems had children counting pairs of shoes, with one shoe left over. This promotes children to think about using the strategies of *doubles* and *doubles plus one*. Then, videos of students from different grade levels were shown, and participants located the strategies used on the *Landscape*.

Innovations Summer Institute 2012: Day 2 – June 19, 2012 (6 hours)

Investigation. Participants revisited the "pizza problem" from the previous day. In this case, someone again ate 2/6 of one pizza and 3/8 of another, except this pizza parlor cut their slices differently. 17/48 of the total pizza was consumed, which is still not the expected answer. After working individually and then making posters in groups describing their process, they debriefed as a whole group about how this problem illuminates some Big Ideas about fractions: *fractions are equal parts of a whole*; *a whole or unit can be divided into equal parts in many different ways*; *a unit may be a single object or a collection of things*.

Common Core State Standards. Participants finished reviewing the Standards. They created lists of action statements describing what teachers can do to give students the opportunity to engage in and develop mathematical ideas. They transcribed these ideas on to posters and had a *gallery walk* of everyone's work. Facilitators instructed participants to focus on common ideas between the posters. Together, they consolidated all of the statements into 3-5 specific plans of action.

Games. Participants applied their number sense and strategic thinking to play *Mancala* and *21*.

Strategies & Routines. Participants described the purpose of daily math routines. Two brief videos of a *rekenrek* attendance routine in PreK and 1st grade were shown. Participants discussed the type of math learning depicted, the amount of time this type of routine requires, and how it can evolve throughout the school year. Teachers then made lists of all the math routines they currently use in their own classrooms. They shared and categorized their responses. Now having several routines across grade levels in mind, teachers worked as individuals or in small groups to outline what sorts of routines they would realistically like to implement in their classroom. They were asked to imagine they were given a $100 budget to spend on routines—how might they allocate these funds to cover everything they want to accomplish?

Getting together in school groups, teachers were asked to look at their daily routines in the perspective of the *Landscape of Learning*. They considered how well their routines support the concepts depicted on their *Landscape*, and what routines they might want to start implementing to strengthen weaker areas in their students.

Innovations Summer Institute 2012: Day 3 – June 20, 2012 **(6 hours)**

Investigation. Each participant was given a card with a fraction printed on it. There were signs on the wall marked with "0," "1/2," and "1." Participants moved to the landmark number they thought their fraction was nearest. They discussed with others how they knew this, as well as how this activity builds number sense for fractions.

Content Analysis. Participants worked on a different "pizza problem," this time about adjusting a precise recipe. After working individually and in groups, they discussed how ideas from the last few days might have affected their problem solving strategies. The idea was brought up that these "pizza problems" exist in a real-world context. These types of problems are useful in helping children garner a deep understanding of fractions, rather than simply memorizing answer-getting rules. They brainstormed other topics that could potentially involve whole and fractional parts that might benefit students in the same way. Finally, they discussed some Big Ideas about number (discussed in an earlier Learning Lab) that lay the groundwork for a solid understanding of fractions: *a quantity (whole) can be decomposed into equal or unequal parts; the parts can be composed to form the whole*; *one can quantify a collection by grouping items into equal sets*.

Common Core. Participants split into groups and looked at some example problems. They described and made posters of what student work for these problems would look like at the novice and proficient level. While each group presented their posters, the other groups tried to find common ideas. They considered what kinds of questions help students engage in and develop these practices, as well as what questions a teacher could ask to scaffold their thinking.

Games. Participants applied their number sense and strategic thinking to play *Mancala* and *21*.

Book Ideas. The whole group read a children's book and took notes about the mathematical ideas present in it. Then, they split into groups, each with a small collection of books to read and analyze in a similar fashion. They compiled ideas on how they could use these books to build mathematical understanding.

Innovations Summer Institute 2012: Day 4 – June 21, 2012 (6 hours)

Investigation. Participants worked on various fraction story problems. They discussed how, even though all dealt with the overall topic of fractions, different models were necessary to solve each problem—an area model worked best with one, a number line worked best with another. They went over some classroom activities that teachers could implement in their own classrooms to get children thinking about these different types of models.

Common Core. Participants were distributed copies of the Common Core's Standards for Mathematical Practice. Teachers broke into small groups and studied individual practices. Then, they viewed a video of a second-grade lesson on fractions, noting whether the teacher did or did not demonstrate their assigned practice. Groups discussed the video amongst themselves and then as a whole group. The discussion included ideas on what instructional shifts might benefit the teacher practice displayed in the video, and how participants could make those shifts in their own classrooms.

Games. Participants applied their number sense and strategic thinking to play *Mancala* and *21*.

Strategies & Procedures. Several math "stations" were set up, and teachers rotated between them in 20-25 minute shifts. Some stations contained games: for example, participants played "Rekenrek Bingo," in which the "teacher" displayed a number on the rekenrek, and "students" strategically covered combinations of addends on their playing board that added up to that number. Other stations provided different types of activities: for example, participants looked at different solution strategies and role-played as student and teacher to practice modeling student work.

Learning Lab 7: August 28, 2012 (6 hours)
Content Focus: Linear Measurement

Greeting Activity. Teachers split into groups and discussed different ways they measured in the last 24 hours.

Investigation. Each table group received a set of 5 or 6 similar objects, such as blocks, books, or rocks. They were asked to put their objects in order from biggest to smallest, using sticky notes to keep track of the order. They sorted their objects two more times (focusing on different attributes) and noted how the order changed. They shared their findings with the whole group, highlighting what they discovered about attributes and comparison.

Video Analysis. Before watching the video, participants were instructed to find an object in the room with a different length than their hand. Then they viewed "Just Right for Me" from the *Focus on the Lesson* video series. With the familiar story "Goldilocks and the Three Bears" in mind, preschoolers search the room for objects that are "just right" (the same size as their hand) and discuss how they can be absolutely sure of the accuracy of their measurement. After the video, participants discussed what the teacher does to support the children's exploration of measurement. Participants then were given pattern blocks to quantify how much longer or shorter their object was when compared to their hand. They discussed the difficulties of using a non-numerical measurement system.

Lesson Analysis. Teachers read through a case study of a lesson designed to elicit disagreements among students. The lesson features a "broken ruler," which begins at 3 inches. When asked to measure an index card, students come up with a multitude of answers. One student even says, "There's two answers!" Participants took notes and discussed what these students understand about measurement as well as where that understanding is fragile.

Investigation. Each participant received a cardstock strip and a pattern block. Using these two items, they were instructed to create their own ruler. Instructors did not give detailed instructions on how to go about doing this; instead, they simply asked, "What makes a ruler work?" After completing their rulers, participants formed pairs and measured objects around the room, comparing their results as they went. As a whole group, they discussed the activity, honing in on what general rules apply to all rulers.

Common Core. The ideas from the day were related to Common Core State Standards. Facilitators showed videos and led discussion on how concepts from measurement overlap with other mathematical topics, especially number sense. To close, the book *A Pig is Big* was read aloud, to reiterate the idea that measurements are precise ways to describe relationships between different objects in quantifiable, reproducible terms.

Learning Lab 8: October 4 or 6, 2012 (3 hours)
Content Focus: Measurement

Greeting. Participants were challenged to visually estimate if everyone present could span the entire perimeter of the room with their arms outstretched. If not, how many additional people would be needed? Participants briefly *turned and talked* to explain and justify their answers. Then, they stood up and tried to see if they could indeed stretch the span of the room!

Investigation. Participants worked in groups to find a solution strategy (not an exact answer) to a problem involving finding the area and perimeter of an irregular shape with only one side length given. They made posters explaining why their solution strategy would work, how they would find an answer, and what additional information they might need. Then, the groups performed a *gallery walk*, observing each other's posters and posting any comments and questions they had. Groups explained their strategies to the others, and facilitators prompted participants from other groups to restate their thinking in their own words.

Some groups suggested using grids to help solve the problem. This bridged into a whole-group discussion about area serving as a measurement model of multiplication. It was discussed how grids (and thus, concepts of measurement, area, perimeter, etc.) can naturally progress from the regular use of landmark-based tools such as rekenreks and particularly 10-frames. It was emphasized that students need to find a balance between pictorial, symbolic, and concrete understandings to fully understand a mathematical concept.

Video Analysis. Before the video, participants were shown a container and cubes of various sizes. They estimated how many cubes of each type it would take to fill the container. They briefly shared answers and discussed how children might progress from understanding area to understanding volume. The video featured a second grade volume activity, in which a teacher observes pairs of students working with cubes. Participants discussed evidence of Common Core mathematical practice in the video and related it to the ideas present in the day's learning lab. Grade-level groups discussed what knowledge students would have to have before they could effectively use manipulatives and work without direct teacher guidance like the students in the video.

New Teacher Induction Training Part 1: October 27, 2012 (6 hours)
Content Focus: Numerosity & Number Sense

Greeting. As participants entered, they placed a craft stick into a rekenrek-styled attendance chart. They then took part in a *Count Around* activity. Standing in a circle, participants chose a target number and counted up to it. When the number was reached, that participant introduced themself and sat down. When finished, participants discussed with a partner, and then with the whole group what could be learned from this activity, and gave "thumbs up" if they heard ideas similar to what they had discussed.

Material Analysis. Participants studied the attendance chart they used when they entered. They discussed how many people were present and how they saw that number from the way the craft sticks were arranged in the chart (grouping clusters of sticks or use landmark numbers, for example). They compared the process of counting with the rekenrek, counting using unifix cubes and counting groups of bundled straws. It was emphasized that different manipulatives and counting tools allow children to conceptualize and make associations in different ways.

Investigation. In groups, participants deciphered a hypothetical counting system—it uses the letters A through E and dashes, rather than the numerals in the decimal system we are more familiar with. The groups used unifix cubes to represent various numbers in this system (*DE* or *C*-, for example) and created a poster describing how they might explain this number system to someone less well versed. When finished, the groups took part in a *Math Congress*, scrutinizing each other's posters and asking for further clarification when needed. Coming back together as a whole group, they debriefed what could be taken away form this investigation.

Big Ideas. Some of the Big Ideas of number sense were covered more deeply. Participants viewed video of these ideas in action and *turned and talked* to discuss what they noticed. It was emphasized that the various concepts of number sense develop in parallel in children. They related these ideas back to their own experience of learning a new number system in the previous investigation. Participants did the *Count Around* activity from earlier in the day; however, this time they used the new number system from the day's investigation.

New Teacher Induction Training Part 2: November 17, 2012 (3 hours)
Content Focus: Visual Number Sense & Operations
Strategy Focus: Landscape of Learning

Greeting. Participants placed a craft stick in a rekenrek-styled attendance chart. Those participants who had attended the first new teacher induction training grabbed two sticky notes and recorded something that impacted their thinking and something that they tried in their classroom since last time. New teachers took one sticky note and recorded something they wanted to learn or try in their classrooms.

After this, participants formed a circle. Starting on a number between 2-8, each person added nine and said that number. These numbers were written on the board, and teachers discussed what pattern they were noticing as well as how they were finding the next number. They discussed with a partner what was or could be learned from this activity and discussed their ideas as a whole group. After all of this, some time was given to reflect on the comments on the sticky notes. Teachers were given the chance to reflect on their experiences since the last meeting.

Material Analysis. Participants looked at the rekenrek attendance rack and counted how many people were in attendance, discussing how they saw the quantity. Facilitators then presented evidence from brain research highlighting the importance of building visual number sense and relationships between numbers in young children—something that the rekenrek excels in. Teachers *turned and talked* about how they would precisely define number sense and what it looks like when a person has good number sense. Coming back together, facilitators discussed the connection between the rekenrek, 10-frames, and dot cards. They discussed the strategies that using the rekenrek emphasizes, and what everyday situations (e.g., attendance, bunk beds, double decker bus) the rekenrek could relate to. They also viewed video demonstrating ways to introduce the rekenrek in the classroom.

Learning Lab 9: December 13 or 15, 2012 (3 hours)
Content Focus: Spatial Reasoning

Greeting. Participants introduced themselves and then traversed a simple obstacle course set up in the room one at a time. Observers repeated a "chant" emphasizing the direction word for each action: "Go *around* the chair, go *between* the tables," and so on. After a certain point, participants who speak languages other than English were invited to lead the chant in that language (Spanish or Polish, for example). Teachers briefly discussed the difficulties students might experience trying to make sense of the often flexible and arbitrary meaning of directional prepositions.

Book Ideas. The book *Rosie's Walk* was read aloud in English and Spanish. As it was being read, teachers drew maps detailing the path that Rosie walked in the story. The maps were briefly discussed. It was emphasized that Rosie's path (which ends the same place it starts) is the perimeter of a closed shape and that this shape encloses a region. Participants were prompted to consider what additional details would be needed to make a more precise map of Rosie's path.

Investigation. On a grid, participants made up a new route for Rosie. When finished, they broke off into groups of two. Hiding their maps from each other, one person described their map while the other person attempted to draw it. The person drawing was only allowed only ask two clarifying questions throughout the process. Each person got a turn in each role. Afterwards participants discussed how accurate their maps were and what additional details would help improve their mapmaking.

HIS-EM. The HIS-EM (High Impact Strategies for Early Mathematics) framework was introduced to participants. While the Common Core lists standards for student practice, HIS-EM defines standards for teacher practice. HIS-EM outlines nine different facets of math teaching that impact student learning outcomes. Teachers were briefed on one of the dimensions, *mathematical representations*, to help analyze the following video.

Video Analysis. The teachers viewed "Walk with Rosie" from the *Focus on the Lesson* series. This lesson builds off of *Rosie's Walk*. A small group of preschool English Language Learners traverse an obstacle course in the classroom. Much like earlier in the day, as one student goes through the course, the others narrate the actions. Then the students make maps of the route they took. During the video, half of the Learning Lab participants took notes on the teacher's instruction choices, and the other half focused on the student's practice. Teachers *turned and talked* with each other and then with the whole group about their observations.

Make and Take. Teachers were instructed how to make *tangrams* from regular sheets of paper. Tangrams are puzzles made from seven shapes that can be combined to make numerous arrangements. After making their tangrams, participants made various arrangements. They discussed the role of visualization when their students work with shapes, puzzles, and blocks.

Learning Lab 10: February 1, 2013 **(3 hours)**

Content Focus: Shapes

Greeting. Upon entering, each participant received a three-dimensional solid. Several sorting instructions were called out to the group—"shapes with a triangular face, move to the left side of the room," for example—and participants shifted around the room as instructed.

Investigation. Participants were given a piece of paper and were instructed to make a replica of their three-dimensional solids. To do this, they had to make a *net*—a two-dimensional outline of their shape, which they could then fold and tape together to make their solid. Participants remarked that they and their students are much more familiar with two-dimensional shapes, even though the objects we interact with on a daily basis are three-dimensional. It was concluded that our schooling system, which is text-and picture-focused, emphasizes those shapes that can be easily displayed on a flat viewing surface. Teachers then took notecards and wrote mathematical descriptions of their three-dimensional shape. The cards were collected, shuffled, and redistributed. Each participant was given a card and searched for the person whose shape matched the description. Once everyone had found each other, they discussed various aspects of this activity: what kinds of words people used to describe the shapes; what descriptions could apply to multiple shapes.

Video Analysis. Before watching the video, in which a teacher leads her second grade class to create three-dimensional shape "skeletons" using straws and twist ties, participants studied the handout that the teacher gave to her students. For each shape they make—cube, triangular prism, etc.—students are instructed to note how many edges, vertices, and faces each shape has. Participants *turned and talked* to one another about what they might see in this video. While watching, one half took notes on the teacher's performance in relation to the HIS-EM standard of *conceptual development*. The other half looked at students' performance in relation to the Common Core standard of *looking for and making use of structure*. Afterwards, the group discussed if the lesson led students to a deeper understanding of concepts and if the teacher helped students generalize what they learned.

Strategies & Procedures. Teachers then considered how students come to recognize two-dimensional shapes and attributes of polygons. They studied the *Van Hiele* model of geometric thought. Some major points that came from this included the fact that various levels of geometric understanding are sequential but not tied to age, and that the use of physical materials, drawings, and computer models are important at every level. Keeping these points in mind, participants looked at a few video case studies. After each example, teachers *turned and talked* about their reactions before debriefing as a whole group. The *Big Idea* emphasized was that *shapes can be defined and classified by their attributes*.

Learning Lab 11: February 28 or March 2, 2013 (3 hours)
Content Focus: Shapes and Attributes

Greeting. On a *rekenrek* chart, teachers answered the question "Have you visited the Early Math Collaborative website in 2013?" (Note: Our current website, <earlymath.erikson.edu>, launched in February 2013.) "Yes" was marked with blue strips of paper; "no" was marked with yellow strips of paper. Without looking, participants then placed a card on their back with a shape pictured on it. They walked around the room asking yes-or-no questions to try to determine what shape they had. Participants were not permitted to use shape names in their questions or answers; for example, "do I have a triangle?" or "no, you don't have a rectangle" would both be forbidden!

Book Ideas. Teachers read *The Important Book of Polygons* by Margaret Wise Brown. They discussed briefly to review the difference between *polygon* and *shape*. Then, mimicking the structure of the book, groups wrote poems about various polygons: rhombus, octagon, acute triangle, etc. The first and last lines of the poems had to be mathematically true and apply to all polygons of the given category. Each poem was written on poster paper and displayed when finished.

Video Analysis. Teachers viewed two videos from the *Focus on the Child* series, in which children sort blocks based on their attributes. Facilitators emphasized that *attribute* is a key concept in all areas of mathematics. As a whole group, they looked at the Common Core State Standards for geometry and summarized that *attribute* and *composing and decomposing shapes* are two of the most common concepts. They discussed why those ideas might come up so often in the standards and considered how they might relate to other areas of math.

Investigation. Participants were given pattern blocks and puzzle frames and were asked to complete the puzzles in multiple ways. Then, in table groups, they made charts of pattern block equivalence. For instance, two of the triangle pattern blocks can be put together to fill the same amount of space as the rhombus pattern block. The "square" and "small rhombus" pieces are unusual in that they do not easily compose or decompose into the other shapes. Teachers were challenged to find their position in the equivalence charts. Teachers *turned and talked*, then shared in the whole group how composing and decomposing shapes could help students build understanding in such topics as fractions, area and multiplication.

Video Analysis. Participants viewed a video of a second grade class representing fractions with pattern blocks. One half of participants focused on the teacher's practice, specifically focusing on her performance in the HIS-EM standard of *student engagement*. The other half of the group looked at student practice through the scope of the Common Core, specifically focusing on their *strategic use of appropriate tools*. They *turned and talked* after the video, and then discussed as a full group. The discussion focused on who was doing more of the thinking, explaining and justification in the video—teacher or student?

Learning Lab 12: April 6 or 8, 2013 (3 hours)
Content Focus: Data & Pattern

Greeting. As in last month's learning lab, teachers used the *rekenrek chart* to mark whether they had visited the new Early Math website. The responses from each month were compared. Participants discussed why it could be useful to have data from two time points, and if the rekenrek works well as a tool for this task.

Investigation. In table groups, participants listed all the footwear they own. They discussed if any obvious trends, as well as what possible questions they could answer using their data. Focusing on one question, they considered how best to display their data: bar graphs, pie charts, tallies, and so on. Once decided, they made posters to share their information. The table groups then did a *gallery walk* of the posters, focusing on how clearly their graph answers their question. Coming back as a whole group, participants discussed the various uses of each type of graph.

Video Analysis. Because lessons on data can be more time consuming than other math topics, facilitators had participants reflect on how much class time they would expect to reserve for set-up, execution, and discussion. Participants were told that the lesson in the video to follow was designed as a two-day activity. They viewed the planning conversation and execution of the lesson "Shoe Graph" from the *Focus on the Lesson* video series, in which preschoolers create categories and graph their shoes. Participants discussed how the teachers in the video adjusted the activity after hearing questions and ideas raised from students. They also discussed the teachers' ability to prompt analysis from their students.

Content Analysis. Focus was then turned to the concept of pattern, which was framed as an underlying concept in math, rather than a unit or series of activities to cover. Teachers started by standing in a circle and counting by 9s, beginning with any number less than 20 and not divisible by 9. Each number was recorded on the board. After everyone had said a number, facilitators asked participants what patterns they noticed from the listed numbers. Participants were shown other types of patterns—numerical patterns, growing patterns, etc.—and concluded that predictability and generalizability are key components of pattern. Participants then discussed what they already do in their classrooms to get students thinking about pattern and how they could emphasize the mathematical ideas in these activities. These ideas were related to standards from the Common Core.

Video Analysis. Participants viewed a video of a multi-day activity for first grade, in which students make packs of ten while creating an inventory of objects in the classroom. They make a chart of their findings, marking down the number of packs (tens) along with how many are left over (ones). Half of the learning lab group took notes on the teacher's practice, focusing on the HIS-EM standard of *planning*. The other half of the group focused on the children, looking for the Common Core standard of *looking for and expressing regularity in repeated reasoning*. Teachers *turned and talked*, then debriefed as a whole group about their conclusions.

Summer Spectacular: August 15, 2013 (6 hours)

Workshops. There were two 90 minute workshop sessions, and participants could choose what topic to explore in each session.
The topics for the 1st session were::

- *What's the Buzz? Orchestrating Productive Mathematics Discussions in Your Classroom*
- *Talking the Talk: Getting the most out of Number Strings*
- *Guided Discovery & Interactive Modeling: Games & Manipulatives*
- *How much is your name worth? Using books to inspire mathematical thinking in the 1st 6 weeks of school*

The topics for the 1st session were:

- *Supporting English Language Learners in the Mathematics Classroom*
- *Starting Strong: Setting Up for Successful Math Routines*
- *Using Reading Comprehension Strategies to Boost Your Students' Math Problem Solving Skills*
- *Making Math Work Stations Work for You*

Plenary Session. All the participants joined together to hear about the project's research plan. They also watched a video about the project by the National Center for Quality Teaching and Learning.

Common Core Content Standards. Participants worked in missed-grade groups to dig into the content shifts in the Common Core Standards, using place value as a focus.

Winter Day of Learning: January 24, 2014 (3 hours)

Greeting. A small presentation was given, celebrating the accomplishments of Innovations teachers at this point in the program. Grade level meetings, lesson study, and the Teacher Leader Initiative were all highlighted.

Analysis. Participants solved the following problem:

> *I went to the library last week. On Monday, I checked out 1 book. On Tuesday, I checked out 2 books. On Wednesday, I checked out 3 books. On Thursday, I checked out 4 books. On Friday, I checked out 5 books. How many books did I check out last week?*

After solving it themselves, they discussed how children might go about solving this problem, thinking specifically of what strategies or misconceptions they might have. Then they viewed videos of two 2^{nd} graders solving the same word problem. They compared and contrasted the work of both students.

Strategies & Procedures. To demonstrate the idea that *there's more to addition and subtraction than just "add" & "take away,"* participants modeled solutions to the number sentence "13-10=3" using various strategies: counting with manipulatives, drawing pictures, "jumping" on the number line, etc. They discussed how different procedures and language can be connected to the same operations and written symbols. Having a broad understanding of addition and subtraction strategies helps children apply operations flexibly to solve a range of problems.

Participants also focused on the idea that *all operations tell a story*. They looked at various types of problems and discussed what makes them more or less challenging. They classified each problem they looked at on a chart of common addition and subtraction situations (for example, "comparison problem with difference unknown" as opposed to "comparison problem with smaller quantity unknown").

Common Core. Facilitators selected eight standards from the Common Core relating to operations and algebra. Table groups worked together to put these standards in order from most complex to least complex. As a whole group, they discussed their conclusions. After revealing the actual order, participants discussed how to help students meet these standards. Ideas included providing students with a variety of problem situations and giving them ample time to discuss the problems and their solutions.

Appendix E

Standards for Mathematical Practice Toolkit

Standards for Mathematical Practice Toolkit

Summer Institute ● 2012

erikson
early mathematics education
INNOVATIONS

COMMON CORE
STATE STANDARDS INITIATIVE
PREPARING AMERICA'S STUDENTS FOR COLLEGE & CAREER

This toolkit was created through a collaborative process involving more than 100 Preschool through 3rd grade teachers from eight Chicago Public Schools. Over a period of four days in June of 2012, these teachers met with staff and instructors of Erikson Institute's Early Mathematics Education Project in Chicago to study and discuss the eight Standards for Mathematical Practice that were published as part of the Common Core State Standards for Mathematics. We hope this toolkit can provide assistance to teachers, both in the Chicago Public Schools and elsewhere in the United States, as they shift their mathematics teaching practices to incorporate the Common Core State Standards.

Thank you to the teachers at our EME Innovations schools:

Brentano Math and Science Academy
de Diego Community Academy
Cleveland Elementary School
Gale Elementary Community Academy
Jordan Elementary Community School
Lorca Elementary School
Manierre Elementary School
Reinberg Elementary School

COMMON CORE STATE STANDARDS FOR MATHEMATICAL PRACTICE

The 8 practice standards describe the behaviors of mathematically proficient students. Mathematics teachers at all levels should seek to develop these behaviors. Every lesson should include all the practices, though any given lesson will emphasize some practices more than others. Student behaviors are connected to teacher behaviors.

Practice #1: Make sense of problems and persevere in solving them

Students:

- Identify the important information needed to make a plan
- Monitor work throughout the process, verifying strategies and solutions
- Keep trying until a clear understanding emerges
- Show patience and a positive attitude

Teachers:

- Model how to pull out important information by asking questions and re-reading the problem carefully
- Encourage the use of different strategies and give time for students to explain strategies to one another
- Make strategic decisions about when to offer students help and when to let them struggle
- Encourage students to continue until they are confident they have done their best

Practice #2: Reason abstractly and quantitatively

Students:

- Identify relevant quantitative information in a problem situation
- Visualize a problem situation mentally
- Represent a problem and solution with pictures, models, numbers and other symbols
- Use numbers and operations flexibly

Teachers:

- Ask questions that help students abstract the math from problem situations
- Ask students to explain their thinking, regardless of accuracy
- Use thinking aloud to model reasoning
- Highlight the flexible use of numbers

Practice #3: Construct viable arguments and critique the reasoning of others

Students:

- Communicate answers and logical thinking processes using words, pictures, acting it out, etc.
- Identify confusions to discover clarity
- Ask clarifying questions to improve understanding
- Actively compare thoughts of others to own ideas

Teachers:

- Plan time for students to share and compare thinking (explain, rephrase, turn & talk, etc.)
- Establish classroom norms for the safe discussion of different ideas
- Model and encourage the asking of questions to clarify thinking
- Use confusion as an opportunity for learning

Practice #4: Model with mathematics

Students:
- See and describe the relationship between a model and an everyday situation
- Select and apply appropriate models to solve problems and represent thinking
- Check that models accurately reflect the situation and revise as necessary

Teachers:
- Plan tasks and problems that involve solving equations in everyday situations (e.g., grocery shopping, sharing)
- Provide time for students to share and discuss their models and how they relate to their thinking about a problem
- Highlight similarities and differences between various models

Practice #5: Use appropriate tools strategically

Students:
- Make reasonable choices about when to use tools
- Demonstrate the correct use of tools while solving problems
- Learn from the use of a tool

Teachers:
- Provide a variety of appropriate mathematical tools and time to explore their use
- Consistently model use of tools during instruction
- Expect students to use mathematical tools to support their reasoning

Practice #6: Attend to precision

Students:
- Specify the steps involved in solving a problem
- Use accurate mathematical language, including symbols, labels, definitions, and units of measure
- Calculate with precision and attention to detail

Teachers:
- Model clarity of explanation by using explicit language and clear mathematical models
- Encourage students to be specific when explaining their thinking
- Have students paraphrase others' thinking to push for precision and confirmation

Practice #7: Look for and make use of structure

Students:
- Look for and recognize mathematical significance
- Generalize relationships within and between problems (e.g., Math-to-Math connections)
- Apply a new idea to related problems

Teachers:
- Plan tasks and problems with patterns (e.g., number strings)
- Ask questions that focus students of the structure the problem
- Highlight different approaches for solving a problem

Practice #8: Look for and express regularity in repeated reasoning

Students:

- Check work for sense, repeatedly
- Notice patterns and connections that help them develop generalizations or "shortcuts"
- Explain what they are doing and why it makes sense
- Explain why a generalization is true and useful

Teachers:

- Ask about possible answers before, and reasonableness during and after computations
- Use thinking aloud to model how to explain what they are doing and why it makes sense
- Ask students to explain what they are doing and why it makes sense
- Ask students if a generalization is always true

Practice #1: Make sense of problems and persevere in solving them

Context: Joining problem with the change unknown

NOVICE	APPRENTICE	PROFICIENT
Student is able to identify part of the information (the given) but not enough to make sense of the problem as a change unknown problem. Student is unaware how to model or solve this type of problem. Student is not yet able to recognize whether the answer is correct or not, or gets frustrated and gives up.	Student can identify the important information in the problem but is unable to determine the correct operation for a change unknown problem. Student chooses a flawed strategy and does not verify that solution makes sense.	Student can recognize the problem type and can use the given and unknown to model the change. Student is able to check his or her work and change strategies until he or she has a clear understanding and a solution.

TEACHER QUESTIONS THAT SCAFFOLD PRACTICE #1 (in context)

- *What are the important words in the problem?*
- *How many did she have to start with and how many did she end up with?*
- *What changed to get to the final number?*
- *What ideas do you have about how to show this problem using manipulatives or pictures and how can they help us solve the problem?*
- *Would another way work better?*

Practice #2: Reason abstractly and quantitatively

Context: Mental math with number strings

NOVICE	APPRENTICE	PROFICIENT
Student may be able to solve initial problem, but cannot apply quantitative understanding to develop a strategy. Student does not find ways to use numbers flexibly to make problem-solving easier.	Student may be able to solve entire string of problems, but not be able to explain how they solved each one or why a strategy worked. Student may be able to use target strategy to solve one string, but not apply the reasoning to a similar string.	Student can successfully solve string of problems by applying their understanding of numbers and operations. Student can explain how they solved the problem and why it works.

TEACHER QUESTIONS THAT SCAFFOLD PRACTICE #2 (in context)

- *For visual number strings: Do you see any groups or patterns that could help you?*
- *For symbolic number strings: Do you see any friendly numbers that could help you, or could you make a friendly number?*
- *Can you use something you already know to help you with this problem?*

Practice #3: Construct viable arguments and critique the reasoning of others

Context: Fair share problem

NOVICE	APPRENTICE	PROFICIENT
Student cannot explain to others how he or she shared. Student cannot justify the fairness of the result. Student does not ask or respond to questions for clarity.	Student can explain some parts of sharing process, but not all. Student may have difficulty coming up with clarifying questions, even when she or he recognizes she or he does not understand. Student may be able to follow the reasoning of others but not be able to compare it to his or her own thought process.	Student communicates the strategy used to divide the total number into equal parts. Student can justify his/her reasoning and compare it to others. Student can ask and respond to questions for clarity.

TEACHER QUESTIONS THAT SCAFFOLD PRACTICE #3 (in context)

- *How do you know your answer is fair?*
- *How is what he/she did different from what you did?*
- *Which part of this doesn't make sense to you?*
- *Is there another way to show how you made it fair?*

Practice #4: Model with mathematics

Context: Classroom inventory problem

NOVICE	APPRENTICE	PROFICIENT
Student may group some objects to count but the groupings do not match his or her counting. The number does not accurately reflect the quantity of objects. Student does not recognize the mismatch between the situation and the models.	Student knows to group objects for counting but does not know why this is helpful or how groups of 10 are related to our number system. When prompted, student is able to revise model for increased accuracy. Student does not generalize models to other situations involving large counts.	Student can group objects into "friendly" number sets (e.g., 5s and 10s) to correctly count in an efficient way. Student can model counting with groups of objects and numbers and explain the relationship between the two models. Student is able to apply models to other everyday situations involving large counts.

TEACHER QUESTIONS THAT SCAFFOLD PRACTICE #4 (in context)

- *Is there an easier way to count this?*
- *Is there a way to keep track of your counting?*
- *Why did you choose to group by 10s? How is that helpful?*
- *How does your model (groupings, numbers, chart, tallies) show the total amount?*

Practice #5: Use appropriate tools strategically

Context: 1- or 2-digit addition

NOVICE	APPRENTICE	PROFICIENT
Student does not know to select a tool to help solve the problem. Student does not know which tool(s) to select in order to solve problem (e.g., blocks, number line, rekenrek). When given the appropriate tool, the student does not know how to use the tool independently.	Student realizes they have a choice of tools, but sometimes chooses a tool that does not help the student work efficiently. Student may choose an appropriate tool, but may not use it strategically (e.g., counts by ones on rekenrek). Student is sometimes able to explain why they chose a tool and how it was helpful.	Student knows how and when to use tools strategically with 1- and 2-digit addition problems. Student can accurately explain their tool choice and how it helped them solve the problem. Student's tool use enhances their understanding of the addition problem. Student is able to solve problems with multiple tools and strategies.

TEACHER QUESTIONS THAT SCAFFOLD PRACTICE #5 (in context)

- *What tool could you use to solve this problem?*
- *Why did you choose this tool?*
- *Could you show me how you used this tool?*
- *Is there another way you could use this tool?*
- *Is there another tool that may be more efficient?*

Practice #6: Attend to precision

Context: Comparison problem

NOVICE	APPRENTICE	PROFICIENT
Student cannot identify specific steps to solve the problem. Student misses key mathematical language to understand the problem. Student miscalculates or details relating to symbols, labels, or units are missing.	Student solves problem accurately, but cannot provide a detailed explanation or produces inaccurate models. Student understands the concept but miscalculates during operation or details relating to symbols, labels, or units are incorrect.	Student calculates and solves problem accurately and all details relating to symbols, labels, and units are correct. Student can precisely explain his/her thinking.

TEACHER QUESTIONS THAT SCAFFOLD PRACTICE #6 (in context)

- *What do we know? What are we trying to find out?*
- *What is the story of the problem or what is happening in this story?*
- *Have you included all the important information to communicate your answer?*
- *Can you tell us exactly what you did?*
- *How could you explain this to someone who is confused?*

Practice #7: Look for and make use of structure

Context: Function Machines (from *Everyday Mathematics* in grades K—3)

NOVICE	APPRENTICE	PROFICIENT
Student does not know to look at more than one in/out pair to find the pattern. Student does not recognize the rule and does not correctly complete the function machine.	Student can figure out the change in at least some of the in/out pairs. Student may be able to complete the function machine but has difficulty generalizing the rule to produce new in/out pairs.	Student knows to look at relationship of the in/out pairs to generalize the rule. Student can produce new in/out pairs or create his or her own function machine to fit a given rule.

TEACHER QUESTIONS THAT SCAFFOLD PRACTICE #7 (in context)

- *Do you notice a change in each pair of numbers that is the same?*
- *Did you look at all the pairs?*
- *If the pairs all change by 3, what would be another example that would work?*

Practice #8: Look for and express regularity in repeated reasoning

Context: Frames and Arrows (from *Everyday Mathematics* in grades 1—3)

NOVICE	APPRENTICE	PROFICIENT
Student cannot explain why or how the rule works. Student inserts numbers into the sequence haphazardly and does not check his or her work for reasonableness.	Student can figure out missing numbers in sequence, but cannot identify a missing rule. Solving sequences with smaller numbers is easier. Even when successful with one problem, student gets confused with new Frames and Arrows problems.	Student can complete or extend a sequence, filling in any missing parts. Student is able to explain what he or she is doing and why it works. Student can create own Frames and Arrows problems.

TEACHER QUESTIONS THAT SCAFFOLD PRACTICE #8 (in context)

- *If I put another number in the sequence, would that help?*
- *Do you see any pattern in the way the numbers change from one frame to the next?*
- *What helped you figure out the rule?*
- *How can you show that the rule works?*
- *How can you use what you learned in problem X to help you in problem Y?*

Everyday Mathematics® Goals for Mathematical Practice

The *Everyday Mathematics* authors have distilled the CCSS-M Standards for Mathematical Practice into a set of 23 "Goals for Mathematical Practice" that are intended to be more usable for elementary school teachers and students. These *Everyday Mathematics* Goals for Mathematical Practice (GMPs) provide a framework for instruction in mathematical practices similar to the framework for mathematical skills and understandings provided by the *Everyday Mathematics* Program Goals and Grade-Level Goals. The chart below provides the full text of each CCSS-M Standard for Mathematical Practice in the left-hand column along with the corresponding GMPs in the right-hand column.

Common Core State Standards for Mathematical Practice	*Everyday Mathematics* Goals for Mathematical Practice
Standard for Mathematical Practice 1: Make sense of problems and persevere in solving them.	
Mathematically proficient students start by explaining to themselves the meaning of a problem and looking for entry points to its solution. They analyze givens, constraints, relationships, and goals. They make conjectures about the form and meaning of the solution and plan a solution pathway rather than simply jumping into a solution attempt. They consider analogous problems, and try special cases and simpler forms of the original problem in order to gain insight into its solution. They monitor and evaluate their progress and change course if necessary. Older students might, depending on the context of the problem, transform algebraic expressions or change the viewing window on their graphing calculator to get the information they need. Mathematically proficient students can explain correspondences between equations, verbal descriptions, tables, and graphs or draw diagrams of important features and relationships, graph data, and search for regularity or trends. Younger students might rely on using concrete objects or pictures to help conceptualize and solve a problem. Mathematically proficient students check their answers to problems using a different method, and they continually ask themselves, "Does this make sense?" They can understand the approaches of others to solving complex problems and identify correspondences between different approaches.	**GMP 1.1** Work to make sense of your problem. **GMP 1.2** Make a plan for solving your problem. **GMP 1.3** Try different approaches when your problem is hard. **GMP 1.4** Solve your problem in more than one way. **GMP 1.5** Check whether your solution makes sense. **GMP 1.6** Connect mathematical ideas and representations to one another.
Standard for Mathematical Practice 2: Reason abstractly and quantitatively.	
Mathematically proficient students make sense of quantities and their relationships in problem situations. They bring two complementary abilities to bear on problems involving quantitative relationships: the ability to *decontextualize*—to abstract a given situation and represent it symbolically and manipulate the representing symbols as if they have a life of their own, without necessarily attending to their referents—and the ability to *contextualize,* to pause as needed during the manipulation process in order to probe into the referents for the symbols involved. Quantitative reasoning entails habits of creating a coherent representation of the problem at hand; considering the units involved; attending to the meaning of quantities, not just how to compute them; and knowing and flexibly using different properties of operations and objects.	**GMP 2.1** Represent problems and situations mathematically with numbers, words, pictures, symbols, gestures, tables, graphs, and concrete objects. **GMP 2.2** Explain the meanings of the numbers, words, pictures, symbols, gestures, tables, graphs, and concrete objects you and others use.

Common Core State Standards for Mathematical Practice	*Everyday Mathematics* Goals for Mathematical Practice
Standard for Mathematical Practice 3: Construct viable arguments and critique the reasoning of others.	
Mathematically proficient students understand and use stated assumptions, definitions, and previously established results in constructing arguments. They make conjectures and build a logical progression of statements to explore the truth of their conjectures. They are able to analyze situations by breaking them into cases, and can recognize and use counterexamples. They justify their conclusions, communicate them to others, and respond to the arguments of others. They reason inductively about data, making plausible arguments that take into account the context from which the data arose. Mathematically proficient students are also able to compare the effectiveness of two plausible arguments, distinguish correct logic or reasoning from that which is flawed, and—if there is a flaw in an argument—explain what it is. Elementary students can construct arguments using concrete referents such as objects, drawings, diagrams, and actions. Such arguments can make sense and be correct, even though they are not generalized or made formal until later grades. Later, students learn to determine domains to which an argument applies. Students at all grades can listen or read the arguments of others, decide whether they make sense, and ask useful questions to clarify or improve the arguments.	**GMP 3.1** Explain both what to do and why it works. **GMP 3.2** Work to make sense of others' mathematical thinking.
Standard for Mathematical Practice 4: Model with mathematics.	
Mathematically proficient students can apply the mathematics they know to solve problems arising in everyday life, society, and the workplace. In early grades, this might be as simple as writing an addition equation to describe a situation. In middle grades, a student might apply proportional reasoning to plan a school event or analyze a problem in the community. By high school, a student might use geometry to solve a design problem or use a function to describe how one quantity of interest depends on another. Mathematically proficient students who can apply what they know are comfortable making assumptions and approximations to simplify a complicated situation, realizing that these may need revision later. They are able to identify important quantities in a practical situation and map their relationships using such tools as diagrams, two-way tables, graphs, flowcharts and formulas. They can analyze those relationships mathematically to draw conclusions. They routinely interpret their mathematical results in the context of the situation and reflect on whether the results make sense, possibly improving the model if it has not served its purpose.	**GMP 4.1** Apply mathematical ideas to real-world situations. **GMP 4.2** Use mathematical models such as graphs, drawings, tables, symbols, numbers, and diagrams to solve problems.
Standard for Mathematical Practice 5: Use appropriate tools strategically.	
Mathematically proficient students consider the available tools when solving a mathematical problem. These tools might include pencil and paper, concrete models, a ruler, a protractor, a calculator, a spreadsheet, a computer algebra system, a statistical package, or dynamic geometry software. Proficient students are sufficiently familiar with tools appropriate for their grade or course to make sound decisions about when each of these tools might be helpful, recognizing both the insight to be gained and their limitations. For example, mathematically proficient high school students analyze graphs of functions and solutions generated using a graphing calculator. They detect possible errors by strategically using estimation and other mathematical knowledge. When making mathematical models, they know that technology can enable them to visualize the results of varying assumptions, explore consequences, and compare predictions with data. Mathematically proficient students at various grade levels are able to identify relevant external mathematical resources, such as digital content located on a website, and use them to pose or solve problems. They are able to use technological tools to explore and deepen their understanding of concepts.	**GMP 5.1** Choose appropriate tools for your problem. **GMP 5.2** Use mathematical tools correctly and efficiently. **GMP 5.3** Estimate and use what you know to check the answers you find using tools.

Common Core State Standards for Mathematical Practice	*Everyday Mathematics* Goals for Mathematical Practice
Standard for Mathematical Practice 6: Attend to precision.	
Mathematically proficient students try to communicate precisely to others. They try to use clear definitions in discussion with others and in their own reasoning. They state the meaning of the symbols they choose, including using the equal sign consistently and appropriately. They are careful about specifying units of measure, and labeling axes to clarify the correspondence with quantities in a problem. They calculate accurately and efficiently, express numerical answers with a degree of precision appropriate for the problem context. In the elementary grades, students give carefully formulated explanations to each other. By the time they reach high school they have learned to examine claims and make explicit use of definitions.	**GMP 6.1** Communicate your mathematical thinking clearly and precisely. **GMP 6.2** Use the level of precision you need for your problem. **GMP 6.3** Be accurate when you count, measure, and calculate.
Standard for Mathematical Practice 7: Look for and make use of structure.	
Mathematically proficient students look closely to discern a pattern or structure. Young students, for example, might notice that three and seven more is the same amount as seven and three more, or they may sort a collection of shapes according to how many sides the shapes have. Later, students will see 7×8 equals the well remembered $7 \times 5 + 7 \times 3$, in preparation for learning about the distributive property. In the expression $x^2 + 9x + 14$, older students can see the 14 as 2×7 and the 9 as $2 + 7$. They recognize the significance of an existing line in a geometric figure and can use the strategy of drawing an auxiliary line for solving problems. They also can step back for an overview and shift perspective. They can see complicated things, such as some algebraic expressions, as single objects or as being composed of several objects. For example, they can see $5 - 3(x - y)^2$ as 5 minus a positive number times a square and use that to realize that its value cannot be more than 5 for any real numbers x and y.	**GMP 7.1** Find, extend, analyze, and create patterns. **GMP 7.2** Use patterns and structures to solve problems.
Standard for Mathematical Practice 8: Look for and express regularity in repeated reasoning.	
Mathematically proficient students notice if calculations are repeated, and look both for general methods and for shortcuts. Upper elementary students might notice when dividing 25 by 11 that they are repeating the same calculations over and over again, and conclude they have a repeating decimal. By paying attention to the calculation of slope as they repeatedly check whether points are on the line through (1, 2) with slope 3, middle school students might abstract the equation $(y - 2)/(x - 1) = 3$. Noticing the regularity in the way terms cancel when expanding $(x - 1)(x + 1)$, $(x - 1)(x^2 + x + 1)$, and $(x - 1)(x^3 + x^2 + x + 1)$ might lead them to the general formula for the sum of a geometric series. As they work to solve a problem, mathematically proficient students maintain oversight of the process, while attending to the details. They continually evaluate the reasonableness of their intermediate results.	**GMP 8.1** Use patterns and structures to create and explain rules and shortcuts. **GMP 8.2** Use properties, rules, and shortcuts to solve problems. **GMP 8.3** Reflect on your thinking before, during, and after you solve a problem.

Appendix F

Attitudes, Beliefs, and Confidence in Early Math (ABC-EM) Items

Two Factor Model

Confidence in Math Teaching (15 Items)

1 Even when I try, I don't teach mathematics as well as I teach many other subjects. *
2 I am confident in my ability to use a variety of assessment techniques to evaluate students' mathematical learning and progress.
3 I am confident in my ability to translate assessment results into mathematics teaching plans.
4 I am confident in my ability to set appropriate math learning goals for my students.
5 I am confident in my ability to anticipate problems and confusions that students might have with particular math topics or concepts.
6 I am confident in my ability to engage students in mathematics problem solving.
7 I am confident in my ability to facilitate students' communication about mathematics (for example, discussions, questions, and journals).
8 I am confident in my ability to encourage students to represent mathematics in a variety of ways (such as drawings, manipulatives, symbols, and language).
9 I am confident in my ability to connect mathematics learning to other curricular areas.
10 I am confident in my ability to help students reason about and prove how they have solved a mathematics problem.
11 I am confident in my ability to locate resources for preparing exciting and engaging math lessons.
12 I believe many times in my class I can get through to even the most difficult or unmotivated students.
13 I am confident in my ability to further students' math knowledge when they make spontaneous math comments or discoveries.
14 I believe that my pre-service education has sufficiently prepared me to teach mathematics.
15 I have the support I need to teach math well.

Positive Math Attitudes (13 Items)

16 I am not a "math person." *
17 I have a hard time quickly calculating arithmetic facts in my head. *
18 I can easily convert fractions into percentages or decimal numbers.
19 I'm good at looking at numeric data and finding patterns.
20 Math was one of my best subjects in school.
21 I am good at math puzzles.

22 Math is my least favorite subject to teach. *
23 I like doing math.
24 Math was one of my favorite subjects in school.
25 Just the word "math" can make me feel nervous. *
26 I enjoy teaching mathematics more than any other subject.
27 I can easily figure out how something would look from another angle.
28 I like teaching math to my students.