Design Research with Educational Systems: Investigating and Supporting Improvements in the Quality of Mathematics Teaching and Learning at Scale

Paul Cobb
Vanderbilt University

Kara Jackson
McGill University

Thomas Smith
Vanderbilt University

Michael Sorum
Fort Worth Independent School District

Erin Henrick
Vanderbilt University

The research project reported on in this chapter was supported by the National Science Foundation under grant Nos. ESI-0554535 and DRL-1119122. Kara Jackson's contributions to the chapter were also supported by the National Academy of Education/Spencer Postdoctoral Fellowship Program. The opinions expressed do not necessarily reflect the views of either Foundation. The work reported on in this chapter has been conducted in collaboration with Dan Berebitsky, Glenn Colby, Anne Garrison, Lynsey Gibbons, Karin Katterfeld, Adrian Larbi-Cherif, Christine Larson, Chuck Munter, Charlotte Munoz, Brooks Rosenquist, Rebecca Schmidt, and Jonee Wilson.

This chapter focuses on the Middle School Mathematics and the Institutional Setting of Teaching (MIST) project, which seeks to develop an empirically grounded theory of action for improving the quality of mathematics instruction at scale. At the time of writing, we are in the fifth year of an eight-year collaboration with urban school districts that focuses on middle-grades mathematics. As part of these collaborations, we test, revise, and elaborate the conjectures about school and district supports and accountability relations for instructional improvement that comprise our evolving theory of action. The current iteration of this theory of action comprises five interrelated components: curriculum materials and associated instructional guidance instruments such as curriculum frameworks; pull-out teacher professional development and school-based teacher collaborative meetings; mathematics coaches' practices in supporting teachers' learning; school leaders' practices as instructional leaders in mathematics; and district leaders' practices in supporting the development of school-level capacity for instructional improvement. (A presentation of the current theory of action is beyond the scope of the chapter. The reader is referrred to Cobb and Jackson (2011) for a discussion of each of these components.)

We limit our focus in this chapter to the first phase of MIST conducted 2007-2011 in which we partnered with four urban districts that served a total of 360,000 students. As we will clarify, the approach we took in collaborating with leaders in the four districts was consistent with the basic tenets of design-based implementation research (DBIR) as articulated by Penuel, Fishman, Cheng, and Sabelli (2011) and involved co-designing improvement strategies that were informed by data.

The motivation for MIST stems in large part from the observation that although research in mathematics education, teacher education, and related fields has made considerable progress in recent years, this work has had little influence on instruction in most US mathematics classrooms (Stigler & Hiebert, 1999). In our view, this limited impact is attributable in large measure to the lack of synergy between research on classroom teaching and learning on the one hand and research on educational policy and leadership on the other hand. Research on teaching and learning in mathematics and in other subject matter areas typically treats classrooms as existing in an institutional vacuum despite the abundant evidence that teachers' instructional practices are profoundly influenced by the school and district settings in which they work (for notable exceptions, see Nelson & Sassi, 2005; Stein & Spillane, 2005). As a consequence, this research has little to say about supports for instructional improvement that extend beyond issues of curriculum and teacher professional development. For its part, research in educational policy and leadership typically treats the classroom as a black box and fails to take a position on what counts as high quality instruction (for notable exceptions, see Coburn & Russell, 2008; Cohen, 2011; Spillane & Thompson, 1997). As a consequence, the resulting recommendations for school and district improvement strategies are relatively global and generic.

Given the limitations of the current research base, school and district leaders who are pursuing ambitious agendas for instructional improvement that involve significant teacher learning necessarily have to go into uncharted territory as they formulate and implement improvement strategies. Although there are coherent bodies of research on teacher professional development and productive teacher collaborative work (Borko,

2004; Horn & Little, 2010), research on content-focused coaching is underdeveloped, research on school instructional leadership in mathematics (and other content areas) provides conflicting advice (Resnick, 2010; Stein & Nelson, 2003), and research on the role of district leadership practices in supporting the development of school capacity for instructional improvement is extremely sparse (cf. Honig, 2008). Our intent in developing an empirically grounded theory of action for instructional improvement in mathematics is to begin to address some of these limitations.

In this chapter, we describe how we are attempting to achieve this goal and highlight how this work contributes to an understanding of the nature of *evidence* in DBIR. Penuel et al. (2011) noted that design-based implementation research is an emerging methodology and observed that the practical aspects of the work can easily overshadow the need for rigorous research. They went on to call for the establishment of

> shared norms and practices regarding theory development and the specification and testing of specific claims or conjectures. In other words, the approach needs to establish a distinctive "argumentative grammar" (Kelly, 2004) for judging the adequacy of data supports for particular claims and theories and for warrants that link claims to data. (Penuel et al., 2011, p. 355)

We first give an overview of MIST. Next, we clarify that the project involves both pragmatic and research-oriented goals, and we describe the research tools that we developed in our work with the four districts. We then focus on the role of evidence in achieving the pragmatic goal of the project as we describe the data collection, analysis, and feedback cycles that we conducted in each of the four partner districts each year in

order to provide them with timely evidence of how their strategies were playing out in schools. Finally, we focus on the research-oriented goals of the project and discuss the evidence on which we drew as we modified and elaborated the conjectures that comprised our emerging theory of action for instructional improvement in mathematics.

Project Overview

Any investigation that focuses on instructional improvement necessarily takes a position on what counts as high-quality teaching, and thus on what is worth knowing and doing mathematically (Hiebert & Grouws, 2007). The goals for students' mathematical learning on which we focused included conceptual understanding of central mathematical ideas, justifying and generalizing solutions, and making connections between multiple representations of mathematical ideas, as well as procedural fluency. Research in mathematics education, the learning sciences, and related fields indicates that students' attainment of these learning goals requires frequent opportunities to solve cognitively-demanding tasks (i.e., non-routine tasks that have the potential to support students in connecting mathematical representations and ideas) (Stein, Smith, Henningsen, & Silver, 2000). In addition, attainment of the goals requires that students participate in whole class discussions in which the teacher presses them to justify their reasoning and to make connections between solutions (Franke, Kazemi, & Battey, 2007; Stein, Engle, Smith, & Hughes, 2008). Instructional practices of this type have been called ambitious because they aim at rigorous goals for students' learning (Lampert, 2001). Research on teacher learning indicates that most teachers' development of ambitious classroom practices involves the reorganization rather than the mere elaboration or extension of current practices (Kazemi, Franke, & Lampert, 2009). The findings of a number of studies

indicate that significant teacher learning of this type requires substantial support for an extended period of time (Darling-Hammond, Wei, & Orphanos, 2009). Our primary research goal of developing a theory of action for instructional improvement in mathematics therefore involved testing and refining conjectures about supports and accountability relations that scaffold and press for teachers' reorganization of their instructional practices.

Recruitment of Partner Districts

Given our focus on ambitious mathematics instruction, our first criterion when recruiting districts was that they were aiming at rigorous goals for students' mathematical learning. Our second criterion was that they were responding to high-stakes accountability pressures by attempting to implement strategies that went beyond "teaching to the test" by supporting and holding teachers accountable for improving the quality of their instructional practices. Recruiting districts was challenging because our planned data collection was quite extensive. Although we promised to provide feedback on how their improvement strategies were playing out in schools, we could not provide examples of such feedback at the outset of the project. After several false starts, we were eventually able to recruit four districts that met our criteria. A leading intermediary organization that supports the work of urban districts, Institute for Learning (IFL), generously facilitated our entry into three of the districts. The final district was participating in another project that was led by a member of the MIST research team. Although there was some variation in the districts' improvement strategies, they all expected teachers to use instructional materials for middle-grades mathematics that aimed at ambitious goals for students' learning and were providing professional development

for teachers and school leaders to support the implementation of these materials.

The four districts were all coping with a range of challenges typical of urban districts including limited financial resources, a high proportion of students from poor communities, high teacher turnover, a high proportion of novice teachers, and highly public Federal and State accountability systems. As an illustration, at the time that we began working with one of our partner districts, Fort Worth Independent School District (FWISD), a new superintendent and her staff were dealing with massive budget cuts that gutted support for schools, negative press surrounding vendors and bond programs, a new testing and accountability system, and serious student achievement issues. In 2005, only 31% of African American and 44% of Hispanic students achieved standard on the first administration of the eighth-grade mathematics Texas Assessment of Knowledge and Skills (TAKS).

In response, FWISD leaders drew on work they were conducting with IFL to develop a comprehensive set of instructional improvement strategies that included: adopting a high-quality standards-based, inquiry-oriented curriculum for middle-grades mathematics; developing an accompanying curriculum framework that identified relevant instructional resources and provided guidance on differentiating instruction; hiring mathematics coaches for each middle-school who taught half day and coached their peers the other half of the day; hiring a small cadre of central office mathematics specialists; and providing ongoing professional development for mathematics teachers and coaches, and for principals and assistant principals to support their development as instructional leaders in mathematics. FWISD leaders had prescribed roles and responsibilities for each "layer" of the organization to support the implementation of these strategies, and were

using IFL tools such as Learning Walks™ to monitor the quality of implementation. The process of recruiting FWISD involved extended consultations both in person and by phone, before district leaders concluded that the proposed collaboration had the potential to add value to their work by clarifying their improvement strategies and by co-constructing an evaluation of the work.

Once leaders in FWISD and the other three districts agreed to work with us, we requested their assistance in selecting approximately six middle-grades schools that were representative across each district in terms of their capacity for instructional improvement. We explained that we wanted to provide them with feedback that would be applicable to middle-grades schools in their districts in general, not only to schools in our sample. Once we had identified schools, we recruited 30 randomly selected mathematics teachers, the mathematics coaches that served these schools, and the school leaders (i.e., the principal and any assistant principals who were responsible for monitoring mathematics instruction), as well as district leaders across central office units, for a total of approximately 50 participants in each district.

Levels of Analysis and Research Tools

Our collaboration with the four districts involved conducting data collection, analysis, and feedback cycles in each district each year. Each cycle (discussed in detail below) involved documenting the district's improvement strategies, collecting and analyzing data to assess how these strategies were being implemented, and reporting the findings to the district and making recommendations about how the strategies might be revised. In conducting this work, we differentiated explicitly between two distinct levels of analysis. The first level was pragmatic and involved providing the districts with

timely evidence of how their strategies were playing out in schools. The second level focused on our primary research goal of developing an empirically grounded theory of action for instructional improvement that is generalizable and could inform instructional improvement in other districts. At the second level, we assembled evidence to test, revise, and elaborate conjectures about supports and accountability relations that we had developed prior to working with the four districts by drawing on the available research literature. These two levels of analysis were interdependent in that insights that we developed while formulating empirically grounded recommendations to the districts informed the revision of our theory of action (see Figure 1). Conversely, the current iteration of the theory of action was an essential research tool at each point in our collaboration with the partner districts and guided our formulation of recommendations about how they might revise their improvement strategies.

[Insert Figure 1 about here]

Interpretive Framework

We found it essential to create a second research tool that we could use to assess the potential of each district's designed or intended strategies to contribute to instructional improvement, and to account for the consequences of the strategies once they were implemented. This tool is an interpretive framework that distinguishes between four general types of supports that capture all the improvement strategies that our four partner districts attempted to implement across the four years (2007-2011): new positions, learning events (including professional development), organizational routines, and tools. In developing the framework, we drew on research in the learning sciences, teacher learning, and related fields to assess the potential of each general type of support

to scaffold teachers', coaches', and school leaders' reorganization of their practices. We clarify the nature of each type of support and its potential to support practitioners' learning in the following paragraphs. As will become apparent, the framework reflects the view that co-participation with others who have already developed relatively accomplished practices is essential when the learning demands of an improvement strategy require the reorganization rather than the extension or elaboration of current practices (Lave & Wenger, 1991; Rogoff, 1997; Sfard, 2008).

New positions. District instructional improvement strategies often include the creation of new positions whose responsibilities include supporting others' learning by providing expert guidance (Bryk, 2009; Spillane & Thompson, 1997). For example, FWISD created the position of a school-based mathematics coach in each middle school whose responsibilities included providing support both for groups of teachers and for individual teachers in their classrooms, and supporting their principals' work as instructional leaders in mathematics. The extent to which the investment in the new position will pay off is likely to be influenced by the expertise of the appointees and the extent to which the appointees and the people they are expected to support co-participate in activities that are close to the intended forms of practice.

Learning events. Professional development sessions for school leaders and members of other role groups are instances of learning events, which we define as scheduled meetings that can give rise to opportunities for participants to improve their practices. Several distinctions proved useful when assessing the four partner districts' intended improvement strategies. First, o*ngoing intentional learning events* are designed as a series of meetings that build on one another, and typically involve a relatively small

number of participants. Because a small number of participants is involved, the group has the potential to evolve into a genuine community of practice that works together for the explicit purpose of improving their practices (Brown & Duguid, 1991; Rogoff, 1994). In contrast, *discrete intentional learning events* include series of meetings that are not designed to build on each other, including one-off professional development sessions. Although, discrete intentional learning events are, by themselves, unlikely to support significant reorganizations of practice (Lave, 1993), they have the potential to support the elaboration or extension of current practices (e.g., training school leaders on a classroom observation tool that is consistent with their current views of high-quality mathematics instruction).

Learning opportunities are not limited to *intentionally* designed learning events. They can also arise *incidentally* as school and district personnel collaborate with others with greater expertise to carry out functions of the school or district. For example, middle-school principals and mathematics coaches in FWISD were expected to meet weekly to discuss the quality of mathematics teaching in their school. It is possible that learning opportunities could arise for principals in the course of these meetings as they discussed their classroom observations with a coach who, presumably, had greater expertise in mathematics instruction. In general, the extent to which interactions with a more knowledgeable other involve significant learning opportunities depends on whether the interactions focus on problems of practice (Ball, Sleep, Boerst, & Bass, 2009; Lampert, 2010). However, incidental learning events might, by themselves, be inadequate to support significant professional learning because interactions are not

oriented by the explicit purpose of improving practice and are therefore unlikely to provide sustained support for the development of particular forms of practice.

New organizational routines. Instructional improvement strategies sometimes include the specification of new organizational routines. Feldman and Pentland (2003) define organizational routines as "repetitive, recognizable patterns of interdependent actions, carried out by multiple actors" (p. 94). As Sherer and Spillane (2011) clarify, the introduction of carefully designed organizational routines can be an important means of supporting learning. As an illustration of an organizational routine, leaders in FWISD expected that middle-school principals would conduct Learning Walks™ with the mathematics coach at their schools on a regular basis. A Learning Walk™ is a repetitive, recognizable pattern of actions that involves determining the focus of classroom observations (e.g., the extent to which teachers maintain the cognitive challenge of tasks throughout the lesson), selecting classrooms to visit, observing a classroom, and then conferring to discuss observations before moving on to the next classroom. Organizational routines in which a more knowledgeable other scaffolds relative novices' learning as they co-participate in activities that are close to practice is a potentially productive means of supporting the reorganization of practice (Grossman & McDonald, 2008; Lampert & Graziani, 2009).

New tools. By tools, we mean material entities that are used instrumentally to achieve a goal or purpose. Large-scale instructional improvement efforts almost invariably involve the introduction of a range of new tools designed to be used in practice, including newly adopted instructional materials and revised curriculum frameworks for teachers, and new classroom observation protocols and data management

systems for principals. The findings of a number of studies conducted in the learning sciences substantiate Pea's (1993) claim that the incorporation of a new tool into current practices can support the reorganization of those practices (Lehrer & Schauble, 2004; Meira, 1998; Stephan, Bowers, & Cobb, 2003). However, the findings of a number of studies of policy implementation and of teaching indicate that practitioners often assimilate new tools to their current instructional practices rather than reorganize their practices as intended (Cohen & Hill, 2000; Remillard, 2005; Spillane, 1999). These findings suggest that the introduction of tools whose effective use requires the reorganization of current practice should be coordinated with supports that involve co-participation with an already accomplished user.

Based on the interpretive framework of general types of supports, we anticipated that strategies that support consequential professional learning would involve some combination of new positions to provide expert guidance, ongoing intentional learning events, carefully designed organizational routines carried out with a more knowledgeable other, and the use of new tools whose incorporation into practice is supported. We did not discount the support that discrete intentional learning events and incidental learning events might provide and took them into account when analyzing the partner districts' designed improvement strategies.

Pragmatic Level of Analysis: Evidence for Feedback to Partner Districts

The data collection, analysis, and feedback cycle that we conducted in each district each year and the role of the two research tools in the cycle are shown in Table 1. In discussing our pragmatic goal of providing the districts with timely feedback to inform the revision of their improvement strategies, we distinguish between the forms of

evidence we used to 1) document each district's designed or intended improvement strategies, 2) assess how those strategies were being implemented in schools, and 3) make recommendations to adjust the strategies.

[Insert Table 1 about here]

Claims About Districts' Designed Instructional Improvement Strategies

The data we collected each October to document each district's designed improvement strategies included audio-recorded interviews conducted with approximately ten leaders from each district, as well as artifacts detailing the planned implementation of these strategies (e.g., district organizational charts, schedules of professional development sessions, curriculum frameworks, school leader and teacher evaluation forms). The leaders we interviewed were from the units of Curriculum and Instruction responsible for developing instructional guidance tools (e.g., curriculum frameworks) and for providing professional development for mathematics teachers and mathematics coaches, and from the Leadership department responsible for supporting and holding school leaders accountable. The interviews focused on the district's goals for middle school mathematics instruction and the strategies the district was implementing to achieve these goals (e.g., developing curriculum frameworks, providing pull-out and job-embedded teacher professional development, scheduling time for mathematics teachers to collaborate during the school day, providing mathematics content expertise through coaching, supporting the development of school instructional leadership in mathematics).

We analyzed transcriptions of these interviews to develop an account of each district's designed instructional improvement strategies. In doing so, we triangulated the responses of different leaders in a district and were usually able to discern broad

consistencies across interviews. We recorded the results of these analyses in *District Design Documents* of 3-4 single-spaced pages in which we named each strategy, specified the envisioned forms of practice that constituted the goal of each strategy, and described the intended supports and accountability relations for the development of the envisioned practices.

As an illustration, one of FWISD's three primary improvement strategies in the second year of our partnership was to support principals' development as instructional leaders in mathematics. The envisioned forms of practice specified in the *District Design Document* included that principals would observe classroom instruction for at least two hours each day and would provide feedback to teachers that would communicate appropriate expectations for instructional improvement. The intended supports included: monthly principal meetings that would periodically focus on mathematics instruction (discrete intentional learning events); weekly meetings between the principal and the coach who served the school to discuss the quality of mathematics teachers' instruction and to plan professional development (incidental learning events); Learning Walks™ in which the principal and coach observed instruction together (new organizational routine); and the provision of curriculum maps that included a pacing schedule for each six-week instructional module together with descriptions of the mathematics concepts being taught, resources teachers should use, and expected student products (new tool). The intended accountability relations included that the members of the Leadership Department who worked directly with principals would hold them accountable for supporting improvements in the quality of teachers' instructional practices (e.g., observing classroom instruction, conducting Learning Walks™).

Each November, we then shared the *District Design Documents* with district leaders to determine whether we had accurately represented their improvement goals and intended strategies. The triangulation of district leaders' responses and this member check constituted the primary evidence for our claims about each district's designed improvement strategies. Leaders in all four districts indicated that they found the documents useful because they provided a succinct codification of their improvement strategies. The documents were therefore important communication tools in achieving a consensual understanding of the districts' instructional improvement initiatives.

We also developed an in-house version of each *District Design Document* in which we included an assessment of possible limitations of the district's intended strategies by assessing: 1) whether the district's strategies were coherent (i.e., mutually reinforcing or at odds with each other), 2) whether each individual strategy was likely to contribute to instructional improvement if members of particular role groups (e.g., school leaders, mathematics coaches, teachers) developed the intended practices, and 3) whether the supports and accountability relations included in each individual strategy were likely to enable members of particular role groups to develop the intended practices. The evidence for our conclusions for the first and second questions initially stemmed from our reading of the literature on instructional policy, professional learning, and teacher learning. In subsequent years, our conclusions were also grounded in the conjectures about school and district supports and accountability relations for instructional improvement that comprised our emerging theory of action (see Table 1). The evidence for our conclusions for the third question stemmed primarily from our use of the

interpretive framework to assess the potential of the districts' improvement strategies (see Table 1).

In the case of FWISD's strategy of principals becoming instructional leaders in mathematics, we noted that the only intended support that might involve co-participation in activities close to the envisioned leadership practices with a more accomplished colleague were the planned Learning Walks™ in which a principal and coach would observe instruction together. Crucially, the intended supports for principals did not include any ongoing intentional learning events. We questioned internally whether Learning Walks™, monthly principal meetings, and weekly meetings with a coach would be sufficient to support the principals' development of more sophisticated visions of high-quality mathematics instruction. In addition, we anticipated that the principals would assimilate the curriculum maps to their current observational practices because no supports for using the maps were planned.

We did not include our in-house assessments of the districts' improvement strategies in the *District Design Documents* that we shared with district leaders in November of each year because it would not have been possible for them to make significant changes to their strategies midway through the school year. However, we did draw on these assessments when we prepared the feedback reports that we shared with district leaders each May after we had documented how their improvement strategies were playing out in schools.

Claims About the Implementation of District Instructional Improvement Strategies

We collected multiple types of data in each district each January-March in addition to conducting interviews with the 200 participants as indicated in Table 1.

However, we limited the data we analyzed to assess how the districts' strategies were being implemented to the January interviews and artifacts so that we could provide the districts with feedback before district leaders began planning strategies for the following school year over the summer. (We describe the other types of data in a subsequent section of this chapter.) The audio-recorded interviews focused on both the formal and informal supports for members of each role group, and to whom and for what they perceived themselves to be accountable. The interviews were designed to assess each construct in our initial conjectures about school and district supports for instructional improvement (Cobb & Smith, 2008). The nine separate protocols we used for teachers, mathematics coaches, and school leaders, and district leaders in different central office units are downloadable at http://vanderbi.lt/mist.

The first step in the analysis of these interview transcripts involved completing an Interview Summary Form (ISF) for each interview. We created separate summary forms that were specific to the issues addressed in the teacher, mathematics coach, school leader, and district leader interview protocols, and customized the forms based on our analysis of each district's intended improvement strategies. For example, the ISF for teacher interviews used in FWISD reported each teacher's responses to the following issues: 1) the frequency and types of interactions that the teacher had with colleagues about mathematics instruction, both informally and in scheduled meetings; 2) the teacher's perceptions of the principal's and coach's instructional expectations, whether the principal observed the teacher's instruction and the nature of any verbal or written feedback received, and any assistance the coach provided (e.g., observing instruction, modeling, co-teaching); 3) the professional development in which the teacher had

participated, and sessions that the teacher found most valuable, 4) the teacher's vision of high quality mathematics instruction and the teacher's assessment of the extent to which other teachers and instructional leaders in the school shared this view, 5) the instructional materials, tools, and resources that the teacher used and how he or she used them when preparing for and conducting classroom lessons, and 6) whether and how the teacher adjusted instruction for struggling students.

The second step in the analysis involved completing mid-level summary forms by synthesizing the ISFs for all participants in a school (i.e., school summary forms) and for each role group across a district (i.e., teachers, coaches, and school leaders). As with the ISFs, the mid-level summary forms were customized for each district. In conducting these syntheses, we triangulated participants' responses. For example, our inferences about how frequently a particular principal observed mathematics instruction and about the nature of his or her feedback were based on a comparison of the principal's and teachers' accounts. We required that all claims made on the mid-level summary forms about the practices of members of a particular role group and about how particular strategies were playing out in a particular school had to be justified by citing specific evidence identified in the relevant ISFs.

The final step in determining how the districts' strategies were being implemented involved looking across the mid-level summary forms for each district to determine both the practices that teachers, coaches, and school leaders were developing, as well as the supports and accountability relations that had been established. Evidence for the resulting claims stems from the ethnographic criterion of trackability (Smaling, 1992). The criterion of trackability is closely related to reliability and indicates when we can

justify findings by backtracking through the successive steps of the analysis to the interview transcripts and audio-recordings.

Recommendations to Districts About Revising Their Improvement Strategies

Before we could make recommendations to the districts about how they might revise their improvement strategies, we had to explain *why* their strategies were playing out in the ways that we had documented rather than as district leaders had intended (Argyris & Schön, 1974, 1978). To develop these explanations, we first assessed whether either the nature or the frequency of practices of members of particular role groups had changed from the previous year. As an illustration, all of our partner districts attempted to support and hold principals accountable for observing mathematics teachers' instruction and providing feedback that communicated appropriate expectations for instructional improvement. The changes we documented in principals' practices during the first phase of MIST (2007-2011) were similar in all four districts. Although the frequency of their classroom observations increased, sophistication of their visions of high-quality mathematics instruction and the nature of the feedback they gave teachers showed little improvement.

For each district each year, we accounted for changes (or the lack of change) in participants' practices by assessing both the learning opportunities and press for improvement afforded by the implemented supports and accountability relations. In making these assessments, we used one of our primary research tools, the interpretive framework that differentiates between four general types of supports (see Table 1). In the case of FWISD's strategy of developing principals as instructional leaders in mathematics, our analysis of the implementation indicated that the monthly principal

meetings did periodically focus on mathematics instruction. Furthermore, all principals met with their coach regularly. However, in half of the participating schools, the discussions between principals and coaches tended to focus on which mathematics topics were being taught and when they were being taught but not how they were taught, and there was little evidence that principals and coaches planned professional development together. It also became evident that only one principal had conducted Learning Walks™ with a coach, due to scheduling difficulties, and that the principals were not using the curriculum maps to guide their classroom observations. In addition, all the principals reported that members of the Leadership department held them primarily accountable for improving student achievement on state assessments, and only secondarily for supporting improvements in the quality of instruction.

When developing the *District Design Document* for FWISD earlier in the year, we had anticipated that the designed supports for school leaders might not be adequate even if they were implemented as intended as there were few potential learning opportunities that involved co-participation with someone more expert in mathematics instruction. Only the planned Learning Walks™ with a mathematics coach had the potential to involve co-participation in activities close to practice. As Learning Walks™ with a coach occurred only rarely and as principals' meetings with a coach often focused only on issues such as pacing, we were not surprised that the sophistication of the principals' visions of high-quality mathematics instruction and the quality of their feedback to teachers had not improved from the prior year. We also concluded that principals' work with coaches might have been less productive than intended, at least in part, because they perceived themselves to be primarily accountable for raising test

scores but not necessarily for supporting teachers in improving the quality of their instructional practices.

Our explanations for the differences between the districts' intended and implemented improvement strategies grounded our formulation of recommendations for how district leaders might revise their improvement strategies. In developing these recommendations, we drew on the conjectures about supports and accountability relations that comprised the current iteration of our theory of action for instructional improvement at scale (see Table 1). In the illustrative example of supporting principals in becoming instructional leaders in mathematics, we recommended that FWISD leaders address the tension that principals reported between improving the quality of instruction in the long-term and raising students' test scores in the short-term by clarifying expectations for principals' intended practices with members of both the Leadership and the Curriculum and Instruction departments. We indicated the importance of district leaders being explicit about what members of the Leadership Department should hold principals accountable for and how members of both departments should support principals in meeting these expectations. We also recommended that these supports include sustained professional development for principals (i.e., ongoing intentional learning events) that focused on discriminating between strong and weak enactments of inquiry-oriented mathematics lessons. In this and in other cases, the evidence for our recommendations stems from the relatively rigorous manner in which we documented the districts' intended and implemented strategies, from our use of the interpretive framework to account for differences between intended and implemented strategies, and from our use of the current iteration of our theory of action to formulate recommendations.

The final products of each annual cycle were *District Feedback and Recommendations Reports* (DFRRs) of approximately 15 single-spaced pages that we prepared for leaders in each district. These reports built directly on the *District Design Documents* and were intentionally structured around the district's major strategies so that they related directly to the work district leaders were attempting to accomplish. For each strategy reported in the *District Design Document*, we reiterated the envisioned forms of practice that constituted the goal of the strategy and described the intended supports and accountability relations for the development of the envisioned practices. We then reported our findings about how that strategy was playing out in schools, explained why this was the case, and made our recommendations for adjusting the strategy. For example, the report we prepared for FWISD leaders in the second year of our collaboration included our findings and recommendations summarized above for their strategy of supporting principals in becoming instructional leaders of mathematics. We submitted the reports to districts leaders in May, approximately one week in advance of a two-hour meeting scheduled in each district with leaders of the departments that were involved in the improvement effort. We requested that these meetings be held in conference rooms and spoke from notes rather than PowerPoint slides as we summarized our findings and recommendations in order to encourage an open discussion. It was clearly understood that district leaders retained the authority to determine the improvement strategies that would be implemented. Consequently, our role was advisory and we viewed our recommendations as proposals that were open to debate. Although our findings were often disappointing for district leaders, they were never defensive and

instead engaged in an open dialogue about the current status of the district's improvement efforts and about possible modifications to those efforts.

The following October, after district leaders had developed plans for the new school year over the summer, we interviewed them again to document their revised instructional improvement strategies. Our analyses of these interviews indicated that leaders in all four districts acted on many of our recommendations. In the illustrative case of the recommendations we made to FWISD about instructional leadership, we found the following year that district leaders required members of the Leadership department to communicate expectations to principals about their role as instructional leaders in mathematics by conducting Learning Walks™ with them. In addition, the monthly principal meetings included ongoing professional development on recognizing high-quality mathematics instruction that was specific to the instructional materials teachers were using (ongoing intentional learning events). This illustration is representative in that leaders in all four districts acted on our recommendations to a significant extent.

In the case of FWISD, district leaders indicated that the work of the partnership held significant sway with them for two reasons. First, the findings shared with them were based on extensive interviews conducted with personnel at all levels of the District—from district leaders in different central office units to teachers. Second, the findings were organized in terms of FWISD's improvement strategies, thereby allowing district and school leaders to triangulate the findings with student results and their own classroom observations. FWISD leaders also clarified that although their work in middle-school mathematics was a small piece of a larger reform agenda, their use of a

single set of coherent strategies across grade levels and content areas made it easy to transpose findings and recommendations to their work in other content areas and grade levels. They were therefore able to scale the partnership work up to high-school mathematics and out to science at both the middle- and high-school levels. In sharing these observations, FWISD leaders stressed that urban school reform is hard, slow work. While the district has celebrated that achievement gaps have narrowed, middle school mathematics results remain painfully low. In 2011, only 54% of African American and 60% of Hispanic students achieved standard, 23 and 16 point increases respectively since 2005. However, as the district was responding to the introduction of a new state test and a new accountability system, leaders were confident that the structures and resources they had put in place were stronger and better suited to adaptation because of the partnership work.

The extent to which district leaders saw value in partnership work and acted on our recommendations is a relatively strong indication that MIST was successful in achieving its pragmatic goal of providing useful and timely feedback to the districts. In this regard, our partnership with the districts exemplified two of the key principles of DBIR identified by Penuel et al. (2011): "a focus on persistent problems of practice" from the perspectives of both practitioners and researchers, and "a commitment to iterative, collaborative design" (p. 332).

Research-Oriented Level of Analysis: Evidence for Revising Theory of Action for Instructional Improvement in Mathematics

To this point, we have focused on the first of the two levels of analysis, in which we attempted to add value to the partner districts' instructional improvement efforts. As

shown in Figure 1, our collaboration with the districts grounded the second level of analysis at which we addressed our primary research question by developing a theory of action for large-scale instructional improvement in mathematics. As described earlier, in its current iteration, the theory of action comprises five interrelated components that range from supports integral to classroom practice such as curriculum materials and associated instructional guidance instruments to district leadership practices. Here, we focus on one component, school leaders' practices as instructional leaders, to illustrate the types of evidence on which we drew as we revised the conjectures that comprise the theory of action.

As we have indicated, all four districts in our study attempted to support school leaders' development as instructional leaders in mathematics. Our initial conjectures about school instructional leadership were relatively global and did not differentiate between the practices of school leaders (i.e., principals and assistant principals) and mathematics coaches. While, these conjectures focused on the importance of members of both role groups having a relatively deep understanding of the mathematical intent of the instructional program, an appreciation of the challenges that teachers face in using the program effectively, and an understanding of the challenges in supporting teachers' reorganization of their instructional practices, they did not specify the relationship between the two role groups. Our current conjectures emphasize the potential value of school leaders and coaches assuming shared responsibility for school instructional improvement, with school leaders pressing and holding teachers accountable for developing ambitious instructional practices and coaches supporting teachers' development of those practices. In addition, our current conjectures specify how school

leaders might press teachers by delineating two leadership practices: observing mathematics instruction and providing feedback that communicates appropriate instructional expectations, and participating in mathematics teacher collaborative meetings. School leaders' development of the first of these practices would be a significant accomplishment as they would have to recognize and distinguish between low- and high-quality enactments of ambitious instructional practices in order to provide feedback that could orient teachers towards these instructional expectations.

The changes in our conjectures about school instructional leadership are representative of the revisions we made to the theory of action more generally. In the course of working with the districts for four years, we elaborated conjectures for all five components by specifying both potentially productive practices for members of various role groups, and supports and accountability relations necessary for the development of those practices. In the following paragraphs, we discuss and illustrate the three types of evidence on which we drew as we revised our conjectures: our findings about how the districts' instructional improvement strategies were being implemented, our reading of the research literature, and the findings of retrospective analyses that we are conducting by drawing on data we collected each year.

Findings About the Districts' Instructional Improvement Strategies

In formulating recommendations to our partner districts about how they might adjust their improvement strategies, we necessarily had to address concrete organizational design challenges by proposing how each district might support and hold members of particular role groups accountable for improving their practices that the districts had the capacity to enact. Addressing these challenges was a primary context for

our learning as we sought to understand what it might take to support instructional improvement in mathematics on a large scale. Our focus when formulating recommendations for a particular district was pragmatic given that the recommendations could be consequential for the mathematics instruction of a significant number of students. However, once we completed each annual data collection, analysis, and feedback cycle, we stepped back and framed our findings about and recommendations for the districts' improvement strategies as cases of attempting to support instructional improvement at scale. In doing so, we considered whether any of our recommendations to a particular district represented refinements or elaborations of our current conjectures, and if they did whether they might have more general implications. Collaborating with four districts was helpful in addressing the question of potential generality as we could consider whether a recommendation made to one district might be feasible for the other districts and, if so, under what conditions.

We illustrate how our findings about the implementation of the districts' improvement strategies constituted evidence for the revision of our conjectures by focusing on our current conjecture concerning school leaders pressing and holding teachers accountable by observing instruction and providing feedback. Current research on school instructional leadership provides contradictory guidance on what principals need to know and do in order to provide effective direct support for instructional improvement in mathematics. Some researchers propose that school leaders should, in effect, act as coaches and that they need a deep understanding of the mathematical content on which instruction focuses, students' mathematical learning, and teacher learning (Nelson & Sassi, 2005; Stein & Nelson, 2003). This proposed line of reasoning

seemed unfeasible in the case of our collaborating districts as they did not have the capacity to support school leaders, most of whom were not mathematics specialists, in developing these capabilities. Other researchers have argued that school leaders can provide teachers with effective feedback if they understand general, content-independent principles of learning and instruction (Resnick & Glennan, 2002).

We initially regarded the extent to which school leaders need content-specific expertise in order to provide effective feedback to mathematics teachers as an open question. However, our findings regarding the impact of districts' improvement strategies proved relevant, as school leaders in three of the districts received ongoing professional development that focused on content-independent characteristics of high-quality instruction. In the interviews we conducted each year, we asked school leaders a series of questions to assess their visions of high-quality mathematics instruction and we questioned teachers about the nature of the feedback they received from these school leaders. Our findings indicated that the school leaders in these three districts could not distinguish between strong and weak enactments of ambitious mathematics instruction after they had received the content-independent professional development, and that the quality of the feedback they gave teachers did not communicate expectations for the kinds of instructional improvement that district strategies are designed to support (Cobb & Jackson, in press).

These findings provided evidence that the content-independent principles on which the professional development had focused were too global for school leaders to connect to instructional practice in a specific content area such as mathematics. We therefore refined our conjectures about school instructional leadership by specifying that

it is important that school leaders be able to distinguish between low- and high-quality enactments of key instructional practices that are integral to ambitious mathematics teaching (e.g., differentiate between high-and low-cognitively demanding mathematics tasks and between whole class discussions that are productive and unproductive in terms of student learning opportunities). As a consequence of this revision, the recommendations that we subsequently made to the districts emphasized that supports for school leaders' learning should be content-specific and should be organized around the instructional materials that teachers were expected to use.

We have noted that the leaders of the four districts attempted to implement many of our recommendations, including those for school instructional leadership. We therefore had the opportunity to investigate the consequences of our recommendations in subsequent years, and thus the conjectures on which they were based. In the process, we became co-designers of district improvement strategies and, in effect, conducted four parallel design experiments at the level of a large district as we annually tested and revised our conjectures about supports and accountability relations for instructional improvement at scale (Cobb & Smith, 2008). In a very real sense, our co-participation with district leaders in the practice of designing to support instructional improvement was a primary context for our learning.

Research Literature

We have noted that current relevant research that can inform the design of instructional improvement strategies becomes increasingly thin the further one moves away from the classroom. Nonetheless, findings reported in the literature have, on occasion, provided evidence for the revision of our conjectures. As an illustration, one of

our current conjectures concerns school leaders' participation in mathematics teacher collaborative meetings. Our delineation of this leadership practice was informed in part by the finding of a synthesis conducted by Robinson, Lloyd, and Rowe (2008) of studies that have investigated the relationship between school leadership practices and student outcomes. This analysis indicated that school leaders' participation in teacher professional development is strongly associated with improvements in student achievement. Our focus on this leadership practice was also informed by our finding that the types of activities in which teachers in the four districts were engaging during collaborative meetings were frequently of insufficient depth to support their development of ambitious instructional practices. School leaders' participation in the meetings has the potential to both signal the importance of teacher collaboration and hold teachers accountable for using collaborative time productively. In addition, this conjecture was informed by the previously mentioned finding that the feedback that school leaders gave teachers after observing instruction was generally not effective in orienting teachers' improvement efforts. By participating in collaborative meetings school leaders might also learn about the specific aspects of instruction on which their teachers are working, thereby enabling the focus of their classroom observations and feedback to be more tightly tied to teachers' problems of practice.

Retrospective Analyses

Earlier in this chapter, we indicated that we collected a wide range of different types of data in each district each January-March in addition to the interviews conducted with the 200 participants. These additional data included: on-line surveys for teachers, coaches, and school leaders that quantified many of the issues addressed in-depth in the

interviews; video-recordings of two consecutive mathematics lessons in each of the 120 participating teachers' classrooms, which were then coded with the Instructional Quality Assessment (Boston, in press); teachers' and coaches' scores on the Mathematics Knowledge for Teaching instrument (Hill, Schilling, & Ball, 2004); video-recordings of select district professional development; audio-recordings of teacher collaborative planning meetings; and an on-line assessment of teacher networks completed by all 300 middle-grades mathematics teachers in the participating schools. In addition, the districts provided us with access to student achievement data on state assessments.

Our work during the first two of the four data collection, analysis, and feedback cycles focused on developing the method we have described for analyzing the 200 interviews in order to provide the districts with empirically grounded feedback and on developing measures of constructs that are central to our conjectures (e.g., measures of the sophistication of participants' visions of high-quality mathematics instruction (Munter, 2009) and of their views of students' mathematical capabilities (Jackson, 2011)). As a consequence, we did not begin conducting retrospective analyses that draw on the additional types of data until the third year of data collection.

The five lines of retrospective analyses that we are currently conducting correspond to the five components of our emerging theory of action for instructional improvement in mathematics. The initial analyses that we completed provided evidence for the revision of our conjectures about supports and accountability relations for instructional improvement. For example, one analysis indicated that principals play a critical role in enabling mathematics coaches to be effective in supporting teachers' improvement of their instructional practices (Gibbons, Garrison, & Cobb, 2011). In

addition, we have also found that teachers' access to a colleague such as a coach who has instructional expertise is one of the strongest predictors of improvement in the quality of instruction (Smith et al., 2012). We modified our conjectures about school instructional leadership in light of these findings to highlight the potential importance of school leaders and coaches working together and assuming shared responsibility for school instructional improvement in mathematics. In this and in other cases where retrospective analyses have informed the revision of our conjectures, the justification for the revisions stems from the methods used to conduct the analyses. In the example of the school leaders assuming shared responsibility for instructional improvement with coaches, the warrant stems from the rigor of a cross-case comparative analysis (Yin, 2003).

Discussion and Conclusion

In this chapter, we have highlighted the evidence for our pragmatic and research-oriented claims. The evidence for our recommendations to the partner districts indicates the importance of conducting member checks about district strategies and developing analytical methods that ensure claims can be tracked back to original data sources. The evidence for the revisions we made to our theory of action indicates the importance of establishing research-practitioner partnerships that involve co-designing, testing, and refining current school and district design conjectures. The empirical grounding for and the interdependence of our pragmatic and research conjectures illustrates a core principle of design-based implementation research articulated by Penuel et al. (2011): conducting systematic inquiry to develop theory related to improving the quality of classroom instruction and student learning at the system level.

Given DBIR's current status as an emerging methodology and the limited guidance for instructional improvement at scale provided by current research, it is reasonable to expect that the systematic inquiry to which Penuel et al. refer will be a bootstrapping process. This aspect of DBIR is evident in the development of our two primary research tools, our theory of action for instructional improvement and the interpretive framework for assessing improvement strategies. As we have illustrated, we used the most current iteration of our theory of action to inform the formulation of the feedback recommendations to the districts each year. We then used what we learned as we worked through each data collection, analysis, and feedback cycle to revise and elaborate the conjectures that comprise the theory of action. Similarly, we developed and refined the interpretive framework during the first two years that we worked with the four districts by organizing the various supports that the districts were attempting to implement into conceptually meaningful categories.

It is worth clarifying that the theory of action and interpretive framework exemplify two general types of research tools that are, in our view, essential for DBIR. The school and district design conjectures that comprise the theory of action both informed our work with the partner districts and are a primary product of our work. The interpretive framework enabled us to make sense of the school and district contexts in which participants were developing and refining their practices. Given the complexity of the school and district settings in which design-based implementation studies are conducted, ongoing interpretations made while conducting a study are necessarily highly selective and reflect assumptions about which aspects of the settings in which participants work are important. As we illustrated in the case of the interpretive framework, these

ongoing interpretations of how improvement strategies are playing out ground decisions about whether to continue with particular strategies, modify them, or abandon them. In our view, the development, explication, and refinement of interpretive frameworks for assessing intended improvement strategies and making sense of their implementation should be central to DBIR.

As our work with the partner districts illustrates, capacity to conduct DBIR is not limited to the development of research tools but also includes tools and routines for communicating with school and district partners. Our key communication tools were the *District Design Documents* in which we shared our understanding of the districts' intended improvement strategies with district leaders and the *District Feedback and Recommendations Reports*. Our key routine was to conduct meetings with the leaders of each district each year to discuss our findings and recommendations. It is worth noting that our primary point of reference when creating the documents and reports each year and when conducting the meetings were districts' improvement goals and the strategies they were attempting to implement to achieve those goals. We contend that this is an appropriate point of reference for DBIR more generally as it places the practitioners' ongoing instructional improvement efforts at the center of partnership work.

In addition to tools and routines, capacity to conduct DBIR includes the skills and dispositions of the research team. Based on the MIST experience, we consider it essential to build interdisciplinary teams. We noted early in this chapter that work of the type that we have illustrated falls at the intersection of research on teaching and learning and research on educational policy and leadership. It was therefore important that the MIST team included members with backgrounds in mathematics education, the learning

sciences, teacher education, educational policy, and educational leadership. Methodologically, we have also found diversity to be a major asset as team members have expertise in a range of methods including ethnography, case study research, design research, social network theory, and quantitative data analysis. In addition, diversity in research interests has proven valuable, with different team members focusing on particular components of our theory of action. To be productive, this diversity requires that team members' work is oriented by a commitment to a common set of problems that they see the value of and are willing to bridge across research specializations. One index of an effective research team is the extent to which individual members become brokers between different research specializations.

With regard to dispositions, DBIR requires that researchers develop new ways of working with practitioners that prioritize the development of trust, take schools' and districts' current improvement goals and strategies as a primary point of reference, and are sensitive to schools' and districts' capacities and constraints. In this respect, DBIR differs significantly from more conventional researcher-practitioner relationships in which researchers work in schools to trial and perhaps refine innovations that they have developed as next steps in their research programs independently of the concerns of practitioners. DBIR requires that researchers' work with practitioners is at the service of practitioners' instructional improvement efforts. Because this was a cardinal principle for MIST, the four-year collaboration with the partner districts proved to be a transformative experience for the MIST lead researchers. In the course of the collaboration with district colleagues, they came to identify with the types of problems that their practitioner colleagues were addressing and came to question the significance of many of the issues

constituted as important in the internal conversations of particular research communities. From this new perspective, there is no "there" there in much educational research because it lacks a grounding in the reality of schooling and the practices of instructional improvement.

A final and often overlooked aspect of capacity to conduct DBIR is effective research team management. Such management involves developing and maintaining relationships with practitioner colleagues, in part by ensuring that that research demands do not overly burden often stressed schools and districts. In addition, it involves balancing the pragmatic agenda of contributing to practitioners' improvement efforts with the research agenda by creating internal team structures, tools, and routines (e.g., interview and school summary forms) that both streamline work and enable team members to see how their contributions contribute to the larger enterprise. In this regard, it is important to establish mechanisms that support team members in developing an overview of research analyses being conducted across the team and in ensuring that potentially related analyses mutually inform each other. Finally, management involves fostering a team culture that takes practitioners' concerns seriously and that provides learning opportunities for graduate students to develop the types of skills and dispositions we have described. Reading about, discussing, and observing others' efforts to conduct DBIR can result in the ability to commentate competently on DBIR (Fish, 1989). However, the development of the capability to conduct design-based implementation studies and the disposition to want to do so requires co-participating in the practice of conducting such studies. A primary management aim is to organize graduate students' participation so that it has the characteristics of research apprenticeship to the greatest

extent possible. The development of a cadre of beginning researchers who can and want to engage in this type of work is critical if DBIR is to become more than a marginal methodology.

References

Argyris, C., & Schön, D. (1974). *Theory of practice*. San Francisco: Jossey-Bass.

Argyris, C., & Schön, D. (1978). *Organizational learning: A theory of action perspective*. Reading, MA: Addison Wesley.

Ball, D. L., Sleep, L., Boerst, T., & Bass, H. (2009). Combining the development of practice and the practice of development in teacher education. *Elementary School Journal, 109*, 458-476.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher, 33*(8), 3-15.

Boston, M. D. (in press). Assessing the quality of mathematics instruction. *Elementary School Journal, 113*.

Brown, J. S., & Duguid, P. (1991). Organizational learning and communities-of-practice: Towards a unified view of working, learning, and innovation. *Organizational Science, 2*, 40-57.

Bryk, A. S. (2009). Support a science of performance improvement. *Phi Delta Kappan, 90*(8), 597-600.

Cobb, P., & Jackson, K. (2011). Towards an empirically grounded theory of action for improving the quality of mathematics teaching at scale. *Mathematics Teacher Education and Development, 13*(1), 6-33.

Cobb, P., & Jackson, K. (in press). Analyzing educational policies: A learning design perspective. *The Journal of the Learning Sciences*.

Cobb, P., & Smith, T. (2008). The challenge of scale: Designing schools and districts as learning organizations for instructional improvement in mathematics. In T. Wood, B. Jaworski, K. Krainer, P. Sullivan & D. Tirosh (Eds.), *International handbook of mathematics teacher education* (Vol. 3, pp. 231-254). Rotterdam, Netherlands: Sense.

Coburn, C. E., & Russell, J. L. (2008). District policy and teachers' social networks. *Educational Evaluation and Policy Analysis, 30*, 203-235.

Cohen, D. K. (2011). *Teaching and its predicaments*. Cambridge, MA: Harvard University Press.

Cohen, D. K., & Hill, H. C. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record, 102*, 294-343.

Darling-Hammond, L., Wei, R. C., & Orphanos, S. (2009). Professional learning in the learning profession: A status report on teacher development in the United States and abroad: National Staff Development Council.

Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly, 48*(1), 1-19.

Fish, S. (1989). *Doing what comes naturally*. Durham, NC: Duke University Press.

Franke, M. L., Kazemi, E., & Battey, D. (2007). Mathematics teaching and classroom practice. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 225-256). Greenwich, CT: Information Age Publishers.

Gibbons, L., Garrison, A., & Cobb, P. (2011, April). *Factors that influence coach centrality*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Grossman, P., & McDonald, M. (2008). Back to the future: Directions for research in teaching and teacher education. *American Educational Research Journal, 45*(1), 184-205.

Hiebert, J. S., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 1, pp. 271-405). Greenwich, CT: Information Age Publishing.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal, 105*(1), 11-30.

Honig, M. I. (2008). District central offices as learning organizations: How sociocultural and organizational learning theories elaborate district central office administrators' participation in teaching and learning improvement efforts. *American Journal of Education, 114*, 627-664.

Horn, I. S., & Little, J. W. (2010). Attending to problems of practice: Routines and resources for professional learning in teachers' workplace interactions. *American Educational Research Journal, 47*(1), 181-217.

Jackson, K. (2011, April). *Exploring relationships between mathematics teachers' views of students' mathematical capabilities, visions of instruction, and instructional practices*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Kazemi, E., Franke, M. L., & Lampert, M. (2009). *Developing pedagogies in teacher education to support novice teachers' ability to enact ambitious instruction.* Paper presented at the Annual Meeting of the Mathematics Education Research Group of Australasia, Wellington, New Zealand.

Kelly, A. E. (2004). Design research in education: Yes, but is it methodological? *Journal of the Learning Sciences, 13*, 115-128.

Lampert, M. (2001). *Teaching problems and the problems of teaching*. New Haven, CT: Yale University Press.

Lampert, M. (2010). Learning teaching in, from, and for practice: What do we mean? *Journal of Teacher Education, 61*(1-2), 21-34.

Lampert, M., & Graziani, F. (2009). Instructional activities as a tool for teachers' and teacher educators' learning. *The Elementary School Journal, 109*(5), 491-509.

Lave, J. (1993). The practice of learning. In S. Chaiklin & J. Lave (Eds.), *Understanding practice: Perspectives on activity and context* (pp. 3-32). Cambridge, UK: Cambridge University Press.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. London: Cambridge University Press.

Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal, 41*, 635-679.

Meira, L. (1998). Making sense of instructional devices: The emergence of transparency in mathematical activity. *Journal for Research in Mathematics Education, 29*, 121-142.

Munter, C. (2009, September). *Defining visions of high-quality mathematics instruction.* Paper presented at the the annual meeting of the North American Group for the Psychology of Mathematics Education, Atlanta, GA.

Nelson, B. S., & Sassi, A. (2005). *The effective principal: Instructional leadership for high-quality learning*. New York: Teachers College Press.

Pea, R. D. (1993). Practices of distributed intelligence and designs for education. In G. Salomon (Ed.), *Distributed cognitions* (pp. 47-87). New York: Cambridge University Press.

Penuel, W. R., Fishman, B. J., Cheng, B. H., & Sabelli, N. (2011). Organizing research and development at the intersection of learning, implementation, and design. *Educational Researcher, 40*(7), 331-337.

Remillard, J. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research, 75*, 211-246.

Resnick, L. (2010). Nested learning systems for the thinking curriculum. *Educational Researcher, 39*(3), 183-197.

Resnick, L., & Glennan, J., T. K. (2002). Leadership for learning: A theory of action for urban school districts. In A. M. Hightower, M. S. Knapp, J. A. Marsh & M. W. McLaughlin (Eds.), *School districts and instructional renewal: Critical issues in educational leadership* (pp. 160-172). New York: Teachers College Press.

Robinson, V. M. J., Lloyd, C. A., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly, 44*, 635-674.

Rogoff, B. (1994). Developing understanding of the idea of communities of learners. *Mind, Culture, and Activity, 1*, 209-229.

Rogoff, B. (1997). Evaluating development in the process of participation: Theory, methods, and practice building on each other. In E. Amsel & A. Renninger (Eds.), *Change and development: Issues of theory, application, and method* (pp. 265-285). Hillsdale, NJ: Erlbaum.

Sfard, A. (2008). *Thinking as communicating: Human development, the growth of discourses, and mathematizing*. New York: Cambridge University Press.

Sherer, J. Z., & Spillane, J. P. (2011). Constancy and change in work practice in schools: The role of organizational routines. *Teachers College Record, 113*(3), 611-657.

Smaling, A. (1992). Varieties of methodological intersubjectivity - the relations with qualitative and quantitative research, and with objectivity. *Quality and Quantity, 26*, 169-180.

Smith, T., Schmidt, R., Berebitsky, D., Garrison, A., Larbi-Cherif, A., & Cobb, P. (2012, April). *Relationships between school and district supports for adopting an inquiry-oriented curriculum and change in the quality of teaching*. Paper presented at the annual meeting of the American Educational Research Association, Vancouver, Canada.

Spillane, J. P. (1999). External reform initiatives and teachers' efforts to reconstruct their practice: The mediating role of teachers' zones of enactment. *Curriculum Studies, 31*, 143-175.

Spillane, J. P., & Thompson, C. L. (1997). Reconstructing conceptions of local capacity: The local education agency's capacity for ambitious instructional reform. *Educational Evaluation and Policy Analysis, 19*, 185-203.

Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. *Mathematical Thinking and Learning, 10*(4), 313-340.

Stein, M. K., & Nelson, B. S. (2003). Leadership content knowledge. *Educational Evaluation and Policy Analysis, 25*(4), 423-448.

Stein, M. K., Smith, M. S., Henningsen, M. A., & Silver, E. A. (2000). *Implementing standards-based mathematics instruction: A casebook for professional development*. New York: Teachers College Press.

Stein, M. K., & Spillane, J. P. (2005). Research on teaching and research on educational administration: Building a bridge. In B. Firestone & C. Riehl (Eds.), *Developing an agenda for research on educational leadership* (pp. 28-45). Thousand Oaks, CA: Sage Publications.

Stephan, M., Bowers, J., & Cobb, P. (Eds.). (2003). *Supporting students' development of measuring conceptions: Analyzing students' learning in social context. Journal for Research in Mathematics Education Monograph No. 12* Reston, VA: National Council of Teachers of Mathematics.

Stigler, J., & Hiebert, J. I. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: Free Press.

Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Figure 1. Relationship Between the Pragmatic and Research-Oriented Levels of Analysis.

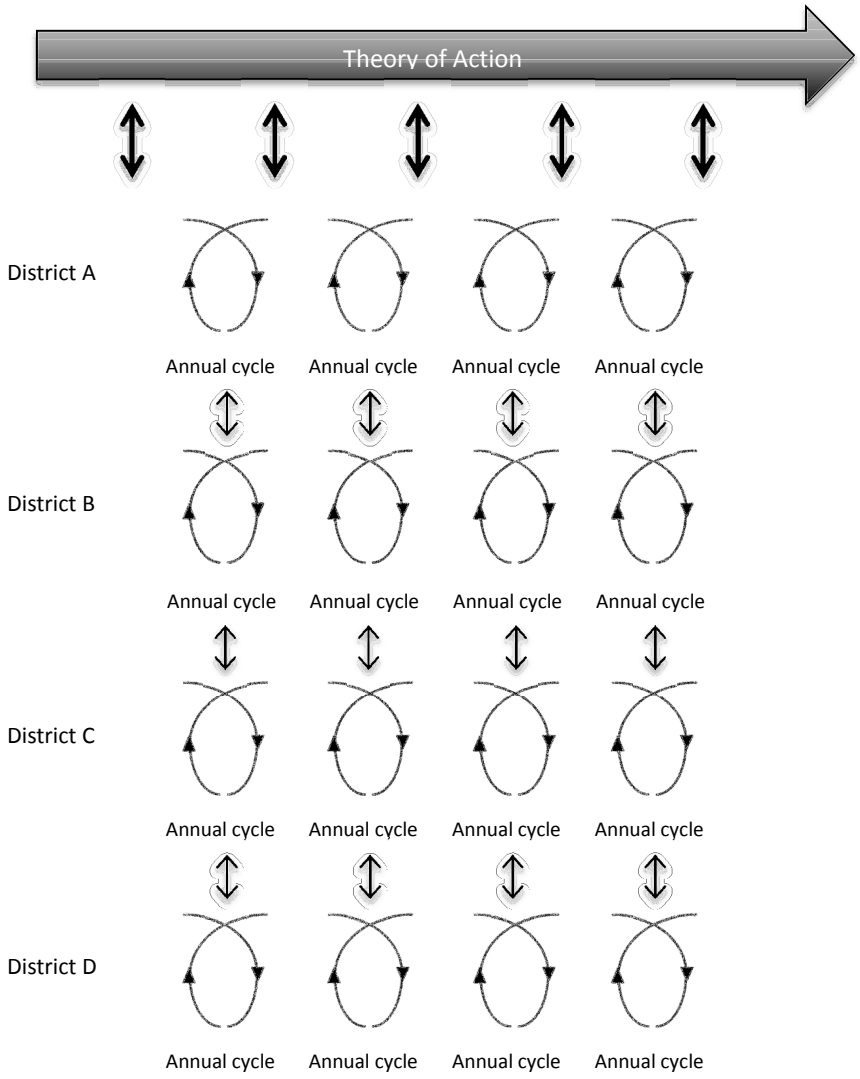

Table 1. Annual Cycle of Data Collection, Analysis and Feedback.

Timeline	Activity	Research Tools Used
October	Interview key district leaders to document strategies for instructional improvement	
October - December	Analyze interviews to create *District Design Document (DDD)* Share DDD with key district leaders and conduct member-checks Create in-house version of DDD	Interpretive Framework, current iteration of Theory of Action
January	Interview teachers, coaches, instructional leaders, and district leaders to document the implementation of the strategies	
February - April	Analyze interviews Create *District Feedback and Recommendations Report* (DFRR)	Interpretive Framework, current iteration of Theory of Action
May	Share DFRR with key district leaders Meet with key district leaders to discuss DFRR	