American Educational Research Journal
February 2020, Vol. 57, No. 1, pp. 148–187
DOI: 10.3102/0002831219854050
Article reuse guidelines: sagepub.com/journals-permissions

Learning From Adaptation to Support Instructional Improvement at Scale: Understanding Coach Adaptation in the TN Mathematics Coaching Project

Jennifer Lin Russell
Richard Correnti
Mary Kay Stein
Victoria Bill
Maggie Hannan
University of Pittsburgh
Nathaniel Schwartz
Tennessee Department of Education
Laura Neergaard Booker
Vanderbilt University
Nicole Roberts Pratt
Instruction Partners (Formerly Tennessee Department of Education)
Chris Matthis
University of Pittsburgh

Attempts to scale up instructional interventions confront implementation challenges that mitigate their ultimate impact on teaching and learning. In this article, we argue that learning about adaptation during the design and implementation phases of reform is critical to the development of interventions that can be implemented with integrity at scale. Through analysis of data generated during a mathematics instructional coaching initiative, we examine the adaptations coaches made to diverse relational and organizational contexts. Findings from two studies of adaptation illustrate the need to attend to the extent to which adaptations are consistent with the core features of a reform. Based on our findings, we posit a generalizable model that supports evidence-based mutual adaptation.

Keywords: adaptation, implementation, instructional improvement, instructional coaching, scale

Learning From Adaptation

The history of education reform efforts is rife with examples of instructional innovations that falter when attempts are made to implement

---

Jennifer Lin Russell holds a joint appointment at the University of Pittsburgh as associate professor in the School of Education and research scientist at the Learning Research and Development Center, 3939 O'Hara St. #808, Pittsburgh, PA 15260; e-mail: *jrussel@pitt.edu*. Her research examines policy and other educational improvement initiatives through an organizational perspective. Her current work examines the way networks and other forms of research-practice partnerships are organized to accelerate systemic improvements that help educators address persistent problems of practice.

Richard Correnti is an associate professor and research scientist at the University of Pittsburgh. His research interests center on the measurement and determinants of teaching and how to improve teaching practice at scale. His current projects focus on analyzing teaching development in the context of interventions such as literacy and mathematics coaching.

Mary Kay Stein holds a joint appointment at the University of Pittsburgh as professor of learning sciences and policy and senior scientist at the Learning Research and Development Center. Her research focuses on classroom-based mathematics teaching and the ways in which policy and organizational conditions shape teachers' practice. Her current work examines the role that cognitive mechanisms play in supporting teacher learning of a set of practices associated with productive discussions in mathematics and literacy classrooms.

Victoria Bill is a senior fellow at the Institute for Learning at the University of Pittsburgh. She is a designer and facilitator of research-based professional development, with a specialization in mathematics coaching, curriculum, assessment, and intervention, as well as a background in school leadership and improvement science.

Maggie Hannan is the associate director for K–12 in Carnegie Mellon's Simon Initiative. She earned her doctorate at the University of Pittsburgh's Learning Research and Development Center, specializing in policy implementation, qualitative methods, and systems theory. Her research focuses on scaling educational change, large-scale collaborative improvement efforts, and fostering equity in the design and implementation of educational technologies.

Nathaniel Schwartz has spent the past 7 years at the Tennessee Department of Education, first as director of research and policy and then as chief research and strategy officer. During his time at the department, he has contributed to the launch of *Tennessee Succeeds*, a strategic plan and vision aimed at increasing postsecondary and career readiness for Tennessee's 1 million students, and to the creation of the Tennessee Education Research Alliance, a state-level research partnership with Vanderbilt University. He has written several articles about state research partnerships and has testified on opportunities to strengthen education research before a subcommittee of the U.S. House of Representatives. He taught high school physics in Illinois and Arkansas and earned his PhD and MPP from the University of Michigan.

Laura Neergaard Booker was formerly executive director of research at the Tennessee Department of Education, where she led the research team and served as a broker for several research-practice partnerships. She serves on the steering committee of the Tennessee Education Research Alliance and cofounded the education research podcast *The TNth Period*. She earned her doctorate from the Department of Leadership, Policy, and Organizations at Vanderbilt University and has recently returned to the same department to serve as a senior lecturer.

Nicole Roberts Pratt is the director of organizational effectiveness at Instruction Partners (formerly the Tennessee Department of Education). She attended mostly rural schools and has a passion for supporting students and educators in rural communities. She started her career in education as a fourth-grade teacher in rural North Carolina. After earning her master's degree in education policy from Vanderbilt University, Nicole spent nearly six years at the Tennessee Department of Education. During her time there, she worked in curriculum and instruction, supporting teachers and leaders in the transition to more rigorous standards and assessments. Nicole also worked closely with the University of Pittsburgh and the Institute for Learning partnership exploring the impact of mathematics instructional coaching and has led the development of a statewide instructional coaching network. Nicole is a practitioner interested in the intersection of policy and implementation to support great teaching and learning for students.

Chris Matthis is a research program manager. He holds a bachelor's degree in biology as well as a master's degree in secondary education from Duquesne University. Chris has contributed to projects related to student motivation in the pursuit of STEM careers, the effectiveness of math coaching, and the formation of networked improvement communities.

them at scale. As instructional innovations are taken from the specific contexts where they were developed, they confront variable systemic capacity to support implementation with integrity (LeMahieu, 2011; Penuel, Fishman, Cheng, & Sabelli, 2011). Our work aims to contribute to research and practice related to the improvement of mathematics instruction at scale through instructional coaching. Instructional coaching is a promising approach to supporting teacher learning and instructional improvement. Well-designed coaching programs have demonstrated significant effects on teaching and even student learning (Kraft, Blazar, & Hogan, 2018).

However, attempts to spread and scale coaching programs have resulted in variable outcomes, in part because coaching is highly context dependent. Education reform efforts have promoted rapid growth in coaching positions in educational systems (Domina, Lewis, Agarwal, & Hanselman, 2015; Neufeld & Roper, 2003). For example, Desimone and Pak (2016) report that the Every Student Succeeds Act (ESSA) includes 11 references encouraging states and districts to utilize coaches to support improvement. As carriers of reform efforts, coaches operate as boundary spanners between federal, state, district, and school contexts, and their roles are often fluid and heavily dependent on contextual factors (Coburn & Russell, 2008; Domina et al., 2015; Kraft et al., 2018). Consequently, effective spread and scale of coaching interventions requires planning explicitly for the adaptation of a coaching approach to diverse local instructional systems.

We engaged in a collaborative project to develop, test, refine, and scale a model for mathematics instructional coaching: the TN + IFL Math Coaching Model.[1] The model provides (1) a framework for evidence-based coaching practices that support teaching improvement, (2) an approach to train coaches to enact the coaching framework, and (3) guidance for schools, districts, and states on how to organize and support a coaching program. The model is designed to be a resource for schools and districts as they support the transition to teaching that is aligned with rigorous, college- and career-ready mathematics standards. As such, the coaching model is a tool to support mathematics instructional improvement at scale. By "scale," in this context, we mean the use of our coaching model to support the uptake of standards-aligned mathematics instructional practices in classrooms throughout a state. In this sense, we evoke the notions of breadth and depth of implementation implicit in the scaling reform concept (Coburn, 2003).

The goal of this article is to describe what and how we learned about adaptation that supports a coaching model that can be deployed at scale. Specifically, we highlight what we have learned about coaches' adaptation of the model to diverse relational and organizational contexts, which represent two key points of variation that coaches confront in their work. Adaptation is evident in the adjustments to the coaching practices that our coaches undertook as they worked with different teachers. By studying these adaptations, we learned about the relative importance of the coaching practices that define our model. Additionally, we present exploratory findings from our more explicit attempt to promote coach experimentation with adaptive integration utilizing practitioner-led methods of improvement research. By adaptive integration, we mean the use of systematic methods to learn how to implement a new practice or work process in a new context, one other than where it was developed (Hannan, Russell, Takahashi, & Park, 2015; Tichnor-Wagner et al., 2018). In both studies, we aimed to learn about productive adaptation and inform the spread of the model to additional coaches and contexts. Ultimately, we argue that explicit attention to adaptation in design projects is critical to the production of scalable interventions, and we offer the insights we learned about the adaptation process, including positing a generalizable model that supports evidence-based mutual adaptation.

Conceptual Grounding

Coaching to Support High-Quality Mathematics Teaching and Learning

Implementing college and career readiness–focused standards, such as those promoted by the Common Core State Standards, requires significant changes in typical teaching practice. For example, in mathematics, students are expected to not only develop fluency in the use of mathematical procedures but also develop a deep conceptual understanding and the capacity to think, reason, and problem solve as they grapple with nonroutine problems. The kinds of environments that support such learning are challenging to set up and, not surprisingly, somewhat atypical in U.S. classrooms (Hiebert & Stigler, 2004). Research tells us that changing instructional practice to make it more rigorous and conceptually demanding is not a trivial undertaking (Correnti & Martinez, 2012; Stein, Correnti, Moore, Russell, & Kelly, 2017; Thompson & Zeuli, 1999). Standard forms of professional development, such as workshops, are generally not sufficient to support this kind of instructional improvement (Birman, Desimone, Porter, & Garet, 2000; Kennedy, 2016; Sandholtz, 2002; Yoon, Garet, Birman, & Jacobson, 2007). As a result, schools and districts have tried to orchestrate more intensive, job-embedded modes of professional development, such as coaching (Desimone & Pak, 2016).

There is an emerging evidence base to suggest that instructional coaching is a promising intervention to support teaching improvement and student learning (Allen, Pianta, Gregory, Mikami, & Lun, 2011; Biancarosa, Bryk, & Dexter, 2010; Blazar & Kraft, 2015; Bryk, Gomez, Grunow, & LeMahieu, 2015; Campbell & Malkus, 2011; Foster & Noyce, 2004; Garet et al., 2011; Killion, 2012; Mangin & Dunsmore, 2015; Matsumura, Garnier, & Resnick, 2010; Matsumura, Garnier, & Spybrook, 2012, 2013; Neufeld &

Roper, 2003; Neuman & Cunningham, 2009; Polly, 2012; Powell, Diamond, Burchinal, & Koehler, 2010; Sailors & Price, 2010). For example, a 3-year randomized control study conducted by Campbell and Malkus (2011) found that coaches positively influenced elementary student mathematics achievement, particularly after coaches had gained experience and skill through extensive professional development. However, studies of mathematics coaching have not always found significant effects on teaching or student learning (see, e.g., Garet et al., 2011). The mixed evidence for coaching can be partly attributed to the complexity of this professional practice and to the fact that coaching roles are structured and enacted in variable ways (Coburn & Woulfin, 2012; Mangin, 2009; Mangin & Dunsmore, 2015). This research suggests that there are important implementation challenges associated with the introduction of coaching programs in schools and districts. Learning from efforts to implement coaching through adaptive integration will help the field understand how and why different coaching models produce different results.

The TN + IFL Math Coaching Model

Prior research on the promise and implementation challenges of coaching informed our work to design a model that could support its effective implementation in diverse local contexts. Our high-level representation of the model is presented in Figure 1. At the center is our coaching practice framework, which is composed of three components: (1) a coach development framework that specifies our method for training coaches, (2) a coaching framework that specifies key coaching practices and routines, and (3) an ethos of continuous improvement that informs how coaches are trained to use disciplined inquiry cycles to adaptively integrate the coaching model into their diverse local contexts. Other key components of the framework include our stance toward mathematics teaching and guidance regarding how schools, districts, and the state can support the implementation of the model. Here, we further elaborate on the coaching framework depicted at the center, because it is critical to understanding the adaptation process that represents the core feature of the model.

Coaching Framework

Given the need to provide a sharper vision of coaching practice as noted in the research literature (Gibbons & Cobb, 2016), we developed a framework to establish a professional vision for what coaching practice should look like. Drawing on the practice-based experience of our Institute for Learning (IFL) colleagues and the research-based insights about mathematics teaching and learning, our team identified three key coaching practices: (1) deep and specific discussions of the instructional triangle, (2) establishing mathematics and pedagogical goals, and (3) evidence-based feedback.

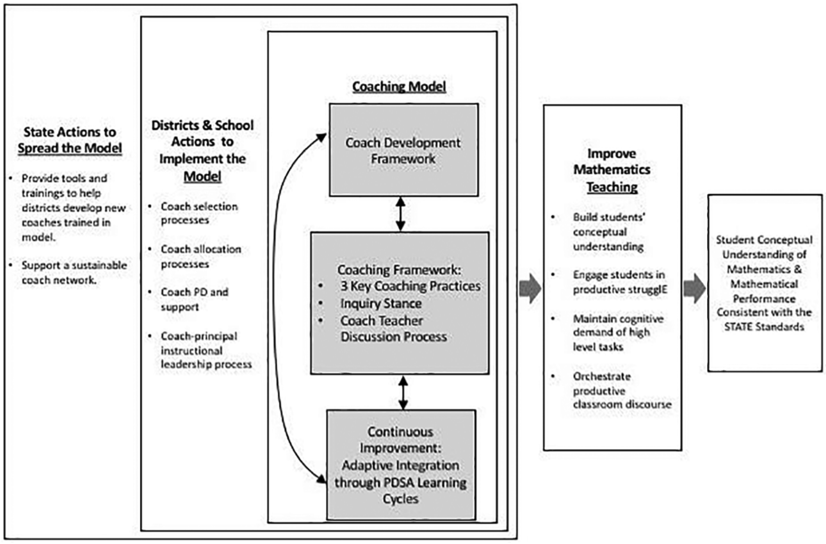

Figure 1. **TN + IFL Math Coaching Model.**
Note. PD = professional development; PDSA = plan-do-study-act.

These practices are rooted in key tenets from the professional learning literature, such as the need for coaching interactions to be content specific (i.e., focused on *mathematics* teaching rather than general teaching strategies) (Desimone, Porter, Garet, Yoon, & Birman, 2002; Matsumura et al., 2012) and the power of substantive talk about content, student thinking, and pedagogy in supporting teachers use of high-quality teaching practices (Coburn & Russell, 2008; Lewis, Perry, & Hurd, 2009; Munter & Correnti, 2017). Finally, guidance on feedback was rooted in research suggesting that coaches often do not take an analytic stance with respect to their colleagues' instruction (Murray, Ma, & Mazur, 2009).

To facilitate the uptake of these practices, we drew on research on organizational routines, which emphasized how routines structure the interdependent work of diverse practitioners (Feldman & Pentland, 2003; Sherer & Spillane, 2011; Stelitano, Russell, & Bray, 2017). We embedded the key coaching practices in a designed routine—the Coach-Teacher Discussion Process (see Figure 2). The process begins with the coach and teacher selecting a high-cognitive-demand mathematics task and then familiarizing themselves with the task independently, including identifying the multiple ways in which students would solve the task. The subsequent steps of the process consist of a preobservation planning conference, a lesson observation, and a postobservation feedback conference. In order to gain competence in

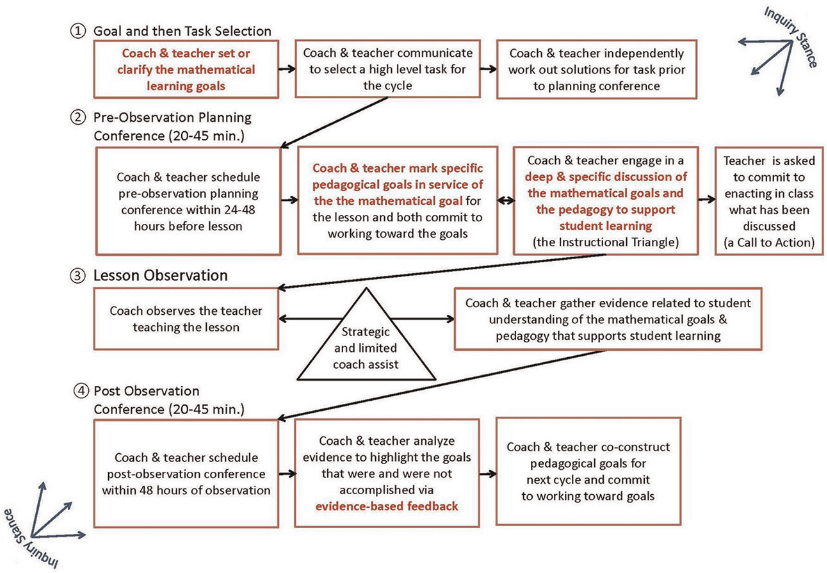

Figure 2. **Coach-Teacher Discussion Process.**

using the routine and to generate data for our research, we asked the coaches to engage in the Coach-Teacher Discussion Process at least three times in Year 1 and two times in Year 2 of the project with two partner teachers (who had agreed to participate in our research).

Finally, the coaching framework defines the *inquiry stance* as the way coaches should interact with the teacher. By specifying that coaches take an inquiry stance with teachers, we highlighted an interactional style that emphasized the coaches relaying what they had "noticed" about classroom practice and raising questions or "wonderings" instead of making more explicit recommendations for what the teachers should do while teaching. The inquiry stance stems from research on professional learning that points to the need for active teacher participation in meaning making around shifts in practice (Opfer & Pedder, 2011).

Coach Development Framework

Prior research and our practice-based partners' experience both suggested that many instructional coaches get limited training in coaching practice (Gallucci, Van Lare, Yoon, & Boatright, 2010). Consequently, a key design task was specifying a coach development framework that included a curriculum and related tools for training coaches to enact the coaching framework. In the first 2 years of the project, the coaches engaged in six face-to-face convenings, each lasting 2 days. The research team presented emerging findings from analyses of data that the coaches generated during action periods between meetings and engaged the coaches in conversations that aimed to make sense of the data-based patterns. For example, our analysis of recorded prelesson planning conferences in an early cycle suggested that only a few coaches were pressing the teachers to articulate specific mathematics content that they wanted students to learn through the lesson they were planning; most of the coaches and teachers were simply identifying broad topics, such as "equivalent fractions." These analyses provided an opportunity for discussion of robust learning goals that help teachers prepare for lessons that support students in developing a conceptual understanding of mathematical ideas. The data analyses informed the design of learning sessions that dove deeper into the key coaching practices in the model, drawing on artifacts from the coaches in the network. The coaches also participated in a series of monthly webinars between the face-to-face meetings.

Continuous Improvement: Learning About Adaptation Through Inquiry Cycles

A final key component of our coaching model is a commitment to continuous improvement. We refined the model through six improvement cycles, each launched by a face-to-face network meeting. During the improvement cycles, the coaches generated and shared data that documented their coaching practice. These data included video-recorded coaching conversations and related artifacts such as instructional tasks and student work. The analysis of each cycle's data contributed to refinement of the coaching model and to the training all coaches received in subsequent network meetings. Additionally, we trained the coaches to utilize an inquiry routine—the PDSA (plan-do-study-act) cycle—to adaptively integrate the model into their local context. Both the research-based adaptations and the PDSA cycles will be discussed further in the section on learning from adaptation.

Evidence of Efficacy of the Model

As we strove for a scalable coaching model, a key consideration was generating evidence that the model had the promise of efficacy that would make it worthy of scaling. In other words, we sought evidence that training coaches in the model contributed to improved coaching and improved teaching practice. In another article, we had examined our first 2 years of data that demonstrated improvements in coaching (effect size [ES] = .71) as well as improvements in teaching (ES = .61) (Russell et al., 2016). Longitudinal models for growth in coaching revealed no significant differences in growth rates among the coaches, suggesting that all the coaches improved the depth and specificity of the preconference discussions they had with their partner teachers at roughly the same rate. Meanwhile, longitudinal growth models for 429 video observations from 103 partner teachers found an average ES of .61 on a per-year basis for growth in teaching aligned with the coaching.[2] Furthermore, teachers with the lowest initial scores on our teaching measures grew the most. These findings provide promising evidence that the coaching model has the potential to contribute to substantive practice improvement. A follow up quasi-experimental study, currently under way, is exploring whether there are significant differences in coaching, teaching, and student learning between coaches and teachers trained in our model and those who engaged in garden-variety coaching.

Learning From Adaptation to Guide Effective Coaching at Scale

In the prior sections, we discussed how research on coaching informed the specification of our model for mathematics instructional coaching. In this section we turn our focus to the central topic of this article: how our project approached studying and understanding the adaptation of the model as it was implemented by a network of coaches in diverse local contexts.

Prior Research About Adaptation

While knowing more about effective interactions between coaches and teachers can contribute to improved coaching programs, improving education at scale is a more complex process than simply identifying optimal practices through research and directing practitioners to implement them (Coburn, 2003; Elmore, 1996). Policy and implementation research has consistently found that when complex interventions are implemented, *mutual adaptation* occurs as educators change their work practices to align with the intervention and make adjustments to make the intervention useable in their context (Berman & McLaughlin, 1978; Coburn, 2004; Cohen et al., 2008; Honig, 2006; Supovitz, 2008). Successful scaling requires that educators make reasoned adjustments to interventions as they become embedded in existing systems (LeMahieu, 2011; McDonald, Klein, & Riordan, 2009; Weinbaum & Supovitz, 2010). However, mutual adaptation tends to be the exception, not the norm, when educational reforms seek significant changes in practice (Reiser et al., 2000). Local adaptation tends to be one-sided rather than mutual in that the intervention goes through more adjustments than the individuals responsible for implementing it (Spillane, 1999). In a notable exception, Siskin (2016) draws on a study of the spread of the International Baccalaureate model to nontraditional school contexts—those serving low-income students of color in an urban area—pointing to the potential for mutual adaptation in action, where feedback from the researchers, context, and coaches enabled the reformers to monitor implementation and adaptation, learn from them, and incorporate productive adaptations into the designs (McDonald et al., 2009; Peurach, 2011).

In the case of our coaching intervention, we expected two main types of adaptation: adaptation to *diverse relational contexts* and adaptation to *diverse organizational contexts*. Prior research on coaching has pointed to potential relational challenges associated with coaching programs (Ippolito, 2010; Neufeld & Roper, 2003). Some teachers respond to coaching with resistance, perceiving coaching as evaluative or an indictment of their teaching practice (Bean & Carroll, 2006; Mangin & Stoelinga, 2011). As a result, coaches often avoid giving the critical, substantive feedback to teachers that challenges them to change their practice (Lord, Cress, & Miller, 2003). Given these relational challenges, we reasoned that coaches would adapt their work with teachers based on their impression of the teacher's openness to coaching. Prior research suggests that coaches vary in how they position themselves in relationship to teachers, with two primary distinctions: responsive versus directive (Deussen, Coskie, Robinson, & Autio, 2007; Dozier, 2006; Ippolito, 2010). While different coaching models advocate different stances, few studies have investigated these claims, and when they have, the studies have not directly investigated the efficacy of these stances (Ippolito, 2010).

Prior coaching studies also point to features of organizational context that shape the implementation of coaching programs. For example, the selection and training of coaches influence the degree to which they are a source of teaching expertise for colleagues (Coburn & Russell, 2008). People who enter coaching roles may be relatively expert teachers, but they need learning opportunities to develop coaching skills and dispositions, such as how to facilitate conversations with teachers about their instructional practices and how to gather and analyze evidence to monitor changes in teachers' instruction (Gallucci et al., 2010; Neufeld & Roper, 2003). Even when coaches are intentionally selected and trained, they may not be effectively utilized in schools. For example, Bean, Draper, and Hall (2010) found that principals often direct coaches to spend a significant proportion of their time engaged in managerial support activities, instead of reserving time for direct work with teachers. Additionally, principals play a critical role in promoting (or hindering) teacher engagement in coaching. In a large-scale study of literacy coaching, Matsumura, Sartoris, Bickel, and Garnier (2009) found that principals influenced coach access to teachers and classrooms through the degree of their endorsement of coaches as sources of expertise. Finally, school and district leaders shape the enactment of coaching roles in consequential ways by framing the purpose and goals of coaching programs (Mangin & Dunsmore, 2015).

Learning From Adaptation in the TN Math Coaching Project

Our project aimed to learn from coach adaptation of the model in two ways. First, researcher-driven inquiry cycles attended to variation in implementation of the model and sought to identify adaptations that were associated with positive coaching and/or teaching outcomes, which could become part of the model's design. McLaughlin (1987) noted that variability offers opportunities to learn about interventions and their implementation, noting that local responses generate natural experiments that should be exploited by analysts. In the course of conducting our traditional research in this project, we sought to exploit those natural experiments to contribute to our learning about the coaching model's implementation. Second, practitioner-driven inquiry cycles provided a disciplined routine for coaches to engage in explicit adaptive integration. In this section, we describe our coach and teacher samples and then provide a broad overview of our approach to each form of learning from adaptation. In subsequent sections, we provide illustrative examples.

Coach and Teacher Samples

Through a competitive process, including a written application and scenario-based interviews, we selected 32 coaches from 21 districts to participate in the network in Year 1 of the study.[3] We selected coaches based on their possession of the basic conditions that would enable engagement in the type of coaching our model promotes, considering factors such as adequate mathematics content knowledge and working in a role that afforded opportunities to engage in intensive one-on-one coaching. However, we strategically selected coaches to represent a range of initial coaching capacity and experience, and variation in basic contextual conditions such as geographic region, urbanicity, and role (e.g., school-based coaches vs. district coaches). For all of the coaches selected, there was a gap between the type of coaching our model promotes and their current practice, and for most of them, that gap was quite significant. For example, our model emphasizes prelesson planning conferences, a practice many coaches did not employ.

The coaches signed a contract with the TN Department of Education accepting the following responsibilities: to attend three 2-day network meetings per year, participate in monthly webinars, document all coaching activity using a coaching log, complete periodic short surveys, select two teachers for intensive study and commit to collect data about their practice, and commit to engage as partners in the continuous improvement of the model. Each coach selected two partner teachers to participate in the study who taught mathematics in Grades 3 through 8, although one teacher left the study midyear. Consequently, our sample consists of 63 partner teachers.

Researcher-Driven Inquiry Cycles

During each cycle, the coaches implemented the coaching model in their work with two partner teachers; specifically, they completed and documented one cycle of the coach-teacher discussion process with each partner teacher. As these data were shared with the research team, the researchers analyzed key constructs related to the uptake of the model and the expected points of context variation, including aspects of the relational context of coaching, such as teacher responsiveness, and the organizational context, such as coach role construction (e.g., school based vs. district based and the number of teachers the coaches supported at any given time).

The research-driven inquiry cycles culminated in evidence-based discussions with the coaches and state leaders, which enabled sense making about the findings. Subsequent training sessions provided ongoing guidance for implementation based on what we were learning from the implementation. This process is consistent with the findings from the New American Schools implementation studies conducted by RAND, which emphasized ongoing communication and support for implementation among designers and practitioners (Berends, 2000; Berends, Bodilly, & Kirby, 2002). Our discussion of Study 1 illustrates an example of researcher-driven inquiry into adaptation to relational context variation.

Practitioner-Driven Inquiry Cycles

Implementation that embraces mutual adaptation expects local variability and aims to give practitioners the time and support to make good choices about how to implement interventions (McLaughlin, 1976). Our approach to helping coaches make good choices in their adaptation of the model drew inspiration from improvement science, which provides tools for learning how to improve practice at scale by building an evidence base generated from disciplined practitioner inquiry (Bryk et al., 2013). Improvement science has a long history in the manufacturing industries, and subsequently the health care fields, and provides a systematic methodology for learning from practice to improve the systems and processes that shape work within organizations (Berwick, 2008; Gawande, 2007; Langley et al., 2009; Lewis, 2015).

Improvement science speaks to the need for adaptive integration of standard work processes into new contexts, which is critical for learning how to scale improvements. As an innovation or change that has worked in one context moves into others, improvement methods are used to learn what it takes to make it work under diverse conditions (Bryk et al., 2015; Tichnor-Wagner et al., 2018). For example, schools participating in a network of districts aiming to reduce beginning teacher attrition—the Building a Teaching Effectiveness Network—used PDSA inquiry cycles to adapt a standard process for coordinating feedback for beginning teachers within the existing constraints of their school systems. The extent to which schools utilized improvement science methods to address the adaptive integration problem predicted their successful enactment of the feedback process (Hannan et al., 2017). One way to plan for adaptive integration is to involve practitioners in designs for implementation (Reiser et al., 2000;

Tichnor-Wagner et al., 2018). We issued invitations to our coaches to codesign the integration of the coach-teacher discussion process into their local contexts through a process of coach-led inquiry cycles. While resource constraints precluded a comprehensive investigation of the uptake of these inquiry cycles, we documented strategically selected cases to begin to explore this process. Our presentation of our exploratory Study 2 provides an illustrative example of coach-driven inquiry into adaptation to organizational context variation.

Study 1: Researcher-Driven Inquiry Into Adaptation to Relational Context Variation

In this study, we demonstrate how we tracked the ways coaches adapted their practice when working with teachers who they perceived to have varying levels of responsiveness to coaching; this is a case of adaptation to the diverse relational contexts of coaching. Our exploratory analysis of Year 1 data revealed evidence that the coaches were adjusting their practice based on their judgments about the type of teacher they were working with: one judged to follow through versus one less likely to follow through. We found that when the coaches rated teachers as relatively less responsive to coaching at the beginning of the year, those same teachers judged coaches to change the nature of their interactions during coaching over the course of the year. Specifically, the coaches became more likely to provide *more explicit directions about teaching*. In contrast, when coaches rated teachers as relatively more responsive to coaching, the same teachers reported that their coaches tended to increasingly *press them to engage in mathematical reasoning* over time. In other words, coaches' perceptions of teacher responsiveness determined whether they emphasized explicit direction about teacher actions or deep conversations about mathematics.

Data Collection

For this study, we drew on three primary data sources: (1) video footage and artifacts from coaching cycles, (2) log entries completed by the coaches for each interaction with their partner teachers, and (3) periodic surveys sent to the partner teachers and coaches. The coaches and their partner teachers participated in three formal coaching cycles in Year 1 using the coach-teacher discussion process, described in the prior section. To capture intensive data on coaching practice, each enactment of the discussion process was documented by the coaches, who gathered videotapes of prelesson planning conferences, lessons, and postobservation feedback conferences; teacher and coach planning and reflection notes; and artifacts such as the instructional tasks. The coaches shared these multiple data sources with the research team.

To get a broader perspective on coach engagement with each partner teacher, we had the coaches complete a web-based coaching log (Coach Tracker) every time they had an interaction with their partner teachers. These log entries consisted of a series of fixed response items related to the date, duration, and type of interaction (e.g., informal conversation or demonstration lesson). The log also provided an opportunity for the coaches to rate the *teachers' responsiveness* to coaching, specifically their judgment of the likelihood that teachers would follow through on the teaching ideas discussed during a given interaction. The coaches could also use a series of open-ended questions to note goals for teachers and reflect on their coaching practice.

Finally, we administered the discussion process survey to the teachers following each time they participated in the coach-teacher discussion process. This survey included a short set of items that aimed to measure uptake of the key coaching practices in our model, such as coaches having deep discussions of mathematics content, student learning, and pedagogy with teachers.

Analytic Approach

Our analyses for Study 1 explored whether the coaches were adapting their practice in response to their perceptions of teacher responsiveness.

Measures

We explored the items on the discussion process survey in order to develop constructs that fit our conceptual model. One factor we call *coaching explicitness* included three items that if endorsed by the teacher would suggest that the coach recommended a prescribed set of teaching practices when coaching teachers. For example, one item asked teachers the extent to which they agreed with the following statement: "As a result of this cycle, I know a set of prescribed behaviors that will always serve me well in the classroom." A second item was "After this coaching cycle, I could articulate the theory of mathematics teaching and learning my coach has." Finally, a third item asked teachers the extent to which their coaching interaction was characterized as follows: "Most of the time was spent talking about the actions I should engage in during the lesson." Endorsement of these three items suggests that the teacher perceived the coach as being explicit about specific teaching behaviors that should be enacted. Considering our coaching model, the *explicitness* construct is complicated to evaluate. On the one hand, the construct could potentially serve the goal of preserving integrity to the goal of increasing the depth and specificity of discussion on pedagogy (one aspect of the instructional triangle), given its focus on teacher actions. On the other hand, it runs counter to the goal of adopting an inquiry stance with teachers—where *co-construction* toward an understanding of how to improve teaching is ideally realized.

A second construct we call *coaching press*, was composed of two items that measured teachers' perceptions of whether they were "pressed" to discuss mathematics. These items asked the teachers the extent to which they agreed with the following statements: "I was pressed to talk about mathematical reasoning" and "I was pressed to make sense of mathematical ideas". There is no ambiguity in how these items are conceptually aligned to our coaching model's aim to build teachers' capacity to think deeply about math content and students' mathematical thinking. In this sense, we viewed the press construct as an indication that coaches were ensuring deep and specific conversations about the instructional triangle, while simultaneously engaging in co-construction of ideas during the inquiry process.

Teacher responsiveness was assessed by the coaches in each log of an interaction with a partner teacher on the Coach Tracker log. Specifically, an item asked the coaches to rate teacher responsiveness using the following scale:

1. Teacher exhibited behaviors that signaled explicit resistance.
2. No explicit resistance, but teacher was not prepared and/or signaled low engagement.
3. Teacher was compliant but wasn't open to suggestions, and/or I don't expect any follow-through.
4. Teacher exhibited behaviors that signaled engagement, yet I am uncertain about the level of follow-through.
5. Teacher exceeded expectations for engagement, and I have high confidence the teacher will follow through.

To generate scores for the coaches' judgments of teacher responsiveness for Cycle 1, we averaged across all coach logs before the end of Cycle 1. The distribution of coach reports of teacher responsiveness indicated a fair degree of variation: 15% of the coaches reported an average at or below a score of 3, another 10% were between 3 and 4, 34% indicated a responsiveness score of 4, 7% were between 4 and 5, and the final 34% were at a score of 5.

A final key construct in our analyses was a proxy for coaching quality and efficiency at Cycle 1. It was important to consider whether the patterns in the relationship between perceptions of teacher responsiveness and coaching decisions about explicitness and/or press were influenced by the quality of the coaching interaction. Therefore, we constructed a preconference efficiency variable that utilized expert ratings of the preconference interactions between coach-teacher pairs (e.g., preconference rigor items included, deep and specific discussion of math content, pedagogy, and student thinking) divided by the duration of the preconference. We then used these scores to predict both status and linear change on explicitness and press over the course of the year. Efficiency was included as a covariate, as a competing explanatory theory of whether and how the coaches adjusted their relational interactions with teachers. Thus, we added this covariate thinking that the coaches are weighing a lot of factors when making decisions about how to adapt their coaching. One mitigating factor could be the coaches' perceptions of the need for rigorous preconference conversations as well as the need for brevity to use time efficiently. As with classroom teaching, we hypothesized that the greater the perceived need for efficiency, the more likely the coach would exhibit more explicitness in coaching conversations.

Exploratory Survey Analyses

We engaged in exploratory analyses to examine whether the coaches adapted their interactions with teachers in response to the variation in the relational context of coaching as measured by the coach's perceptions of *teacher responsiveness*. We looked for evidence of adaptation by examining whether the teachers reported changes in the coach-teacher interactions. We employed multilevel growth models for both the coaching *explicitness* and the coaching *press* scales as outcomes, nested in three time points over the year, with the time points nested in the coach-teacher pairs (for a full description of the model, see the appendix).[4] We hypothesized that if the coaches adapted their coaching to individual teachers, it might be based on their judgments about how responsive the teacher appeared to be in their initial interactions (i.e., whether they believed that the teacher would follow through with the ideas they discussed together). Specifically, in an initial model, we used the measure of the coaches' initial perceptions of teacher responsiveness to predict changes in the teachers' perceptions of the coaching they received (explicitness and press for math reasoning). We ran a second model, adding another independent covariate measuring the efficiency of the first preconference. We also included several covariates as controls in these models. First, we included each of the items to adjust for any differences in item difficulty within each scale. Next, we included the length of the task, or the amount of time it took for the teacher to implement a given task as measured through videotapes of classroom instruction, as a time-varying covariate.[5] We assumed that this was a proxy for the nature of the task being implemented. We included this because we further assumed that the nature of the task itself could influence partner teacher perceptions of the focus on explicitness or press, with higher-cognitive-demand tasks (as evidenced by the longer time spent on the task) providing more fruitful opportunities for the coaches to press teachers for mathematical reasoning.

Qualitative Case-Based Analyses

To explore whether the statistical phenomenon was evident in interactions between coaches and teachers, we examined a limited number of qualitative cases, analyzing the transcripts of coach-teacher interactions in the preconferences. Cases were selected that most represented the pattern in our survey data. Three explicitness cases included coach-teacher pairs where the coach rated the teacher as relatively less responsive in the first cycle and then the teachers subsequently reported that the coaches provided explicit teaching direction in later cycles. The two press cases included teachers rated as highly responsive by their coaches, where the teachers reported that the coaches increasingly pressed them to engage in math reasoning. We coded evidence of responsiveness in the first-cycle preconferences and then coded for evidence of explicitness and press for mathematical reasoning in the second and third cycles. We present examples from one case representing each pattern. These qualitative cases should be interpreted cautiously, not as evidence that this pattern of adaptation was present in all coach-teacher interactions but rather that we could see what these phenomena looked like in a select set of cases.

Survey Results

We first explored the psychometric properties of our primary measures in a fully unconditional measurement model. Our results indicated that the teacher reports of explicitness (λ_1 = .752) and press (λ_2 = .818) form reliable estimates of between-teacher differences in status[6]—which is a relative measure of the variance between teachers at a given time point compared with the measurement error (σ^2). These hierarchical linear model reliabilities demonstrate that there is variance in how the teachers answered the items on each scale and, further, that the variance is relatively large when compared with the variance between the items on each scale within teachers. Interestingly, the correlations between the slope estimates for explicitness and press indicate a negative correlation ($r = -.60$), suggesting that the two scales are somewhat independent but that for at least some teachers while one scale was increasing, the other scale was decreasing.

Exploratory findings on the press slope. The top half of Table 1 provides evidence that the coaches adapted their coaching based on their early judgments about their partner teachers. Take for example, the findings for teacher responsiveness on the construct *press* in both Models 1 and 2. Here, we find that when the coaches deemed teachers to be more responsive at Cycle 1, the partner teachers they coached reported a higher slope for changes in press (β_{111} = 0.48, standard error [*SE*] = 0.26, p = .074 in Model 1; β_{111} = 0.48, *SE* = 0.27, p = .079 in Model 2). Thus, when the coaches deemed teachers to be more responsive initially, the nature of their interactions with these teachers changed over the course of the year. Teachers reported greater amounts of press over time on the discussion process survey. In Model 2, the variable measuring efficiency of the preconference at Cycle 1 is not significant on the press slope.

Table 1
Effects of Coach Judgments on Partner Teacher Perceptions of Change in Coaching Interactions

	Partner Teacher Report of Press for Mathematical Ideas/Reasoning			
	Coeff.	*SE*	Coeff.	*SE*
At first cycle				
Average (β_{100})	.01	.18	.00	.18
Tch. Resp. (β_{101})[a]	–.06	.27	–.06	.28
PC Effcncy (β_{102})[b]			.09	.52
Growth slope				
Average (β_{110})	–.03	.17	–.06	.17
Tch. Resp. (β_{111})[a]	.48	.27	.48**	.26
PC Effcncy (β_{112})[b]			–.50	.56
Task length (π_{102})	.01	.00	.01*	.00
	Partner Teacher Report of Coach Explicitness			
	Coeff.	*SE*	Coeff.	*SE*
At first cycle				
Average (β_{200})	–.06	.12	–.06	.11
Tch. Resp. (β_{201})[a]	.31*	.16	.31*	.15
PC Effcncy (β_{202})[b]		.31	–.88*	.29
Growth slope				
Average (β_{210})	.05	.16	.06	.15
Tch. Resp. (β_{211})[a]	–.55*	.20	–.54*	.19
PC Effcncy (β_{212})[b]		.51	1.16*	.48
Task length (as implemented)	.01	.00	.01	.00

Note. SE = standard error; Tch. Resp. = teacher responsiveness; PC Effcncy = preconference efficiency.
[a]The coaches' reports of the likelihood that the teacher follows through on the reported interaction in Coach Tracker during all the reported interactions in the first cycle.
[b]Efficiency of the preconference rigor score. The sum of items on the preconference rubric divided by the length of time in preconference. More efficient ratings tended to be short preconferences.
$^{*}p < .05$. $^{**}p < .10$.

Exploratory findings on the explicitness slope. Findings on the explicitness slope provide a nice complement to those on the press slope. The less likely the coaches deemed teachers to follow through with the ideas discussed in their initial interactions, the higher the slope for changes in coach explicitness over the year ($\beta_{211} = -0.53$, $SE = 0.20$, $p = .011$ in Model 1; $\beta_{211} = -0.53$, $SE = 0.19$, $p = .008$ in Model 2).[7] Once again, we see evidence of adaptation based on coach judgments of teacher responsiveness. The pattern of findings for the coach explicitness and press scales suggests that the coaches are adapting their coaching practice in response to their judgments of teacher receptivity to coaching when they begin working with an individual teacher.

Finally, the variable measuring the efficiency of the preconference at Cycle 1 had a similar effect and interpretation, but in the opposite direction (β_{212} = 1.14, *SE* = 0.48, *p* = .019). The more efficient the initial preconference was, the more the partner teacher reported the coaching interactions to increase in explicitness over the course of the year. It is interesting that both lower teacher responsiveness and a higher measure of efficiency predict higher rates of explicitness in coaching over the year. This suggests that each covariate predicts some unique variance in the slope, and it underscores the point that the coaches likely attend to multiple factors as they make decisions about how to adapt their interactions with teachers in their effort to help them improve their instruction.

Qualitative Results

Our qualitative analyses revealed evidence of the survey-based findings. While we did not systematically analyze the full transcript data set, we present evidence in these strategically sampled cases to show what this type of adaptation looks like in practice. In a case that we believe represents our explicitness findings, a coach rated her third-grade partner teacher relatively low on responsiveness during the first cycle: The coach's evaluation of teacher responsiveness for each interaction in the first cycle averaged 3.33, or the equivalent of a 3 on the rating scale, signaling, "The teacher was compliant but wasn't open to suggestions, and/or I don't expect any follow-through." We saw some evidence of this teacher's possible compliance orientation to coaching in the first-cycle preconference: 37% of the teacher's turns in talk were one-word responses such as "Mhhm" or "OK."

Perhaps adapting to her perception that the teacher would not follow through, the coach's moves in the second and third cycles are much more explicit than in the first cycle. Throughout the second and third conferences, the coach tightly controls the flow of the conversation and makes specific recommendations about what the teacher should do when teaching the upcoming lesson. For example, the coach presses the teacher to articulate a pedagogical goal for the lesson by saying, "Alright, so what about you, personally? What are your goals as a teacher for this lesson?" The teacher responds by talking about the students, saying "Well, my goal is for them to be able to obviously answer all four questions, but show me with the manipulatives provided how much of each color yarn you need." The coach then redirects the teacher to think about the goals for her teaching: "Alright, but for your teaching practice?" The teacher replies by articulating a goal: "OK, for me, I've looked over this, and I think that what I need to work on is a lot of times I give struggling students not enough think time, so I need to support productive struggle in learning math." The coach accepts this goal and discusses it with the teacher but then suggests another goal for the teacher: "And another part of that, may be an action plan for you would be to be purposeful in planning your assessing and advancing questions."

Additionally, the coach makes several direct suggestions phrased in a way that suggests that the coach may have been trying to hold the teacher accountable for following through on these pedagogical goals. For example, after a discussion of the second pedagogical goal in the conference referenced above, the coach recaps by saying, "Alright, so advancing questions and then walking away are going to be your action steps tomorrow to let them struggle." This discussion continued to the third cycle conference, when the coach said,

> In the share phase, and we talked about this during your last coaching cycle, we talked about how you wanted to take more of a backseat and let the kids lead the share phase. So, as we went through all those solution paths, what do you think about kind of having a checklist, like I wrote down here all the solution paths, and be intentional about looking for those, so when you get to the share phase, you know who you're calling on, and you're thinking about, "Ok, I want to pick a kid who showed the arrays. I want to pick a kid who showed the area model. I want to pick a kid who simply wrote the equations."

As these passages illustrate, the interactions in this conference reflect a focus on what the teacher should do, prompted by explicit coach suggestions.

In contrast, our second focal case illustrates a coach working with a third grade teacher who she judged as highly responsive to coaching, receiving a rating of 5 on all interactions in the first cycle, which corresponds with "The teacher exceeded expectations for engagement, and I have high confidence she will follow through". In this case, we see a very different pattern of interaction, one with fewer coach moves that suggest explicit actions the teacher should take in the classroom and more focus on a general press for teacher and student mathematical reasoning. In the first-cycle preconference discussion, the positive rapport between coach and teacher was evident in their ease in communication, which included a strong coach orientation toward listening to what the teacher says and a general upbeat and positive interactional style. The coach repeatedly presses the teacher to articulate her thinking about mathematics and reason about students' mathematical thinking. For example, when the coach first asks the teacher about her mathematical goals for students, the teacher says,

> I believe the last time we had done this we had at least one group that really were able to show that they had a conceptual understanding of what we were doing with decimals. . . . I expect to see good representation of fraction models. I expect to see some—and hear some

> really great explanations of thinking. And then also—and this is going back to evaluating others—some really good constructive peer feedback.

In this conference, the teacher is making general statements about the mathematical goals of the lesson. However, the coach does not accept these general statements and goes on to press the teacher throughout the conference to get deeper and more specific. For example, in this passage the coach and teacher co-construct a deeper statement about the mathematical goals:

> *Coach:* It really wrestles with interpreting a fraction as division of the numerator by the denominator because students so many times don't see that. They see three fourths—as a fractional form.
> *Teacher:* Yes.
> *Coach:* And don't understand that represents . . .
> *Teacher:* We are struggling with getting them to conceptualize doing—how to actually divide a line into equal sections. That is . . . honestly, that is a very important skill that they struggle the most on. They can read one with all the lines already there, but getting them to conceptualize how to split it up is very difficult.
> *Coach:* And . . . yeah, it's . . . they don't understand that three fourths means three . . .
> *Teacher:* Over four.
> *Coach:* . . . *divided* by four.

At the end of a conference where the teacher has been repeatedly pressed to articulate the mathematics goals of the lesson and how students will make sense of the mathematics, we later see the teacher taking the lead in this kind of talk.

> *Teacher:* So I think in a lot of ways this is really going to tell us who is still at the elementary level in terms of who do we really need to go back and work on the conceptual understanding of partial amounts. And then I think have a clear distinction as to who has it, who doesn't, and where it is that they're missing . . .
> *Coach:* So the first thing, what's the big mathematical idea here that they have to . . .
> *Teacher:* They have to compare values.
> *Coach:* But to do that, what do they have to do?
> *Teacher:* They're gonna have to come up with like denominators.
> *Coach:* Yeah. Equivalent fractions.

While we cannot be certain about the coach's intentions, we speculate that the coach's perception that the teacher was responsive to coaching, may have contributed to her focus on pressing the teacher to focus on mathematical reasoning, rather than specific pedagogical moves the teacher should make in the classroom.

In sum, these two cases provide some indication of what these survey-based adaptation patterns look like in the interaction between coaches and teachers. In both cases, the coach plays an active role in shaping the conference. In the first case, however, the coach is more directive in terms of what behaviors she expects to see the teacher do during the lesson, while in the second case, the coach exerts her influence in terms of urging the teacher to uncover the mathematical ideas that are undergirding the lesson itself and student thinking about those ideas.

Relationship to Teaching Improvement

We found it interesting that both the explicitness and the press construct showed evidence of coach adaptation. However, while adaptation toward press was viewed theoretically as unambiguously positive, adaptations toward explicitness were more indefinite. Thus, while we were heartened to find that some coaches were increasingly pressing teachers for mathematical reasoning, we were less certain about adaptations toward explicitness because it could be an indication that some coaches were not implementing the model's inquiry stance with integrity. As a result, we examined how teachers' perceptions of coach explicitness were related to teaching improvement trajectories.

For this analysis, we first examined the teachers' reports on the discussion process survey and categorized them based on their pattern of responses. Nineteen teachers were categorized as reporting increasingly greater explicitness in the interactions with their coach over the course of the year. We then examined whether the teachers who reported increasing explicitness had different patterns of teaching improvement—on our video-based measure of students' opportunities to engage in conceptual thinking—compared with the remaining teachers. We found, however, no differences in growth patterns for these 19 teachers relative to all other teachers.[8] This suggests that, under the right conditions, explicitness may be efficacious; as such, it called for reconsideration (or elaboration) of the principle of taking an inquiry stance with teachers, a component of the coaching model's design.

Study 2: Coach-Driven Inquiry Into Adaptation to Organizational Context

In the first year of the study, we focused primarily on training coaches to take up our coaching model and iteratively refining our perspective on key coaching practices based on the data the coaches submitted capturing their work with their partner teachers. As the coaching framework coalesced around three key coaching practices and a central coaching routine, the work in our second year actively engaged the coaches in the adaptive integration of the coaching framework in their diverse local contexts. In this exploratory study, we reviewed documentation, albeit limited, to explore strategic cases of robust coach use of disciplined inquiry to support model adaptation.

Data Collection

In preparation for our adaptive integration work, we collected several data sources to learn about the variations in local contextual conditions that might influence implementation of the coaching model. At the beginning of Year 1, we had coaches write essays describing their coaching role and context. In a context survey at the end of Year 1, we asked the coaches and their partner teachers to reflect on the way the school and district context shaped their use of the coaching model. We triangulated the survey-based accounts with interviews of the partner teachers, also at the end of Year 1; these interviews focused on their experiences working with a coach who was implementing our model. In addition, we gathered information about an important consideration for adaptive integration of the model—the way schools and districts defined coaching roles—through our ongoing interactions with the coaches throughout the year.

Finally, we documented the coaches' work in PDSA inquiry groups, where they experimented with adaptive integration. We encouraged the coaches to document their engagement in PDSA cycles by completing a PDSA tracking form. A limitation of our study was the incomplete documentation and submission of these forms. In general, project demands on the coaches were high. They documented the coaching cycles by gathering and uploading video recordings and related artifacts, such as assignments and student work, and completed the Coach Tracker log and several surveys throughout the year. The addition of the PDSA tracking form came on top of these other documentation expectations. We documented collaborative work in inquiry groups through observation and artifacts. These data suggested that most of the teachers were engaging in PDSA cycles, despite incomplete documentation and submission of forms, but the observational data did not provide us with consistently detailed information about their PDSA work. As a result, we focused on selecting cases where we had the most complete data. We do not claim these cases as representative of the coaches' work in the project but rather as illustrating the productive use of the PDSA inquiry routine.

Analytic Approach

We laid the groundwork for our adaptive integration work by analyzing the school and district factors that shaped the coaches' work. First, we categorized the coaches with salient role-based descriptors based on what they wrote in their context essays, including, school or district based, the number of schools they worked with, and the defined focus of their work (e.g., teacher support vs. student intervention). We conducted descriptive analyses of the context survey administered to the coaches and teachers, and thematic analyses of the interviews with the partner teachers were undertaken to identify other salient dimensions of context variation. Then, we generated case studies of individual coaches' engagement in adaptive integration through within-coach analysis of data and follow-up conversations with these coaches.

Results

We begin by describing our findings related to variation in coaching roles and the most influential school and district conditions reported by the coaches. Our coaches varied considerably in the way their roles as instructional coaches were defined and operationalized. In Year 1 of the project, most of the coaches (50%) were school-based mathematics coaches working with one or more schools. However, the remaining half not only were responsible for mathematics coaching but also had broader job descriptions, such as school-based instructional coaches responsible for coaching in both mathematics and literacy or district-based coaches working with a number of schools and sometimes multiple content areas. While 13 out of the 32 coaches worked with a single school, the remaining 19 worked with teachers in two or more schools, with nearly half of the 19 working with three or more schools. Two thirds of our coaches worked exclusively with elementary grade mathematics teachers, but the remaining third (N = 13) worked with teachers in Grades K–12 and/or middle and high school teachers, resulting in greater task complexity due to more complex or a wider range of instructional content. The fact that the coaches had duties in more than one content area, working with a wide range of disciplinary content and/or multiple organizations, contributes to the complexity of the type of coaching work we are promoting and in so doing places demands on the coaches to find ways to integrate the coaching model's key practices and routines into already complex job responsibilities.

Overall, the coaches and partner teachers tended to report alignment with the coaching model and the priorities expressed by school and district leaders. That said, one third of the coaches reported lack of agreement between principals' visions of effective mathematics instruction and the vision promoted by our coaching model. This lack of alignment was also voiced by a sizeable minority of our partner teachers through their responses to the surveys and interviews. For example, one teacher said, "The goals were different. My principal's goal is we need to make our growth [on state achievement tests], and my coach's goal for me was to deepen my understanding in mathematics content and work on teaching in that Common Core style." The most salient contextual challenge that emerged from both

Table 2
Adaptive Integration Groups and Distribution of Coaches

	No. of Coaches
Supporting beginning teachers	8
Supporting veteran teachers	6
Managing time to enable intensive coaching interaction	6
Coaching teachers to attend to student conceptual understanding in more procedurally focused contexts	6
Adapting the coaching model for use in literacy	6

teachers and coaches was the beliefs and resistance of other teachers in their schools. While the coaches generally reported that their partner teachers supported the style of teaching promoted in the coaching framework, both coaches and teachers reported that other teachers in their schools tended to emphasize "more procedural" or "traditional" teaching. This suggests an important consideration for the scale up of our coaching model, as coaches will need to work with teachers with a wider range of buy-in and beliefs.

Responding to our emerging understanding of the contextual challenges faced by our coaches, we actively engaged our coaches in experimentation aimed at promoting the network's learning about how the model can be adaptively integrated into diverse local contexts. We trained the coaches to use the PDSA inquiry routine to guide their experimentation with ways to adapt their use of the model while preserving its integrity to the key coaching practices.[9] The coaches volunteered to join groups organized around contextual challenges, which were derived from their compiled survey responses. We initially presented the coaches with a list of contextual challenges we had generated from the survey analysis, and the coaches decided which implementation challenges they wanted to work on. See Table 2 for the list of groups that were formed and the number of coaches selected to work in each group. In their inquiry groups, the coaches collaboratively planned inquiry cycles and shared what they learned from them in subsequent meetings.

In the next section, we provide one illustrative case of PDSA work undertaken by a coach working in the *negotiating time challenges* group. When given a choice of an implementation challenge to address in work with a collaborative group, 80% of coaches at the fourth network meeting indicated *negotiating time challenges* was one of their top three problems of coaching practice, and 23% said this was their number one implementation challenge. We highlight this case for several reasons. It illustrates a coach working on a salient implementation challenge for the network, and several coaches reported trying some of the strategies our focal coach had designed and tested. Additionally, this coach was relatively more complete in her documentation of her iterative PDSA cycles, giving us a rich portrait of the way coaches might use disciplined inquiry to support uptake of the model with integrity. We emphasize that this case represents exceptional engagement in the PDSA process in our network, rather than being representative of the work that all coaches did. However, exceptional cases can generate important learning, particularly for informing our understanding of emergent phenomena such as the use of improvement science in education.

Managing Time to Enable Intensive Coaching Interactions

The coaches in our network reported complex and diffuse role responsibilities. In essays about their coaching role and local context that we asked the coaches to write at the beginning of the project, they described engaging in a wide array of tasks, including direct support for teacher development through observing and providing feedback, co-teaching, orchestrating professional development sessions, intervention support for students, curriculum and assessment development, data analysis and management, as well as general administrative tasks. Our model prioritizes coach engagement in intensive one-on-one coaching cycles, which require approximately 2 or 3 hours of direct interaction with each teacher per cycle, plus approximately 1 hour of additional preparation time. Consequently, we heard from some coaches that they struggled to prioritize time for engagement in the coach-teacher discussion process and ended up having to frequently cancel and reschedule conferences or observations during cycles.

For example, Martha was one of the coaches in our network who identified time constraints as her biggest challenge to implementing the model. As the only mathematics coach in her small rural school district, Martha was charged with working with all mathematics teachers in the three schools in her district. Martha's role included supporting teaching improvement through modeling, coteaching, providing feedback after observations, and leading professional learning community meetings. But she also provided direct instruction during intervention time for struggling students in Grades 4, 6, and 8; developed common assessments and analyzed student data; and provided professional development on the district's math curriculum. In fact, she reported that she routinely spent 50% or more of her time on administrative tasks rather than direct support for teachers and students. Martha received some direction from school and district leaders to focus on providing support to novice teachers, teachers with the lowest evaluation scores, and teachers in Grades 2 through 8. Like most of the coaches in our network, Martha's job was complex, and few structures were defined to support her management of this role complexity.

As Martha began to utilize the coach-teacher discussion process to structure her one-on-one support for teachers, she found that she was having trouble reliably enacting the routine. For example, she completed a preobservation lesson planning conference with a teacher but was not able to observe the teacher teaching the associated lesson because her principal pulled her into a meeting with another teacher without advance notice. Having missed observing the teacher teaching the lesson, the coaching cycle had to be restarted because it is anchored in planning, teaching, and reflecting on a specific mathematical task. Another time, Martha had to cut short her time allotted for a postobservation feedback conversation, which did not enable her to have the deep and specific conversation she had hoped with the teacher.

Facing this challenge, Martha was eager to join the time constraints PDSA work group to learn how she could reserve time for engagement in the coach-teacher discussion process. Through a series of linked PDSA cycles, Martha identified, tested, and refined an adaptation to the coaching model. She first tested a weekly calendar (posted in Google Doc) that specified her coaching schedule for the week, which she shared with school and district administrators. While the administrators acknowledged receiving the calendar, in a subsequent week she missed six scheduled coaching events. Five missed events resulted from administrators directing her to other tasks (e.g., preparing assessment copies), despite having access to her calendar. Consequently, Martha added a step to the scheduling routine that included a brief weekly report sent via email to school and district administrators at the end of each week, summarizing interruptions to her coaching schedule. Martha theorized that seeing a connection between the competing demands made on her time and missed opportunities for coaching would help administrators think more carefully about whether to interfere with scheduled coaching events. As predicted, the administrators began consulting her web-based coaching schedule, and their requests to miss coaching events diminished over time.

The schedule routine became a key component of Martha's coaching practice, helping her to protect scheduled coaching time from interruption by school and district administrators, coordinate her multiple responsibilities, and take up the coaching model with integrity. Martha found that systematically tracking and sharing the number of schedule disruptions helped her iteratively refine the scheduling routine and know when she arrived at an adaptation that improved her uptake of the model with integrity.

Additionally, there is some evidence that Martha improved her consistency in both the quantity and the quality of coaching she provided. For example, her preconferences in Year 1 ranged from 10 to 26 minutes long, with an average of 19.65 minutes, suggesting she was struggling to preserve time for deep and specific conversations. However, once she began PDSA work in Year 2, her preconferences were all 24 minutes or longer, with an average of 26 minutes. Perhaps reflecting her struggle to reserve time for coaching, the depth and specificity of her preconferences in Year 1 were inconsistent, with two conferences scoring 7 points out of 9 on our rigor rubric but others scoring in the 2- to 4-point range. This suggests that

Martha had the skill to conduct preconferences with deep and specific conversations but could not consistently enact this practice, perhaps due to implementation constraints. However, by the end of Year 2, she was consistently engaging teachers in deep and specific conversations in the preconferences.

Martha shared her successes with the routine in a network meeting, and subsequently other coaches incorporated a similar routine into their practice. In this way, calendaring routines became an adaptation to the model, or more aptly an elaboration, that addressed the challenge of integrating intensive one-on-one coaching into diverse local contexts.

In conclusion, we have evidence that PDSA cycles can help coaches learn how to enact the model with integrity in schools with varying contextual conditions. While we cannot make claims about the pervasiveness of the productive use of coach-driven PDSAs for model adaptation in our project through this exploratory study, exceptional cases such as Martha's suggest that this is a potential way to support adaptation and the network's capacity to learn how to scale up the coaching model.

Discussion and Implications

Education studies going back to at least the 1970s acknowledge the importance of adaptation during implementation, emphasizing the need for both designs and local practice to change in ways that are in keeping with the core intent of the policy or innovation, but few studies have found instances of this kind of mutual adaptation in practice (for an exception, see Siskin, 2016). Instructional coaching is a promising intervention to support mathematics instructional improvement at scale, but prior research on coaching implementation points to variable uptake, due in part to contextual variation that presents implementation challenges. This history suggests that effective expansion of coaching programs likely requires that local educators change their practice in tandem with the adaptation of designed coaching programs to varying context conditions. Our research and design process exemplifies a dynamic approach to the implementation and scaling of a coaching model that can support mutual adaptation.

Contributions to Understanding the Role of Adaptation in Scaling Up Coaching Programs

Our findings point to key features of a scalable coaching model. First, coaching models must identify the key coaching practices that constitute the essence of the model. Clarity about key coaching practices provides a focus for supporting changes in coaching practice that the model aims to promote. But identifying key coaching practices should be a dynamic process unfolding as changes in coaching practice are monitored during implementation and scaling, and design conjectures about the relationship between key coaching practices and valued outcomes are tested empirically.

Implementation research should attend to the adaptations that coaches make to core practices when they confront implementation challenges, and it should systematically test the robustness of practice adaptations by assessing whether they contribute to or inhibit valued outcomes (e.g., teacher learning and improvement). Through iterative cycles that test the adaptations coaches make in practice, we can learn what truly constitutes the essence of the model.

In our project, we organized iterative research and design cycles that enabled us to test whether our design conjectures about the key coaching practices were associated with improvements in teaching. We were also able to identify the ways in which the coaches adapted the core practices during implementation. This work was enabled by a comprehensive measurement system that systematically documented coaching and teaching practice, and rapid analytic cycles during our design phase. Rather than assume that the coaches' adaptations of the key coaching practices were lethal mutations, we sought to interrogate our assumptions about what practices contributed to the teaching changes we sought to support.

Our findings from Study 1 illustrate one form of adaptation the coaches employed when implementing our model. The coaches adapted their use of the key coaching practice that promoted taking an inquiry stance during coaching conversations, utilizing questioning to promote teacher practice change rather than more directive statements about what teachers should do when teaching. We found that coaches adapted their use of the inquiry stance in response to perceptions of teachers' responsiveness to coaching. When coaches judged teachers to be more responsive initially, the teachers reported that coaches increasingly pressed them to reason about mathematics and student learning during the prelesson planning conferences (i.e., coaches were observed adopting an inquiry stance). However, when coaches judged teachers to be less responsive, adaption toward greater explicitness was observed.

In our exploratory analyses of teaching, we show that the coaches' adaptation toward explicitness does not appear to be a lethal mutation, as growth among those teachers was consistent with that of all other teachers in the sample. Analyses, such as this example, that explore coaching adaptations as they relate to teaching improvement help us to think about which adaptations preserve (or violate) the integrity of the coaching model. Through iterative examination of the uptake and adaptation of the model's key coaching practice, we gained insights into the threshold of adaptation where the integrity of the model might be lost. In this case, we see evidence that deep and specific conversations about the instructional triangle is the key coaching practice to preserve in the spread of the coaching model, providing guidance for efforts to spread and scale mathematics coaching programs.

The finding that coaches adapted their coaching to perceptions of teacher responsiveness also contributes to our understanding of how coaches adapt their work to accommodate different types of teachers, or what we have called the relational contexts of coaching, perhaps pointing to an implicit developmental theory of stages of coaching. For example, depending on judgments of where teachers are in their thinking and their capacity to follow through, coaches might provide teachers with clearer and more specific directions about how they can develop their teaching practice in the short term. These interactions could remain fairly brief and could be precursors for further learning about how to teach more conceptually in the future. On the other hand, teachers deemed highly responsive might be ready to engage in deeper learning discussions, where coaches press teachers for their mathematical reasoning, in the hope that by doing so, teachers will engage in similar practices with their students. This finding adds nuance to the literature on the relational contexts of teaching, which has tended to advocate either a directive or a responsive stance in coaching (Costa & Garmston, 2002; Dozier, 2006; Deussen et al., 2007). We should interpret our exploratory findings with caution given that they are correlational, not causal, and primarily based on teachers' perceptions of coaching, but they have value in providing a novel relationship that can be explored in future studies. As we continue to explore the efficacy of more or less explicit forms of coaching, we contribute to the field's understanding of the microprocesses of effective coaching practice (Gibbons & Cobb, 2016; Huguet, Marsh, & Farrell, 2014; Mudzimiri, Burroughs, Luebeck, Sutton, & Yopp, 2014) and much-needed theories of mathematics teaching development (Adler, Ball, Krainer, Lin, & Novatna, 2005; Gellert, Espinoza, & Barbé, 2013; Goldsmith & Schifter, 1997; Munter & Correnti, 2017).

Like our study of variation in teacher responsiveness, Study 2 explores another common coaching implementation challenge: managing varying and competing professional responsibilities. Martha's experience is consistent with what we know about the complexity of the coaching role stemming in part from the need to find ways to efficiently support multiple teachers while managing time constraints and multiple demands (Galey, 2016; Mangin & Dunsmore, 2015; Mudzimiri et al., 2014). Many coaches need to continuously manage their relationship with administrators to carve out the maximum amount of time they have to work with teachers, because they are asked to fill other administrative, noncoaching duties (Bean et al., 2010; Camburn, Kimball, & Lowenhaupt, 2008; Gallucci et al., 2010; Mangin, 2009). Even when they work in supportive environments, coaches are faced with issues of efficiency. How many teachers can they reasonably expect to work with? How much time does it take on average to maintain productive contact with teachers? Do some teachers need more or less intense contact in order to make measurable improvement? Balanced with issues of efficiency are issues of effectiveness. What are the most effective progressions for taking teachers of a particular type through a learning progression? Does that learning progression vary for teachers of the same type?

Coaches likely confront these questions, and more, subconsciously every time they engage with teachers.

Martha's work illustrates how coaches might use methods of disciplined inquiry to learn how to manage difficult implementation conditions (Cohen-Vogel et al., 2015; Tichnor-Wagner et al., 2018). By making small and incremental changes to the coaching system in her district, Martha provides insight into the ways coaches can proactively engage in the spread of a coaching program. By making barriers to her deep engagement with teachers visible to administrators whose actions were interfering with her time dedicated to coaching teachers, Martha's inquiry cycles enabled her to engineer better contextual supports for rigorous coaching practice. While our case study approach provides some insight into the adaptive integration process, a key limitation is that we do not systematically explore the efficacy of the adaptions that our coaches generated in the PDSA groups. Further research that explores how adaptive integration can contribute to improvement outcomes is warranted.

Contributions to Our Understanding of Reform Adaptation More Generally: An Evidence-Based Approach to Promoting Mutual Adaptation

Based on our work in the TN Math Coaching Project, we posit a potentially generalizable process for evidence-based approaches to promote mutual adaptation, in the context of research-practice partnerships. Our experience suggests the potential power of beginning a collaborative reform effort with continuous-improvement research aimed at identifying the core features of a reform that contribute to valued outcomes. In the TN Math Coaching Project, we closely monitored and documented implementation of our coaching model and then studied how variation in implementation related to coaching and teaching improvement. These iterative investigations facilitated identification of the essential core of our reform: deep and specific coaching conversations about pedagogy, content, and student thinking. Subsequently, we monitored implementation by tracking how coaches maintained the integrity of the model through enactment of this key coaching practice.

Once the essential core of a reform or intervention is better understood, our experience suggests it is then important to build educator capacity for productive adaptation. In our project, we defined productive adaptations as those that could be enacted while still maintaining integrity to the key coaching practice, and we continued to train and support coaches in enactment of this practice so that they developed a deep understanding of the core features of the model. In tandem, we trained coaches in a principled way to experiment with supportive elaborations of the model that enabled them to head off potential implementation barriers that could result in lethal mutations. Martha's case suggests that practitioners can learn how to implement research-based interventions under varying contextual conditions, through iterative inquiry cycles. More comprehensive studies of adaptive integration would contribute to our understanding of the adaptive integration process.

This process has several implications for promoting mutual adaptation in research-grounded reform efforts. The process model suggests that researcher-driven and practitioner-driven learning about adaptation can work in tandem to produce designs that are well positioned to be taken to scale. Mutual adaptation is one way to conceptualize scale, by focusing on variation in implementation and its contribution to local effectiveness (Coburn, Catterson, Higgs, Mertz, & Morel, 2013). Researcher-driven investigation of educators' adaptations can identify what constitutes the critical elements of an intervention. Understanding the essential core of a reform or intervention clarifies what educators should be trained to do and what should be monitored during implementation (Kisa & Correnti, 2015). To do so requires a comprehensive, practical measurement infrastructure to support rapid analytics and also suggests the value of iterative research and development cycles (Bryk, Gomez, & Grunow, 2011; Tichnor-Wagner et al., 2018; Yeager, Bryk, Muhich, Hausman, & Morales, 2013).

Practitioner-driven learning from adaptation can generate potential elaborations of an intervention that enable its implementation with integrity in varying contexts. This points to the need to build educator capacity to systematically identify and test adaptations. Accordingly, adaptive integration grounded in disciplined inquiry may be a critical professional skill that enables educators to take a productive stance in managing implementation challenges (Cohen-Vogel et al., 2015; Siskin, 2016). But harnessing this capacity for supporting intervention scaling requires that we develop a learning infrastructure including feedback routines that enable research-practice partnerships to identify and spread productive adaptations.

In conclusion, our continuous-improvement research approach to designing a scalable instructional coaching model led us to important insights about mutual adaptation of the intervention during implementation. Our experience points to the need to systematically monitor practice adaptation during implementation. Studying adaptation can help researchers and designers to identify the essential core features of an intervention, spot potentially lethal mutations, and even identify productive adaptation that should be incorporated into the intervention design. Building educator capacity for principled adaptation that maintains the integrity of an intervention is a critical form of professional practice in this reform era that can be facilitated through improvement research methods. Together, these findings provide a potential process model for an evidence-based approach to mutual adaptation that can be explored in future design-based research-practice partnerships.

Appendix

Before running the models, item subscores from the partner-teacher survey were standardized to have a mean of 0 and standard deviation of 1 (see, e.g., Raudenbush, Rowan, & Kang, 1991). The psychometric phase of these models allowed us to examine the multivariate outcomes and properties of the estimable variance between teachers. The analytic phase of these models also allowed us to examine the differences in teacher perceptions of change on the *press* and *explicitness* scales, in addition to examining which covariates predict higher scores at Cycle 1, as well as higher growth slopes over the course of the three cycles.

The Level 1 model is

$$\text{Item Score}_{mij} = \psi_{1ij} * \left(\text{Press}_{mij}\right) + \psi_{1ij} * \left(\text{Explicitness}_{mij}\right) + \varepsilon_{mij}, \tag{A1}$$

where Item Score_{mij} is the Z-scored item for scale m at time i for teacher-coach cycle j; Press_{mij} is a dummy indicator demarcating two items for the scale Press; ψ_{1ij} is the average score on the Press scale at time i for teacher-coach pair j; $\text{Explicitness}_{mij}$ is a dummy indicator demarcating three items for the scale Coach Explicitness; $\psi_{1\text{ij}}$ is the average score on the Explicitness scale at time i for teacher-coach pair j; and ϵ_{mij} is the measurement error for dimension m at time i for teacher-coach pair j.

The Level 2 model is written as follows:

$$\begin{aligned}\psi_{1ij} &= \pi_{10j} + \pi_{11j} * (\text{Time}) + \pi_{1pj} * \left(A_{pi}\right) + e_{1ij} \\ \psi_{2ij} &= \pi_{20j} + \pi_{21j} * (\text{Time}) + \pi_{1pj} * \left(A_{pi}\right) + e_{2ij},\end{aligned} \tag{A2}$$

where π_{10j} is the Press scale score at time 1 (because time is centered at the first cycle) for teacher-coach pair j; Time is coded as one-third at Cycle 1 and 1 at Cycle 3; π_{11j} is the growth slope, in years, for the Press scale; A_{pi} is a set of (p) time-varying covariates measured at each time point; π_{1pj} is the effect of time-varying covariates on the Press scale at each time point; e_{1ij} is residual error normally distributed with mean of 0 and standard deviation of unity; π_{20j} is the Explicitness scale score at time 1 (because time is centered at the first cycle) for teacher-coach pair j; *Time* is coded as one third at Cycle 1 and 1 at Cycle 3; π_{21j} is the growth slope, in years, for the Explicitness scale; A_{pi} is a set of (p) time-varying covariates measured at each time point; π_{2pj} is the effect of time-varying covariates on the Explicitness scale at each time point; e_{2ij} is residual error normally distributed with mean of 0 and standard deviation of unity.

The Level 3 model is written as

$$\pi_{10j} = \beta_{100} + \sum_{q=1}^{Q} \beta_{10q} X_q + r_{10j}$$
$$\pi_{11j} = \beta_{110} + \sum_{q=1}^{Q} \beta_{11q} X_q + r_{11j}$$
$$\pi_{20j} = \beta_{200} + \sum_{q=1}^{Q} \beta_{20q} X_q + r_{20j} \qquad \text{(A3)}$$
$$\pi_{21j} = \beta_{210} + \sum_{q=1}^{Q} \beta_{21q} X_q + r_{21j}$$

where β_{100} is the average *Press* scale score at Cycle 1 across all teacher-coach pairs; X_q is a set of (q) teacher-coach covariates; β_{10q} is the effect of teacher-coach covariates on the *Press* scale score at Cycle 1; r_{10j} is residual error normally distributed with a mean of 0 and a standard deviation of unity; β_{110} is the average linear growth in the *Press* scale over the year across all teacher-coach pairs; X_q is a set of (q) teacher-coach covariates; β_{11q} is the effect of teacher-coach covariates on linear growth in the *Press* scale over the year; r_{11j} is residual error normally distributed with a mean of 0 and a standard deviation of unity; β_{200} is the average *Explicitness* scale score at Cycle 1 across all teacher-coach pairs; X_q is a set of (q) teacher-coach covariates; β_{20q} is the effect of teacher-coach covariates on the *Explicitness* scale at Cycle 1; r_{20j} is residual error normally distributed with a mean of 0 and a standard deviation of unity; β_{210} is the average linear growth in the *Explicitness* scale over the year across all teacher-coach pairs; X_q is a set of (q) teacher-coach covariates; β_{21q} is the effect of teacher-coach covariates on linear growth in the *Explicitness* scale over the year; r_{21j} is residual error normally distributed with a mean of 0 and a standard deviation of unity.

Notes

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305H140112 to University of Pittsburgh. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

[1]The coaching model built on the practical knowledge accumulated by the Institute for Learning (IFL), an outreach of the University of Pittsburgh's Learning Research and Development Center (LRDC). Fellows from the IFL worked in collaboration with Tennessee (TN) Department of Education staff and LRDC researchers to iteratively refine and test the coaching model that ultimately became the TN + IFL Math Coaching Model.

[2]Scoring of the videos led to a construct measuring students' opportunities to engage in conceptual thinking during the lesson.

[3]Our partners from the TN Department of Education distributed an announcement to all the school districts. Interested coaches submitted a written application, which included statements of interest and experience and effectiveness as a coach, and a performance task including identification and analysis of a high-level math task. We received 62

applications. The applicants then participated in a performance-based oral interview, which included analysis of two scenarios of mathematics instruction and role-playing coaching interactions. We scored each component of the process and selected 32 coaches who not only scored relatively higher but also represented variation in prior experience and training, and coaching context (e.g., urban, suburban, and rural.

[4]In this model, we treat the coach-teacher pairs as independent even though each coach worked with two partner teachers. In addition to the models we present, we also ran four-level models where the coach-teacher pairs were nested in coaches. The parameter estimates were roughly the same. Given the sparseness of our data, that is, two teachers per coach, with some teachers missing data on the partner-teacher surveys, we chose to go with the three-level model. Additionally, if coaches do adapt their coaching based on the teacher, then it could be argued theoretically that each coach-teacher interaction represents an independent observation.

[5]We also examined models with additional proxies such as grade level and the cognitive demand of the task as written for each cycle. These models did not change the parameters and were ultimately eliminated for the sake of model parsimony.

[6]Simultaneously, the model estimates of reliability for between-teacher differences in slopes (or linear change over time) were lower, likely due to the combination of a limited number of items per construct, a limited number of time points per teacher, and a low intraclass correlation *coefficient*. However, because the items were primarily chosen to theoretically represent the constructs explicitness and press, we treat this as a potential limitation of our analyses. If anything, the low reliability suggests that we will have a more difficult time identifying covariates that predict variance in those linear slopes.

[7]While there is an association between higher teacher responsiveness and explicitness at Cycle 1 (β_{201} in Table 1), teacher responsiveness does not predict status in subsequent models with time centered differently.

[8]Although the trajectories look slightly different initially, the estimated curves look similar over time. There are no significant differences in the model-based linear ($\beta_{11} = -0.54$, $SE = 0.82$, $p = .508$), quadratic ($\beta_{21} = 1.59$, $SE = 2.18$, $p = .468$), or cubic ($\beta_{31} = -0.92$, $SE = 1.41$, $p = .515$) growth for the two groups of teachers.

[9]PDSA cycles are a structured inquiry routine used to test a change idea, such as a new practice or tool, or a modification to an existing work process (Langley et al., 2009). Each PDSA is, in essence, a mini-experiment involving the introduction of some change, a prediction of what the outcome will be, and the collection of evidence to examine whether the change appears to be an improvement (Hannan et al., 2017). The cycle concludes with analysis and reflection, leading to identification of next steps and often another inquiry cycle.

References

Adler, J., Ball, D. L., Krainer, K., Lin, F. L., & Novatna, J. (2005). Reflections on an emerging field: Researching mathematics teacher education. *Educational Studies in Mathematics*, *60*, 359–381.

Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science*, *333*, 1034–1037.

Bean, R. M., & Carroll, K. E. (2006). The literacy coach as a catalyst for change. In C. Cummins (Ed.), *Understanding and implementing Reading First initiatives: The changing role of administrators* (pp. 139–152). Newark, DE: International Reading Association.

Bean, R. M., Draper, J. A., & Hall, V. (2010). Coaches and coaching in Reading First schools: A reality check. *Elementary School Journal*, *111*, 87–114.

Berends, M. (2000). Teacher-reported effects of New American Schools designs: Exploring relationships to teacher background and school context. *Educational Evaluation and Policy Analysis*, *22*, 65–82.

Berends, M., Bodilly, S., & Kirby, S. N. (2002). *Facing the challenges of whole-school reform: New American Schools after a decade*. Santa Monica, CA: RAND.
Berman, P., & McLaughlin, M. (1978). *Federal programs supporting educational change: A model of education change: Vol. 8. Implementing and sustaining innovations*. Santa Monica, CA: RAND.
Berwick, D. M. (2008). The science of improvement. *JAMA Journal of the American Medical Association*, *299*, 1182–1184.
Biancarosa, G., Bryk, A., & Dexter, E. R. (2010). Assessing the value-added effects of literacy collaborative professional development on student learning. *Elementary School Journal*, *3*, 7–34.
Birman, B. F., Desimone, L., Porter, A. C., & Garet, M. S. (2000). Designing professional development that works. *Educational Leadership*, *57*(8), 28–33.
Blazar, D., & Kraft, M. A. (2015). Exploring mechanisms of effective teacher coaching: Results from two cohorts of an experimental evaluation. *Educational Evaluation and Policy Analysis*, *37*, 542–566.
Bryk, A. S., Gomez, L. M., & Grunow, A. (2011). Getting ideas into action: Building networked improvement communities in education. In M. Hallinan (Ed.), *Frontiers in sociology of education* (pp. 127–162). New York, NY: Springer.
Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. (2015). *Learning to improve*. Cambridge, MA: Harvard Education Press.
Bryk, A. S., Yeager, D. S., Hausman, H., Muhich, J., Dolle, J. R., Grunow, A., & Gomez, L. (2013, June). *Improvement research carried out through networked communities: Accelerating learning about practices that support more productive student mindsets* (A White Paper prepared for the White House meeting on Excellence in Education: The Importance of Academic Mindsets). Retrieved from https://www.carnegiefoundation.org/wp-content/uploads/2014/09/improvement_research_NICs_bryk-yeager.pdf
Camburn, E. M., Kimball, S. M., & Lowenhaupt, R. (2008). Going to scale with teacher leadership: Lessons learned from a district-wide literacy coach initiative. In M. M. Mangin & S. R. Stoelinga (Eds.), *Effective teacher leadership: Using research to inform and reform* (pp. 120–143). New York, NY: Teachers College Press.
Campbell, P. F., & Malkus, N. N. (2011). The impact of elementary mathematics coaches on student achievement. *Elementary School Journal*, *111*, 430–454.
Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher*, *32*(6), 3–12.
Coburn, C. E. (2004). Beyond decoupling: Rethinking the relationship between the institutional environment and the classroom. *Sociology of Education*, *77*, 211–244.
Coburn, C. E., Catterson, A. K., Higgs, J., Mertz, K., & Morel, R. (2013). *Spread and scale in the digital age: A memo to the John D. and Catherine T. MacArthur Foundation*. Prepared for a convening at the MacArthur Foundation, November 7 and 8, 2013. Retrieved from https://www.informalscience.org/sites/default/files/spread_and_scale_in_a_digital_age_12-31-13_to_share.pdf
Coburn, C. E., & Russell, J. L. (2008). District policy and teachers' social networks. *Educational Evaluation and Policy Analysis*, *30*, 203–235.
Coburn, C. E., & Woulfin, S. L. (2012). Reading coaches and the relationship between policy and practice. *Reading Research Quarterly*, *47*, 5–30.
Cohen, D. J., Crabtree, B. F., Etz, R. S., Balasubramanian, B. A., Donahue, K. E., Leviton, L. C., . . . Green, L. W. (2008). Fidelity versus flexibility: Translating evidence-based research into practice. *American Journal of Preventive Medicine*, *35*(5 Suppl.), S381–S389.
Cohen-Vogel, L., Tichnor-Wagner, A., Allen, D., Harrison, C., Kainz, K., Socol, A. R., & Wang, Q. (2015). Implementing educational innovations at scale: Transforming

researchers into continuous improvement scientists. *Educational Policy*, *29*(1), 257–277.

Correnti, R., & Martinez, J. F. (2012). Conceptual, methodological, and policy issues in the study of teaching: Implications for improving instructional practice at scale. *Educational Assessment, 17*(2–3), 51–61.

Costa, A., & Garmston, R. (2002). *Cognitive coaching: A foundation for renaissance schools* (2nd ed.). Norwood, MA: Christopher-Gordon.

Desimone, L., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis*, *24*(2), 81–112.

Desimone, L. M., & Pak, K. (2016). Instructional coaching as high-quality professional development. *Theory Into Practice*, *56*, 3–12.

Deussen, T., Coskie, T., Robinson, L., & Autio, E. (2007). *"Coach" can mean many things: Five categories of literacy coaches in Reading First* (REL 2007-No. 005). Washington, DC: Regional Educational Laboratory Northwest.

Domina, T., Lewis, R., Agarwal, P., & Hanselman, P. (2015). Professional sense-makers: Instructional specialists in contemporary schooling. *Educational Researcher*, *44*, 359–364.

Dozier, C. (2006). *Responsive literacy coaching: Tools for creating and sustaining purposeful change*. Portland, ME: Stenhouse.

Elmore, R. F. (1996). Getting to scale with good educational practice. *Harvard Educational Review*, *66*(1), 1–26.

Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly*, *48*, 94–118.

Foster, D., & Noyce, P. (2004). The Mathematics Assessment Collaborative: Performance testing to improve instruction. *Phi Delta Kappan*, *85*, 367–374.

Galey, S. (2016). The evolving role of instructional coaches in U.S. policy contexts. *The William & Mary Educational Review*, *4*, 54–71.

Gallucci, C., Van Lare, M. D., Yoon, I. H., & Boatright, B. (2010). Instructional coaching: Building theory about the role and organizational support for professional learning. *American Educational Research Journal*, *47*, 919–963.

Garet, M. S., Wayne, A., Stancavage, F., Taylor, J., Eaton, M., Walters, K., & . . . Warner, E. (2011). *Middle school mathematics professional development impact study: Findings after the second year of implementation* (NCEE 2011-4024). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Gawande, A. (2007). The checklist. *The New Yorker*, *83*(39), 86–95.

Gellert, U., Espinoza, L., & Barbé, J. (2013). Being a mathematics teacher in times of reform. *ZDM: The International Journal on Mathematics Education*, *45*(4), 535–545.

Gibbons, L. K., & Cobb, P. (2016). Content-focused coaching practices implicated in designing potentially productive coaching activities to support teachers' learning. *Elementary School Journal*, *117*, 237–260.

Goldsmith, L. T., & Schifter, D. (1997). Understanding teachers in transition: Characteristics of a model for the development of mathematics teaching. In E. Fennema & B. S. Nelson (Eds.), *Mathematics teachers in transition* (pp. 19–54). Mahwah, NJ: Lawrence Erlbaum.

Hannan, M. Q., Russell, J. L., Takahashi, S., & Park, S. (2015). Improving feedback and support for beginning teachers: The case of the Building a Teaching Effectiveness Network. *Journal of Teacher Education*, *66*, 494–508.

Hiebert, J., & Stigler, J. W. (2004). A world of difference. *Journal of Staff Development*, *25*(4), 10–15.

Honig, M. I. (Ed.). (2006). *New directions in education policy implementation: Confronting complexity*. Albany: State University of New York Press.

Huguet, A., Marsh, J. A., & Farrell, C. C. (2014). Building teachers' data-use capacity: Insights from strong and developing coaches. *Education Policy Analysis Archives*, *22*(52), 1–28.

Ippolito, J. (2010). Three ways that literacy coaches balance responsive and directive relationships with teachers. *Elementary School Journal*, *111*, 164–190.

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research*, *86*, 945–980.

Killion, J. (2012). Coaching in the K–12 context. In J. Fletcher & C. A. Mullen (Eds.), *The SAGE handbook of mentoring and coaching in education* (pp. 273–295). Thousand Oaks, CA: Sage.

Kisa, Z., & Correnti, R. (2015). Examining implementation fidelity in America's Choice schools: A longitudinal analysis of changes in professional development associated with changes in teacher practice. *Educational Evaluation and Policy Analysis*, *37*, 437–457.

Kraft, M. A., Blazar, D., & Hogan, D. (2018). *The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence* (Working paper). Brown University, Providence, RI and Harvard University, Cambridge, MA.

Langley, G. J., Moen, R., Nolan, K. M., Nolan, T. W., Norman, C. L., & Provost, L. P. (2009). *The improvement guide: A practical approach to enhancing organizational performance*. San Francisco, CA: Jossey-Bass.

LeMahieu, P. G. (2011). *What we need in education is more integrity (and less fidelity) of implementation*. Stanford, CA: Carnegie Foundation for the Advancement of Teaching.

Lewis, C. (2015). What is improvement science? Do we need it in education? *Educational Researcher*, *44*, 54–61.

Lewis, C. C., Perry, R. R., & Hurd, J. (2009). Improving mathematics instruction through lesson study: A theoretical model and North American case. *Journal of Mathematics Teacher Education*, *12*, 285–304.

Lord, B., Cress, K., & Miller, B. (2003). *Teacher leadership as classroom support: The challenges of scale and feedback in mathematics and science education reform*. Newton, MA: Centre for Leadership and Learning Communities Education Development Center.

Mangin, M. M. (2009). Literacy coach role implementation: How district context influences reform efforts. *Educational Administration Quarterly*, *45*, 759–792.

Mangin, M. M., & Dunsmore, K. (2015). How the framing of instructional coaching as a lever for systemic or individual reform influences the enactment of coaching. *Educational Administration Quarterly*, *51*, 179–213.

Mangin, M. M., & Stoelinga, S. R. (2011). Peer? Expert? Teacher leaders struggle to gain trust while establishing their expertise. *Journal of Staff Development*, *32*(3), 48–51.

Matsumura, L. C., Garnier, H. E., & Resnick, L. B. (2010). Implementing literacy coaching: The role of school social resources. *Educational Evaluation and Policy Analysis*, *32*, 249–272.

Matsumura, L. C., Garnier, H. E., & Spybrook, J. (2012). The effect of content-focused coaching on the quality of classroom text discussions. *Journal of Teacher Education*, *63*, 214–228.

Matsumura, L. C., Garnier, H. E., & Spybrook, J. (2013). Literacy coaching to improve student reading achievement: A multi-level mediation model. *Learning and Instruction*, *25*, 35–48.

Matsumura, L. C., Sartoris, M., Bickel, D. D., & Garnier, H. E. (2009). Leadership for literacy coaching: The principal's role in launching a new coaching program. *Educational Administration Quarterly*, *45*, 655–693.

McDonald, J. P., Klein, E. J., & Riordan, M. (2009). *Going to scale with new school designs: Reinventing high school*. New York, NY: Teachers College Press.

McLaughlin, M. (1976). Implementation as mutual adaptation: Change in classroom organization. *Teachers College Record*, *77*, 342–344.

McLaughlin, M. (1987). Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis*, *9*, 172.

Mudzimiri, R., Burroughs, E. A., Luebeck, J., Sutton, J., & Yopp, D. (2014). A look inside mathematics coaching: Roles, content, and dynamics. *Education Policy Analysis Archives*, *22*, 1–32.

Munter, C., & Correnti, R. (2017). Examining relations between mathematics teachers' instructional vision and knowledge and change in practice. *American Journal of Education*, *123*, 171–202.

Murray, S., Ma, X., & Mazur, J. (2009). Effects of peer coaching on teachers' collaborative interactions and students' mathematics achievement. *Journal of Educational Research*, *102*, 203–212.

Neufeld, B., & Roper, D. (2003). *Coaching: A strategy for developing instructional capacity: Promises and practicalities*. Washington, DC: Aspen Institute Program on Education.

Neuman, S. B., & Cunningham, L. (2009). The impact of professional development and coaching on early language and literacy instructional practices. *American Educational Research Journal*, *46*, 532–566.

Opfer, V. D., & Pedder, D. (2011). Conceptualizing teacher professional learning. *Review of Educational Research*, *81*, 376–407.

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*, 331–337.

Peurach, D. J. (2011). *Seeing complexity in public education: Problems, possibilities, and success for all*. New York, NY: Oxford University Press.

Polly, D. (2012). Supporting mathematics instruction with an expert coaching model. *Mathematics Teacher Education and Development*, *14*, 78–93.

Powell, D. R., Diamond, K. E., Burchinal, M. R., & Koehler, M. J. (2010). Effects of an early literacy professional development intervention on Head Start teachers and children. *Journal of Educational Psychology*, *102*, 299–312.

Raudenbush, S. W., Rowan, B., & Kang, S. J. (1991). A multilevel, multivariate model for studying school climate with estimation via the EM algorithm and application to US high-school data. *Journal of Educational Statistics, 16*(4), 295–330.

Reiser, B. J., Spillane, J. P., Steinmuller, F., Sorsa, D., Carney, K., & Kyza, E. (2000). Investigating the mutual adaptation process in teachers' design of technology-infused curricula. In B. Fishman & S. O'Connor-Divelbiss (Eds.), *Proceedings of the Fourth International Conference of the Learning Sciences* (pp. 342–349). Mahwah, NJ: Lawrence Erlbaum.

Russell, J. L., Correnti, R. C., Stein, M. K., Bill, V., Booker, L., & Schwartz, N. (2016). *Coaching to improve mathematics instruction in Tennessee* (Working paper). University of Pittsburgh, Pittsburgh, PA.

Sailors, M., & Price, L. R. (2010). Professional development that supports the teaching of cognitive reading strategy instruction. *Elementary School Journal*, *110*, 301–322.

Sandholtz, J. H. (2002). Inservice training or professional development: Contrasting opportunities in a school/university partnership. *Teaching and Teacher Education*, *18*, 815–830.

Sherer, J. Z., & Spillane, J. P. (2011). Constancy and change in work practice in schools: The role of organizational routines. *Teachers College Record*, *13*, 611–657.

Siskin, L. S. (2016). Mutual adaptation in action. *Teachers College Record*, *118*(13), 1–18.

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

Stein, M. K., Correnti, R., Moore, D., Russell, J. L., & Kelly, K. (2017). Using theory and measurement to sharpen conceptualizations of mathematics teaching in the Common Core era. *AERA Open*, *3*(1). Retrieved from https://journals.sagepub.com/doi/pdf/10.1177/2332858416680566

Stelitano, L., Russell, J. L., & Bray, L. E. (2017, April). *Organizing for inclusion: Exploring the routines that shape student supports*. Paper presented at the American Educational Research Association annual meeting, San Antonio, TX.

Supovitz, J. (2008). Instructional influence in American high schools. In M. M. Mangin & S. R. Stoelinga (Eds.), *Effective teacher leadership* (pp. 144–162). New York, NY: Teachers College Press.

Thompson, C. L., & Zeuli, J. S. (1999). The frame and the tapestry: Standards-based reform and professional development. In L. Darling Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 341–375). San Francisco, CA: Jossey-Bass.

Tichnor-Wagner, A., Allen, D., Socol, A. R., Cohen-Vogel, L., Rutledge, S. A., & Xing, Q. W. (2018). Studying implementation within a continuous continuous-improvement process: What happens when we design with adaptations in mind? *Teachers College Record*, *120*(5).

Weinbaum, E. H., & Supovitz, J. A. (2010). Planning ahead: Make program implementation more predictable. *Phi Delta Kappan*, *91*(7), 68–71.

Yeager, D., Bryk, A., Muhich, J., Hausman, H., & Morales, L. (2013). *Practical measurement*. Palo Alto, CA: Carnegie Foundation for the Advancement of Teaching.

Yoon, K. S., Garet, M., Birman, B., & Jacobson, R. (2007). *Examining the effects of mathematics and science professional development on teachers' instructional practice: Using professional development activity log*. Washington, DC: Council of Chief State School Officers.

Manuscript received March 23, 2017
Final revision received February 28, 2019
Accepted April 17, 2019