THE NATIONAL CENTER ON SCALING UP EFFECTIVE SCHOOLS

Using Data for Improvement to Support Implementation At-Scale:

Adaptive Integration in the TN Mathematics Coaching Project

Jennifer Lin Russell | Richard Correnti | Mary Kay Stein
Maggie Hannan | Victoria Bill | Nate Schwartz | Laura Booker | Nicole Roberts Pratt

Conference Paper
October 2015

VANDERBILT PEABODY COLLEGE

EDC Learning transforms lives. | Florida State University 1851 | Georgia State University | The University of North Carolina at Chapel Hill | Wisconsin University of Wisconsin–Madison

The National Center on Scaling Up Effective Schools (NCSU) is a national research and development center that focuses on identifying the combination of essential components and the programs, practices, processes and policies that make some high schools in large urban districts particularly effective with low income students, minority students, and English language learners. The Center's goal is to develop, implement, and test new processes that other districts will be able to use to scale up effective practices within the context of their own goals and unique circumstances. Led by Vanderbilt University's Peabody College, our partners include The University of North Carolina at Chapel Hill, Florida State University, the University of Wisconsin-Madison, Georgia State University, the University of California at Riverside, and the Education Development Center.

This paper was presented at NCSU's second national conference, Using Continuous Improvement to Integrating Design, Implementation, and Scale Up. The conference was held on October 7-9, 2015 in Nashville, TN. The authors are:

Jennifer Lin Russell
Richard Correnti
Mary Kay Stein
Maggie Hannan
Victoria Bill
University of Pittsburgh

Nate Schwartz
Laura Booker
Nicole Roberts Pratt
Tennessee Department of Education

This research was conducted with funding from the Institute of Education Sciences (R305C10023). The opinions expressed in this article are those of the authors and do not necessarily represent the views of the sponsor or the National Center on Scaling Up Effective Schools.

This work was supported by an Institute of Education Sciences, Continuous Improvement Research in Education grant (#R305H140112). The views and content represented in the paper are solely those of the paper authors.

The history of education reform efforts is rife with examples of instructional innovations that falter when attempts are made to implement them at scale. As instructional innovations are taken from the specific contexts where they were developed and tested 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. We are engaged in a collaborative project to test and refine a model for mathematics instructional coaching that is designed to be a resource for schools and districts throughout the state of Tennessee as they support the transition to teaching that is aligned with rigorous, college and career ready mathematics standards. Our project team strategically integrates the expertise and systemic positioning of policymakers in the Tennessee Department of Education (TDOE), and researchers from the Learning Research and Development Center and professional development practitioners from the Institute for Learning, both at the University of Pittsburgh.

Our joint work focuses on mathematics instructional coaching because coaching work lies at a critical intersection in the process of policy implementation. Coaches operate as boundary spanners between district and school contexts, and their roles are often fluid and heavily dependent on unique and specific contextual factors. Coaching is shaped by the demands of multiple levels of the educational system, and varies significantly from school to school and district to district (Coburn & Russell, 2008). Consequently, in order to be effective, coaching policy interventions must plan explicitly for the adaptive integration of a coaching approach into diverse local instructional systems. By invoking the adaptive integration concept, we signal the inspiration we are drawing for our work from the networked improvement community concept which describes an organizational structure for engaging researchers, practitioners and policymakers in disciplined inquiry to support practice improvement (Bryk, Gomez, Grunow & LeMahieu, 2015; Hannan, Russell, Park & Takahashi, 2015). In our case, the networked improvement community is made up of 32 instructional coaches representing 21 school districts spread throughout the state of Tennessee working in collaboration with policymakers at the TDOE, researchers from the University of Pittsburgh, and professional development practitioners from the Institute for Learning.

Adaptive integration is an essential approach to instructional improvement at scale given the structure of the educational system in the United States. In our case, our partners from the TDOE are trying to develop policy that can affect instruction in diverse settings by utilizing the instruments available to a state department of education and remaining cognizant of their limited control over key instructionally relevant decisions made at the local level, such as teacher and coach hiring and curriculum selection. TDOE's efforts to impact practice through the decisions and actions of local agents who operate in diverse settings define a fundamental *problem of education policy*. At the school and district level, coaches likely work within a more coherent system but confront *problems of practice*, such as the need to be responsive to individual teachers who need very different levels of support. Given this complex system in which coaching occurs, a rigidly structured coaching program that must be implemented with fidelity is not likely to be successful. Instead, our work draws on the people doing the work (the coaches) to engage in systematic inquiry (as led by our research team) to test and refine a framework for coaching that proves successful across individuals and contexts. Years of implementation studies suggest that most large-scale programs do not successfully address the above problems of practice and policy simultaneously. We believe that the combination of NICs and a deep research partnership that includes all the right stakeholders might help us do so.

From a research perspective, understanding and guiding such a practice improvement effort further confronts *problems of research practice*. Traditionally, the research community has sought to establish the efficacy of promising interventions through design-based research approaches, only to encounter implementation challenges when trying to get those innovations into practice at a large scale. In our efforts to support both policy and practice improvement efforts, we are confronted with the need to simultaneously build an evidence base to support and refine the efficacy of our coaching framework, while simultaneously learning about how to get more and more practitioners to take it up in practice. In our joint work, we marry traditional research approaches that establish the efficacy of our coaching framework with more practical improvement research methods that explore how to embed the framework in a model for taking the framework into practice at scale. In particular, we explore the feasibility of developing a practical measurement system that practitioners can use to monitor the extent to which their efforts are both aligned to deep principles of the coaching framework and responsive to their contexts.

The goal of this paper is to describe how our project is approaching the challenges of testing, implementing and scaling our coaching model. Specifically we highlight how we are explicitly promoting the adaptive integration of the coaching model into diverse local contexts. Adaptive integration work is evident in the micro-level adjustments to coaching practices that our coaches undertake as they work with different teachers, and our more explicit attempt to promote experimentation with adjustments to the coaching framework utilizing the methods of improvement research. In so doing we employ novel analytic approaches and leverage the social resources of a networked improvement community.

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 mathematics teaching practice. For example, in mathematics, students are expected to not only develop fluency in the use of mathematical procedures but also to develop deep conceptual understanding and the capacity to think, reason, and problem solve as they grapple with non-routine problems. The kinds of learning environments that support such learning are challenging to set up and, not surprisingly, somewhat atypical in US classrooms (Hiebert & Stigler, 2004).

Research tells us that changing instructional practice to become more rigorous and conceptually demanding is not a trivial undertaking (Neergaard et al., 2011; Thompson & Zeuli, 1999). Teachers do not learn to enact more ambitious forms of instruction simply by attending workshops based on general principles of good instruction; rather this kind of transformative shift in teaching practice is best achieved with training that is directly tied to the discipline (mathematics) and to the specific curricula or instructional tasks that teachers use day-to-day in their classrooms. It is under these conditions that teachers come to appreciate the relationship between the pedagogical strategies, the mathematical content and skills to be taught and the ways in which students are or are not making sense of the mathematical ideas.

Workshop-based, short-term trainings are the most common form teacher professional development (Birman, Desimone, Porter, & Garet, 2000; Sandholtz, 2002). When well designed, these kinds of training sessions have value for increasing teacher knowledge, which is important for improving teaching practice (Yoon et al., 2007). In an internal study of the effectiveness of teacher trainings, TDOE researchers found that teachers that taught in the same school as one of the teachers that had been specially trained to provide state-sponsored professional were more likely to improve in their instructional practices. However, there is general agreement in the field that full mastery of complex skills requires *engaging in the skills to be learned* with guidance from a more expert practitioner (Greeno, Collins, & Resnick, 1996). Highly skilled instructional coaches can provide this guidance to teachers.

There is an emerging evidence base to support mathematics coaching (c.f. Foster & Noyce, 2004; Polly, 2012), though the field has not come to consensus about what constitutes effective coaching practice. A 3-year randomized control study found that over time coaches positively affected elementary student achievement, particularly after coaches gained experience and skill through extensive professional development (Campbell & Malkus, 2011). In contrast, a small-scale experimental study of a peer mathematics coaching initiative found that coaches did not take an analytic stance with respect to their colleague's instruction and found coaching discussions lacked depth (Murray, Ma, & Mazur, 2009). In addition, coaches often resist providing the critical, substantive feedback to teachers that challenges them to change their practice (Lord, Cress & Miller, 2003).

The challenge of producing effective coaching suggests the need to provide extensive training to teachers who transition into this role. This is consistent with others that have argued that while people who enter coaching roles may be relatively expert teachers, they should be considered learners who need to develop coaching skills and dispositions (Gallucci, Van Lare, Yoon, & Boatright, 2010). Others have found that the selection and training of coaches influences the degree to which they are a source of teaching expertise for colleagues (Coburn & Russell, 2008). However, empirical data on how coaches learn to take on this new role is quite limited (Gallucci et al., 2010). Many TN districts have instructional coaching positions, but there is reason to believe that the way coaching roles and practices are enacted varies considerably, oftentimes departing from what research says about effective coaching practices. As a result, practitioners and policymakers aiming to implement coaching programs are confronted with a situation in which we lack both an empirical base that identifies best practices and the knowledge about how to get what we do know into practice. Despite the ubiquity of this dual challenge, most educational reform strategies address only one challenge at a time. For example, traditional research methods might be used to establish best practices and then improvement science steps in to tackle obstacles to instantiating those practices into the diverse work practices and settings in which practitioners find themselves. Nevertheless, practitioners do not have the luxury of waiting for traditional science to establish "what works" before beginning improvement efforts. Here, we attempt to address both challenges simultaneously.

Networked Improvement Communities and the Adaptive Integration

The networked improvement community (NIC) concept aims to address this dual challenge by building an evidence base about both productive practices and knowledge of implementation processes to address persistent problems of practice and policy. Bryk and colleagues (2015) argue that a NIC "unites the conceptual and analytic discipline of improvement science with the power of networked communities to innovate and learn together." Improvement science has a long history in the manufacturing industries and subsequently the healthcare fields, and provides a disciplined methodology for learning from practice to improve systems and processes that shape work within organizations (Berwick, 2008; Gawande, 2007; Langley et al., 2009; Lewis, 2015). Improvement research provides a disciplined approach to learning from practice, which can cumulate in practical knowledge (Bryk et al., 2013; Hiebert, Gallimore, & Stigler, 2002). While the primary goal of traditional research is contribution to the field's broader knowledge, inquiry in the context of a NIC is principally aimed at contributing to the ongoing implementation and improvement of interventions. Improvement science carried out through networked forms of organizing accelerates the capacity to learn from practice to improve it (Bryk, et al., 2015). A networked improvement community is marked by four essential characteristics that exemplify their function as a scientific learning community (Bryk, Gomez, Grunow & LeMahieu, 2015). A NIC is:

- focused on a well-specified common aim;
- guided by deep understanding of the problem, the system that produces it, and a shared working theory to improve it;
- disciplined by the methods of improvement research to develop, test, and refine intervention; and
- organized to accelerate interventions into the field and to effectively integrate them into varied educational contexts.

In practice, the work of a networked improvement community has three primary components: developing a strong conceptualization of the problem of practice, its root causes and a theory for how to improve the problem; designing and testing better work processes (tools, work roles, supportive norms) in accordance with a theory of improvement; and focusing explicit attention to the adaptive integration of promising work processes into diverse local contexts (Bryk, Gomez, Grunow & LeMahieu, 2015).

As they have sought to bring the NIC concept to life through experimentation with the concept in action, Bryk and colleagues have pointed to the way NICs can support the adaptive integration of innovative change ideas developed by the network into diverse local contexts, which we argue is a key aspect of implementation and scaling innovations. Adaptive integration involves using improvement research methods to integrate a standard work process 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). NICs facilitate the use of improvement research methods, such as disciplined inquiry methods, to adaptively integrate promising innovations into diverse local contexts. For example, schools participating in the Building a Teaching Effectiveness Network used PDSA cycles to adapt a standard process for coordinating feedback for beginning teachers into the existing constraints of their school systems; the extent to which schools took an improvement science mindset to the adaptive integration problem predicted their successful enactment of the feedback process (Hannan, Russell, Takahashi & Park, 2015). In the context of our work, coaches are engaged in the adaptive integration of the coaching framework into their diverse local contexts.

Methodological Approach

After being awarded a Continuous Improvement Research in Education grant from the Institute of Education Sciences in spring of 2014, project leads from LRDC, TDOE and IFL took rapid action to launch a networked improvement community – the TN Mathematics Coaching Network – to organize the implementation and testing of the TN Mathematics Coaching Model. The network defines the community of coaches that partner with us in the testing and iterative refinement of our coaching framework.

Setting

Tennessee has taken a unique approach to the transition to more rigorous standards. At the outset, the TDOE focused on what could be put in place to support the state's transition to college- and career-ready standards by influencing the instruction that students receive every day. In state-wide training, the state has focused on providing high-quality, instructionally focused professional development to both teachers and leaders. This support began in 2012 with a focus in mathematics and an initial partnership between the state and the IFL at the University of Pittsburgh.

IFL played a vital role in three components of the TDOE's theory of change: quality training and support, strong models of content materials and resources, and transparent and aligned assessments. IFL and TDOE worked together to create training that was aligned with high-quality materials and updated, more rigorous assessments. The combination of these components created a tightly-aligned vision of how to support teachers' enactment of lessons to meet the expectations of the standards and assessments.

Additionally, Tennessee implemented statewide teacher evaluation in the 2011-12 school year, which set the stage for a common vision of instructional excellence – a necessary foundation for effective coaching work. After the initial implementation, the state began working to build the educator workforce toward that vision. Over the past four years, educator training provided by the TDOE has focused primarily on teachers and principals. Instructional coaches often participated in both principal and teacher trainings, but no support was specifically targeted for the skills needed for effective instructional coaching. The state's experience shows that coaches are hungry for this support, though often unlikely to receive it.

At the same time, many districts across the state used funds from Race to the Top to increase the number of instructional coaches. During this time of rapid increase in the number of positions, there was little to no clear guidance or support around the role of an instructional coach. These factors led to the department's growing interest and focus on instructional coaching.

Coach and Teacher Sample

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 one of the study. We selected coaches based on their general capacity to engage 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 versus district coaches). For all of the selected coaches, there was a gap between the type of coaching our model promotes and their current practice, and for most that gap was quite significant.

Coaches signed a contract with the TDOE accepting the following responsibilities: attend three 2-day Network meetings in Nashville per year in years 1 and 2; participate in periodic Webinars between meetings; 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 Instructional Coaching Model. Two coaches left the network at the end of year one – one due to retirement and the other due to a change in job responsibilities – and were replaced with two coaches from the original applicant pool. Each coach selected two partner teachers to be participants in the study who teach mathematics in grades 3 through 8. In year two, coaches were asked to continue working with their year one partner teacher that had the most room for improvement in their teaching practice and replace the second original year one partner teacher with a new partner teacher for the second year of the project. To date, our sample consists of 106 partner teachers.

Data Collection

Our approach to data collection documents coaching and teaching practice through a combination of traditional intensive research measures and an emerging set of practical measures. As we prepare to take the coaching model to scale in year 3, the practical, scalable measures become increasingly important as they provide signals about the efficacy of the model to both practitioners and researchers. In this paper, we draw on three primary data sources: (1) video footage and artifacts from coaching cycles; (2) log entries completed by coaches for each interaction with their two partner teachers; and (3) periodic surveys sent to partner teachers and coaches.

Coaches and their partner teachers participate in three formal coaching cycles using the Coach-Teacher Discussion Process, which is a key part of our coaching framework. Figure 1 describes the steps in the process. The bolded statements highlight three key coaching practices, which have been identified through our iterative work examining variation in coaching interactions in dialogue with theory about effective mathematics instructional coaching practices. Coaches completed the Discussion Process with each of their partner teachers at least once during each of three improvement cycles in year one of the study. In order to capture intensive data on coaching practice, each enactment of the Discussion Process was documented by coaches who gathered: videotapes of pre-observation conferences, lessons and post-observation conferences; teacher and coach planning and reflection notes; and instructional artifacts such as the task and five pieces of student work representing a range of learners. Coaches shared these multiple data sources with the network hub by uploading them to a shared folder.

In order to get a broader perspective on coach engagement with each partner teacher, we developed a web-based coaching log – Coach Tracker – that coaches were asked to complete every time they had an interaction with one of their two partner teachers. These log entries consisted of items that asked coaches 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 coaches to rate teachers responsiveness to coaching, specifically their judgment of the likelihood that teachers would follow through on the teaching ideas discussed during this interaction, and reflect on the efficacy of their use of key coaching practices. Coaches could also use a series of open-ended questions as a way to note goals for teachers and reflect on their coaching practice. Coaches had access to past entries in Coach Tracker making it a potential tool for tracking and reflecting on their ongoing work with their partner teachers over time.

Finally, we administered periodic surveys to coaches and teachers throughout year one of the study. Partner teachers were asked to complete the *Discussion Process Survey* at the end of each time they participated in the Coach-Teacher Discussion Process. This survey asked teachers to respond to sets of items that aligned with key coaching practices in our framework: the extent to which they had deep and specific conversations about mathematics, student learning and pedagogy; the extent to which they set goals for teaching improvement; and the nature of the feedback they received from their coach. Other items tapped into the relational components of the coach-teacher interactions measuring constructs such as trust and collaboration. Coaches and teachers completed an *Instructional Survey* at the beginning and end of the school year, which measured beliefs about mathematics teaching particularly with respect to their stance toward engaging students in productive struggle and prioritizing student conceptual understanding of mathematics. The teacher version also included a set of items aiming to get broader information about the social supports for mathematics teaching that teachers have access to from colleagues, coaches and administrators. An end of year *Context Survey* asked coaches and teachers about the extent to which their school and district context supported or hindered their engagement in the type of coaching promoted by our model. It also gathered standard information about each coach's local defined role such as the number of schools they work with and the nature of their responsibilities. Finally, partner teachers were given an opportunity to volunteer to participate in an interview about their experience with the coaching project and approximately half of teachers opted to be interviewed.

Analytic Approach

By systematically tracking the implementation and impact of the Model in diverse contexts throughout the state of TN, we are building an evidence base to support inferences about what coaching practices work under which contextual conditions. Our analytic work is organized around key components of our coaching model and the relationship between components (see Figure 2). Specifically, we engaged in ongoing analysis of teachers' instructional beliefs and practices, the nature and quality of coaching interactions, and the role of context in shaping coaching practices.

Instructional beliefs and practices. In order to identify changes in teaching practices throughout the year, we coded the classroom videos from each of the three instances of the Coach-Teacher Discussion Process for each partner teacher for maintenances of the cognitive demand of the instructional task. Our coaching framework focuses on building the capacity of teachers to orchestrate high level mathematics tasks in the classroom, because prior research suggests that high level tasks provide opportunities for students to build conceptual understanding by engaging in productive struggle (Stein, Grover, & Henningsen, 1996; Stein & Lane, 1996). By coding the extent to which the instructional tasks are high level and then what happens as teachers set up the task with students and support student engagement in tasks we can see the extent to which students have opportunities to build conceptual understanding and engage in reasoning around mathematics. We employed the procedures described by Stein and Kaufman (2010) to code cognitive demand as written in the task, at lesson set up and during lesson enactment. Because coaching cycles were intended to engage only with high cognitive demand tasks, we were primarily interested in the ability of the partner-teachers to maintain high cognitive demand throughout the lesson. Prior research has shown that lessons tend to decline in cognitive demand, with only a minority of teaching maintaining high cognitive demand all the way through the lesson (Stein, Grover, & Henningsen, 1996). We measured the *maintenance of cognitive demand* from the task-as-written to task-as-setup (rubric score from 1 to 4) and from task-as-setup to task-as-enacted (rubric score from 1 to 4), and then summed them creating a scale from 2 to 8. Our partner-teachers exhibited significant growth (ES=.44)[3] in their maintenance of cognitive demand over the three cycles (see Figure 3). Importantly, they accomplished this while the intensity of assistance provided by coaches during classroom observations declined in a qualitatively significant way by the third cycle.

Further information about teachers' instructional beliefs and practices was revealed through our analysis of the beginning- and end-of-year administrations of our *Instructional Survey,* which includes factors associated with teachers' endorsement of teaching for conceptual

[3] Assuming that there was progress from baseline to the end of cycle 1, this estimate may be conservative because it includes only the interval from the end of cycle 1 to the end of cycle 3 (i.e., the linear growth parameter estimate is 1.3, but the interval for which we have measured is only .87; the raw standard deviation averaged across all time points is 2.00, with a notable decrease from 2.55 to 1.87 from cycle 1 to cycle 3). Scoring of the videos was conducted by a group of mathematics researchers, primarily consisting of assistant professors, highly familiar with the Task Analysis Guide and features of high cognitive demand teaching. Videos were randomly assigned to coders who were blind to the cycle being scored. For further information about the reliability of scoring and the controls added to the analyses please contact the lead author.

understanding and productive student struggle. Our analyses of the instructional survey explored the extent to which teachers became more likely to endorse conceptual understanding and productive struggle items at the end of the year when compared to the beginning of the year, as well as the extent to which their responses became closer to those of their coaches. Here, too, we found evidence across all partner teachers with pre- and post-surveys (N=49) that their self-reported practices and beliefs had significantly changed in expected ways after one year of coaching.

Coaching practices. In order to explore coaching practices we engaged in successive waves of analysis of videos and later transcripts of pre- and post-observation coaching conferences. These analyses ranged from exploratory review of the nature of coaching interactions to more formalized scoring of conferences using a rubric we developed and iteratively refined to identify variation in the extent to which coaches and teachers engaged in deep and specific talk that aligned with the three components of the instructional triangle, mathematics content, student thinking and pedagogy. Exploratory analyses of the pre- and post-observation conferences demonstrated the following: 1) growth in the rigor of pre-conference conversations of the course of the year (ES=.64 for the 4-item scale see Figures 4 and 5 for examples of 2 items in the scale) and 2) an association between the growth in pre-conference rigor and individual change in maintenance of cognitive demand.

These analyses also contributed to refinement of our coaching framework to focus on three key coaching practices: deep and specific conversations about the instructional triangle, evidence-based feedback, and goal setting. More formalized scoring of pre- and post-conferences for enactment of the three key coaching practices is underway to examine the extent to which the key coaching practices predict/explain additional changes in teachers' practices and instructional beliefs. In our discussion of emerging findings, we present the results of analyses that track the relationship between coaches' perceptions of teacher responsiveness (i.e., the coaches' judgment of the likelihood of teacher follow-through of ideas discussed in their interaction) that were entered into the Coach Tracker log and teachers' perceptions of the degree to which their coaching (over time) was characterized by the coach pressing them for reasoning about mathematical ideas (press) and/or the extent to which their interactions conveyed specific and clear actions teachers' should enact in their teaching (prescriptiveness).

Context conditions. Alongside the above traditional research, another critical component of our work involves understanding how the coaching model can be implemented in diverse local systems comprised of school and district factors. Perhaps the most important context that varies for coach-teacher pairs is the nature of the interaction between them, which is what we explore in our initial (quantitative) analyses. In addition, descriptive analyses of the *Context Survey* administered to coaches and teachers and thematic analyses of interviews with partner teachers were undertaken to identify other salient dimensions of context variation. These context factors are now being utilized as predictors in models exploring the relationship between coaching practice and teaching practice. Descriptive analyses also informed the formation of improvement teams comprised of groups of coaches grappling with common contextual challenges, which we describe in greater detail in the second example of adaptive integration reported in the next section.

Adaptive Integration in Action

We offer two examples of adaptive integration that highlight a range of implicit to explicit attention to adaptive integration. Adaptive integration work is evident in the micro-level adjustments to the implementation of the practices specified in our coaching framework that our coaches undertake as they work with different teachers. It is also evident in our more explicit attempt to promote experimentation with adjustments to the coaching framework utilizing the methods of improvement research.

Implicit Attention to Adaptive Integration: Coaches Adapt to Teacher Responsiveness

Our exploratory analysis of the year one data from Coach Tracker and the Discussion Process Survey revealed evidence that coaches are making adjustments to their practice based on their diagnosis of teachers' needs. In this sense the coaches' judgments about the type of teacher they are working with (one judged to follow through versus one not likely to follow-through) is one context factor that requires adaptive action on the part of the coach. Specifically, we found that when coaches rated teachers as relatively less responsive to coaching, those same teachers judged coaches to be more prescriptive over time in their interactions. In contrast, when coaches rated teachers as relatively more responsive to coaching, teachers reported that coaches tended to *press* teachers for their mathematical reasoning more over time We examined both status (at cycle 1) and change in teachers' perceptions of two constructs, *prescriptiveness* and *press*. Teachers' self-reports of changes in their coaching interactions, in combination with coaches' judgments about the likelihood of teacher follow-through, suggest that coaches adapted their coaching based on their initial judgments at the beginning of the year.

We explored items on the partner teacher survey in order to develop constructs that fit our conceptual model. One factor we call coaching prescriptiveness included three items that if endorsed by teachers would suggest more specific attention to a prescribed set of teaching practices. 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." The three items hung together and conceptually, endorsement of the items suggests that teachers' perceived coaches as being explicit about specific teaching behaviors they think teachers should enact in their teaching.

A second construct we call coaching press, was composed of two items that measured teachers' perceptions of whether or not they were "pressed" to engage with mathematical ideas or in mathematical reasoning by their coach. 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". Results from the psychometric phase of our measurement model indicate that teacher reports of prescriptiveness and press form reliable estimates of between-teacher differences in status at a particular point in time, but low reliability estimates for between-teacher differences in slopes (or linear change over time), likely due to a combination of a limited number of items per construct, a limited number of time points per teacher and a low ICC. This suggests we might have a hard time identifying predictors influencing change in teachers' self-reports of these two elements of their coaching interaction. Interestingly, correlations between slope estimates for prescriptiveness and press indicate a modest negative correlation (r=-.36), suggesting the two scales develop independently of one another.

Teacher responsiveness was assessed by coaches in each log of an interaction with a partner teacher on the Coach Tracker. Specifically, an item asked them 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) Exhibited behaviors that signaled engagement, yet I am uncertain about their level of follow through; and (5) Exceeded expectations for engagement and I have high confidence they will follow through. To generate scores for 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 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".

We engaged in exploratory analyses to examine evidence for adaptive integration on the part of the coaches in working with their partner-teachers. We sought to understand the extent to which coaching interactions varied in status and in linear growth over the year between the 64 coach-teacher pairs. We hypothesized that if 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 the teacher would follow-through with the ideas they discussed together). Thus, we employed a measurement model with items for both the prescriptive and press scales, nested in three time points over the year, and time points nested in coach-teacher pairs (for a full description of the model, see Appendix A)[4].

We 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 implemented task as a time-varying covariate[5]. We assumed that this was a proxy for the nature of the task being implemented. We thought the nature of the task itself could

---

[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 in our data, i.e., 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 model parsimony.

influence partner-teacher perceptions of the focus on prescriptiveness or press, with higher cognitive demand tasks providing more fruitful opportunities for coaches to press teachers for mathematical reasoning.

Our primary interest in these models was the addition of two independent variables; one derived from coach tracker (teacher responsiveness), the other derived from expert ratings of the pre-conference interactions between coach-teacher pairs (the efficiency of the pre-conference rigor score – the sum of rubric scores divided by the length of the pre-conference). In both cases, we used the scores from cycle 1, our earliest measure, to predict both status and linear change over the course of the year.

Table 1 displays the results of our exploratory analyses. Our exploratory findings do provide evidence that coaches engage in adaptive integration based on their early judgments about their partner-teachers. Take for example, the findings for teacher responsiveness on the prescriptive factor. While there is an association between higher teacher responsiveness and prescriptiveness at cycle 1 (β_{201} in Table 1), teacher responsiveness does not predict status in subsequent models with time centered differently. What is more interesting is the relationship between coaches' judgments of responsiveness and teacher perceptions of prescriptiveness. The less likely coaches deemed teachers to follow-though with ideas discussed in their initial interactions, the higher the slope for changes in coach prescriptiveness over the year (β_{211} in Table 1). Likewise, the variable measuring the efficiency of the pre-conference at cycle 1 had a similar effect and interpretation, only in the opposite direction (β_{212} in Table 1) – the more efficient (i.e., the shorter) the initial pre-conference was, the more likely the partner-teacher perceived the coaching interactions to be more prescriptive over the course of the year.

These findings were complemented by an opposite set of findings for our second construct, press. Here, we find that when coaches initially deemed teachers to be more responsive at cycle 1, their partner teachers reported a higher slope for changes in press (β_{111} in Table 1). The variable measuring efficiency of the pre-conference at cycle 1 is not significant in this model, but it is in the expected direction – that is, less efficient pre-conferences were likely to lead to a positive slope for press. The pattern of findings for the prescriptiveness and press scales tell an interesting story, we compare and contrast these findings, and their implications, in the discussion.

Explicit Attention to Adaptive Integration: Overcoming Contextual Challenges with PDSAs

In the first year of the study, we focused primarily on training coaches to take up our coaching framework and iteratively refining our perspective on the key coaching practices based on the data that coaches submitted capturing their work with their partner teachers. As the coaching framework has coalesced around three key coaching practices and a central coaching routine, the work in our second year has more actively engaged coaches in the adaptive integration of the coaching framework in their diverse local contexts.

In preparation for this adaptive integration work, we collected various data sources to learn about the variations in local contextual conditions that were influencing implementation of the coaching framework. In an end of year one survey on their coaching contexts, we asked coaches to report on the way that school and district context shaped their use of our coaching framework in their first year working with the project. In addition, we gathered information about an important consideration for adaptive integration of the model – the way that schools and districts defined coaching roles – through our ongoing interactions with coaches throughout the year. We briefly describe some of our findings related to variation in coaching roles and the most influential school and district conditions reported by coaches.

Our coaches vary considerably in the way their role as instructional coaches are defined and operationalized. In year one of the project, the majority of coaches (50%) were school-based mathematics coaches working with one or more schools. However, the remaining half of coaches were responsible for mathematics coaching, but had broader job descriptions such as school-based instructional coaches responsible for coaching in both mathematics and literacy or district-based mathematics coaches working with a number of schools and sometimes multiple content areas. While 13 out of 32 coaches worked with a single school, the remaining 19 worked with teachers in 2 or more schools, with nearly half working with 3 or more schools. Two-thirds of our coaches work with elementary grades mathematics teachers, but the remaining third (N=13) work 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 coaches have 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 coaches to find ways to integrate complex practices into already complex job responsibilities.

Overall, 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 coaches reported lack of agreement between the principal's vision of effective mathematics instruction and the vision promoted by our mathematics instructional coaching model. This lack of alignment was also voiced by a sizeable minority of our partner teachers through their responses to 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 teachers and coaches was the beliefs and resistance of other teachers in their schools. While coaches generally found that their partner teachers – those participating in this study – were generally responsive to the style of teaching promoted in the coaching framework, coaches and teachers reported that other teachers' in their schools tended to emphasize "more procedural" or "traditional" teaching methods. This suggests in 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 about teaching.

Responding to our emerging understanding of the contextual challenges faced by our coaches, we actively engaged our coaches in experimentation aimed at promoting our learning about how the model can be adaptively integrated into diverse local contexts. We have trained coaches to use the Plan, Do, Study, Act inquiry routine to guide their experimentation with ways to adapt their use of the model while preserving its integrity to the key coaching practices. Coaches volunteered to join groups organized around contextual challenges. In these groups they are collaboratively planning PDSAs and sharing what they are learning from them. These groups focused on the following conditions that are salient challenges in their coaching work: adapting the coaching model to support beginning teachers; adapting the coaching model to support veteran teachers; managing time to enable intensive coaching interactions; coaching teachers to attend to students' conceptual understanding in more procedurally focused contexts; and adapting the coaching model for use in literacy coaching. In the next section we provide one illustrative example of PDSA work undertaken by a coach in the working in the "procedural contexts" group.

Adapting the coaching framework to procedural contexts. Our coaching framework focuses on coaching practices that are best actualized when utilized in conjunction with the planning and teaching of high cognitive demand mathematics tasks, because these tasks allow for discussions of relationships and connections to underlying mathematical ideas, multiple solution paths, and the use of multiple representations. However, our coaches find themselves working in some school and district contexts where an emphasis is placed on more procedurally focused mathematics activities. While the state has made efforts and continues to encourage districts to use curricula that contain cognitively demanding math tasks, some districts, schools and teachers have not fully made this transition. Coaches facing these contextual challenges are working to engage in adaptive integration that enables them to coach in a way that focuses on supporting teachers to use teaching practices that emphasize our end goal, student conceptual understanding, while working in a context that privileges the use of more procedurally-focused instructional materials.

One project coach was charged with coaching a teacher who felt obligated to use the school's mathematics curriculum comprised primarily of low-level tasks. The curriculum also included detailed, scripted lesson guides, which the teacher understood that she was to follow when teaching. The coach was in a bind because coaching with low-level tasks does not adhere to the TN Instructional Coach Model. In order to move forward in the work with the teacher, the coach first conferred with the principal to surface his expectations about teacher adherence to the curriculum's scripted lessons, and to ask permission to work with the teacher to make some modifications in order to press students to think more conceptually and to share their reasoning related to important mathematical ideas related to the curricular tasks. The principal agreed that students should be pressed to share their reasoning, although he stipulated that deviation from the curriculum's tasks should be limited. Consent to make some modifications to the curriculum was enough to give the coach permission to engage the teacher in planning related to the identification of important mathematics ideas related to tasks that could be found in the teacher's curriculum.

The coach planned a PDSA cycle in order to learn about how they could integrate the coaching framework into more procedural contexts: doing the Coach Teacher Discussion Process routine with a modified, low-level task. The first test focused on engaging the teacher in a discussion of ways to use a low level task from the curriculum but to also insert "questions to prompt for thinking and reasoning" related to the deeper mathematical idea that undergirded the task.". Without these questions, the mathematical relationships and connections would not be surfaced. Through discussion of the way students would likely solve the task and the misconceptions that students might have in solving the procedural task, the coach and teacher still managed to discuss the underlying mathematics in the task. The coach and teachers also generated questions that the teacher could ask students to engage them in discussion related to the mathematics.

As this coach continues to work with this teacher, she is testing new ways to bring a focus on conceptual understanding into discussions that involve tasks that start out as low-level. As the coach engages in disciplined experimentation with the model and shares the results with the network hub, we gain insights into how the model can be used by coaches who are working with teachers whose contexts contain mostly low-level procedural tasks rather than tasks with a high-level of cognitive demand.

In conclusion, we have emerging evidence that a set of promising change ideas that address specific problems of coaching practice will emerge from the coaches' PDSAs. By facilitating routines and structures for sharing the results of individual coach's PDSAs with our coaching network, we hope that these change ideas will be taken up and tested by other coaches under both similar and dissimilar conditions, and that over time the PDSAs will help us to learn and share our learning about how to integrate the coaching model into different kinds of schools and districts throughout the state.

Discussion

Our work is part of a broader trend in education reform and research that aims to utilize research-practice partnerships to facilitate systemic educational improvements. For example, the design-based implementation research approach promotes explicit attention to implementation dynamics throughout the research and development process (Penuel et al, 2011). Similarly, the networked improvement community concept draws on traditions from improvement science to guide the work of a network of practitioners, researchers and designers working collaboratively to address problems of practice (Bryk, Gomez & Grunow, 2011; Bryk, Gomez, Grunow & LeMahieu, 2015; Dolle, Gomez, Russell & Bryk, 2013). Consistent with the core principles underlying DBIR and the networked improvement community concept we have engaged a diverse colleagueship of expertise representing researchers, practitioners, and policymakers, to work together to test a coaching framework that aims to support instructional improvement at scale.

Although it would have been easier for us to focus on a "main effects" story with our Year 1 data, we only presented these findings very briefly as context. Even the "main effects" story is complex because not all teachers began at the same level and therefore, estimates of improvement are limited by the outcome chosen and the opportunities for teaching growth on those outcomes. Nevertheless, our project's aim was not to build a coaching model, that with intensive resources and effort could train 30 coaches a year in order to improve teaching at-scale through coaching. Instead, the project's aim is to envision how we can imagine improvement at-scale when we have limited resources. A key assumption in our thinking is that in order to generate large-scale change with limited resources it will require the expert judgments of on-the-ground professionals.

Given these aims, we chose to shift our focus from a "main effects" story because, ultimately, what we need to learn is how successfully trained coaches employ expert judgments in their coaching when they are confronted with diverse contexts. We have not assumed that our coaches have reached an end point – i.e., that they are "successfully trained" – but the evidence of their growth as well as the growth in teaching we observed from their partner-teachers suggests they are on their way. Instead, we believe we are in the midst of micro and macro iterative measurement and design improvements.

Our work employs improvement science alongside traditional research activities to work in interaction to address complex problems of policy and practice. Our focus on adaptive integration acknowledges that innovative practices must be adapted to local conditions in order to be implemented beyond sites that resemble initial development sites. We have presented findings from our efforts to learn about how coaches have approached their diverse contexts as we have begun to build a measurement system. These findings have included retrospective accounts, perhaps describing some of the micro-decisions coaches make as they confront some of the complexity in their work with teachers. At the same time, we have presented some of our initial takeaways from efforts to train users to employ a measurement system (e.g., through PDSA cycles) to aid the improvement of teaching through improved coaching.

Although just exploratory, our findings paint an interesting picture of coach adaptation in their interactions with different types of teachers. These findings could point toward an implicit 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 be most interested in providing teachers with clear and specific directions about how they can develop their teaching practice in the short-term. These interactions could remain fairly brief and could be pre-cursors 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 deep learning discussions, where coaches press teachers for their mathematical reasoning, with the hope that by doing so, teachers will engage in similar practices with their students.

While these findings are just exploratory and need to be replicated with larger samples of teachers and coaches, we think they are instructive about the complexity of the improvement process. The literature is replete with examples of street-level-bureaucracy, typically used to explain why change in the teaching process seldom happens (e.g., Sarason, 1982; Weatherley & Lipsky, 1977; Weick, 1976), Yet, little has been documented about how instructional coaches successfully manage diverse contexts. When thinking about the complexity of their role there are several barriers to achieve improvements in teaching efficiently. Forget for a minute the fact that many coaches need to frequently manage their relationship with administrators to carve out the maximum amount of time they have to work with teachers (because they are consistently asked to fill other administrative, non-coaching duties). But, even when they work in supportive environments, coaches are faced with issues of efficiency. How many partner-teachers can they reasonably expect to work with? How much time does it take on average to maintain contact with partner-teachers? What is the variance among different partner-teachers? 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 their partner-teachers. Our contention has been that in order to help coaches make progress on these questions we need to engage with them in building a measurement system with the potential to; 1) surface important relationships that would help build theories of teaching development within specific contexts, and 2) help coaches (and teachers) monitor different aspects of their teaching as they work toward improvements in teaching together.

The example of explicit adaptive integration in procedural school contexts that we discussed in the last section is significant in providing new insight into facilitating instructional improvement through rich methods of disciplined inquiry even under difficult implementation conditions. By making small and incremental changes to both coaching and teaching practice, such as getting permission from school leadership to make minor curriculum adjustments and constructing new planning and questioning routines, this coach is providing essential insight into hard to implement contexts that are often the point at which policy interventions fail. By using our many measures of context and practice to understand the environment in which this work is taking place, honoring the ideas and needs of the coach, and carefully documenting the process that facilitates making changes under these conditions, the network is building knowledge about how to increase capacity for change in school environments that, in more traditional policy and implementation studies, would simply be incompatible with the intervention at hand. This example illustrates the powerful potential of studying implementation in a context-specific fashion. Marrying context and practice as objects of study and actively collaborating with practitioners are emerging as key elements of this kind of work.

This partnership affords us opportunities to take up and advance the networked improvement community concept in several ways. Our work is focused on a persistent problem of practice, in keeping with the NIC concept, but we are also grappling with problems of policy, such as how a state can meaningfully craft a coaching initiative that works in 140 different school districts. We have taken this challenge up as an integral part of our collaborative work and this paper aimed to illustrate how we are using our work with a network of coaches to iteratively refine our coaching framework based on the data generated by the framework's use in practice and incorporating framework refinement to develop a model coaching system.

Furthermore, our TDOE partners provide access to a natural vehicle for scaling the innovations that emerge from the network. As a result our project aims to address not only problems of local educational practice but also problems of policy. For example, we actively grapple with questions regarding the actions a state department of education can take in a decentralized educational governance structure that reinforces the authority of districts over key decisions such as the extent to which instructional coaching is a priority, or even a position, and the ways in which coaching roles are structured and enforced.

Finally, our research team's expertise and interests position traditional research as co-equal to practical measurement and prioritizes the pursuit of practical measures that investigate the deep concepts, grounded in our conceptual model, rather than surface level features of implementation. As a result, we simultaneously grapple with ongoing inquiry work driven by coaches that employ more practical, but less validated measures, in combination with the types of measures employed in traditional social science research. We believe, that it is in the synergistic space between these two inquiry activities that we build the capacity for rapid, meaningful improvement at scale.

References

Berwick, D. M. (2008). The science of improvement. *JAMA*, *299*(10), 1182-1184.

Birman, B. F., Desimone, L., Porter, A. C., & Garet, M. S. (2000). Designing professional development that works. *Educational leadership*, *57*(8), 28-33.

Bryk, A. S., Gomez, L. M., & Grunow, A. (2011). Getting ideas into action: Building networked improvement communities in education. In *Frontiers in sociology of education*, (Ed.) Maureen Hallinan. New York, NY: Springer Publishing.

Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. (2015). *Learning to improve*. Cambridge, MA: Harvard Education Press.

Campbell, P. F., & Malkus, N. N. (2011). The impact of elementary mathematics coaches on student achievement. *The Elementary School Journal*, *111*(3), 430-454.

Coburn, C. E., Russell, J. L., Kaufman, J. H., & Stein, M. K. (2012). Supporting sustainability: Teachers' advice networks and ambitious instructional reform. *American Journal of Education, 119*(1), 137-182.

Dolle, J. R., Gomez, L. M., Russell, J. L., & Bryk, A. S. (2013). More than a network: Building professional communities for educational improvement. *National Society for the Study of Education Yearbook*, *112*(2), 443-463.

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

Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. Handbook of educational psychology (pp. 15–46).

Foster, D. & Noyce, P. (2004). The mathematics assessment collaborative: Performance testing to improve instruction. *Phi Delta Kappan, 85*(5), 367-374.

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*(4), 919-963.

Hannan, M., Russell, J. L., Takahashi, S., & Park, S. (2015). Using Improvement Science to Better Support Beginning Teachers The Case of the Building a Teaching Effectiveness Network. *Journal of Teacher Education*, 0022487115602126.

Hiebert, J., Gallimore, R., & Stigler, J. W. (2002). A knowledge base for the teaching profession: What would it look like and how can we get one?.*Educational researcher*, *31*(5), 3-15.

Hiebert, J., & Stigler, J. W. (2004). A world of difference. *Journal of staff development*, *25*(4), 10-15.

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. Retrieved from http://www.carnegiefoundation.org/blog/what-we-need-in-education-is-more-integrity-and-less-fidelity-of-implementation/

Lewis, C. (2015). What is improvement science? Do we need it in education? *Educational Researcher, 44*(1), 54-61.

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

Neergaard, L., Smith, T., Hochberg, E., & Desimone, L. (2011, September). Impact of organizational supports for math instruction on the instructional quality of beginning teachers. Paper presented at the annual conference of the Society for Research on Educational Effectiveness, Washington, DC.

Murray, S., Ma, X., & Mazur, J. (2009). Effects of peer coaching on teachers' collaborative interactions and students' mathematics achievement. *The journal of educational research*, *102*(3), 203-212.

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

Polly, D. (2012). Supporting Mathematics Instruction with an Expert Coaching Model. *Mathematics teacher education and development*, *14*(1), 78-93.

Sandholtz, J. H. (2002). Inservice training or professional development: Contrasting opportunities in a school/university partnership. *Teaching and teacher education*, *18*(7), 815-830.

Sarason, S. B. (1982). *The culture of the school and the problem of change*. Boston: Allyn and Bacon.

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American educational research journal*, *33*(2), 455-488.

Stein, M. K., & Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in a reform mathematics project. *Educational Research and Evaluation*, *2*(1), 50-80.

Stein, M.K., & Kaufman, J.S. (2010). Selecting and supporting the use of mathematics curricula at scale. *American Educational Research Journal, 47*(3), 663-693.

Thompson, C. L., & Zeuli, J. S. (1999). The frame and the tapestry: Standards-based reform and professional development. *Teaching as the learning profession: Handbook of policy and practice*, 341-375.

Weatherley, R., & Lipsky, M. (1977). Street-level bureaucrats and institutional innovation: Implementing special-education reform. *Harvard educational review*, *47*(2), 171-197.

Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative science quarterly*, 1-19.

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*.

Figure 1: Coach-Teacher Discussion Process

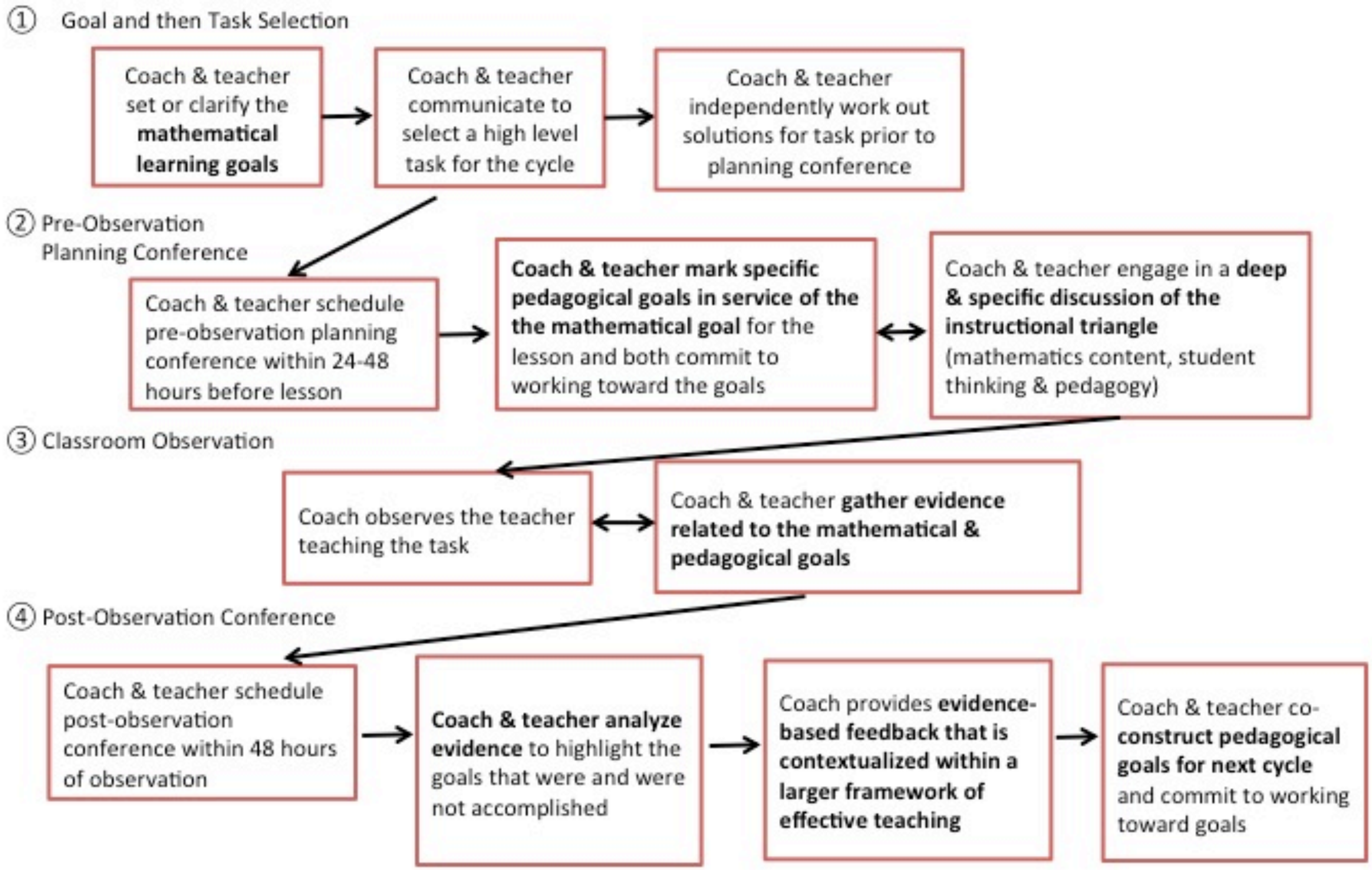


Figure 2: TN Instructional Coaching Model

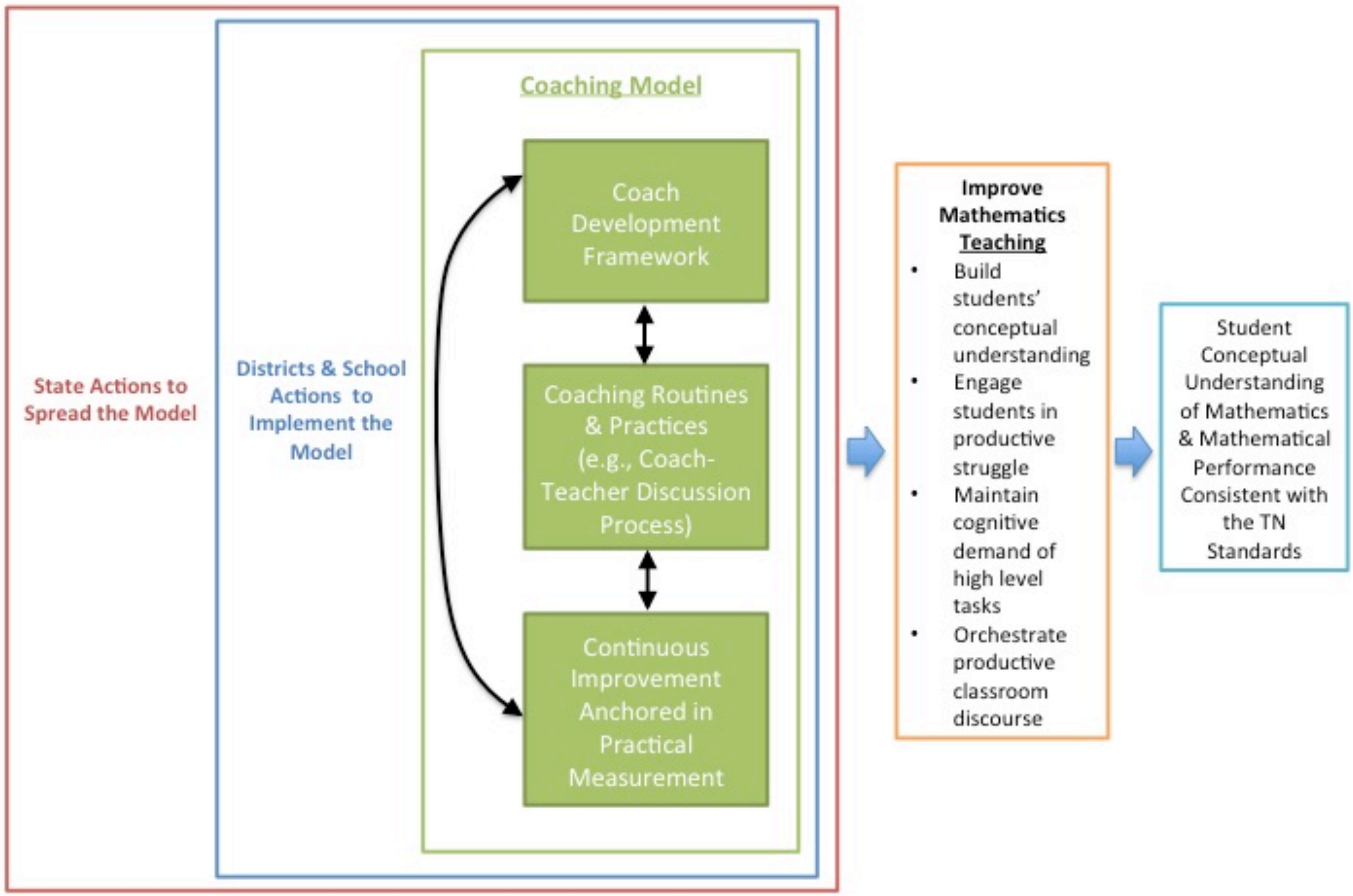
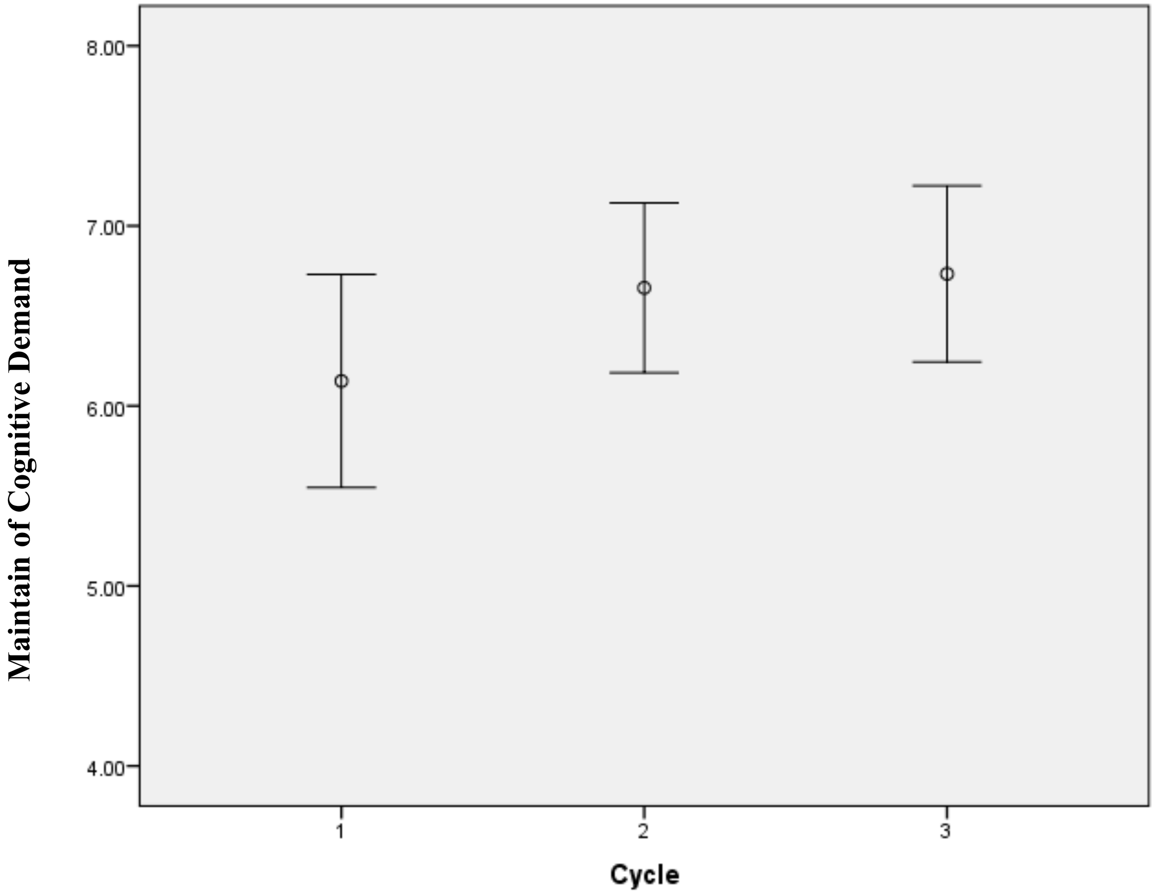
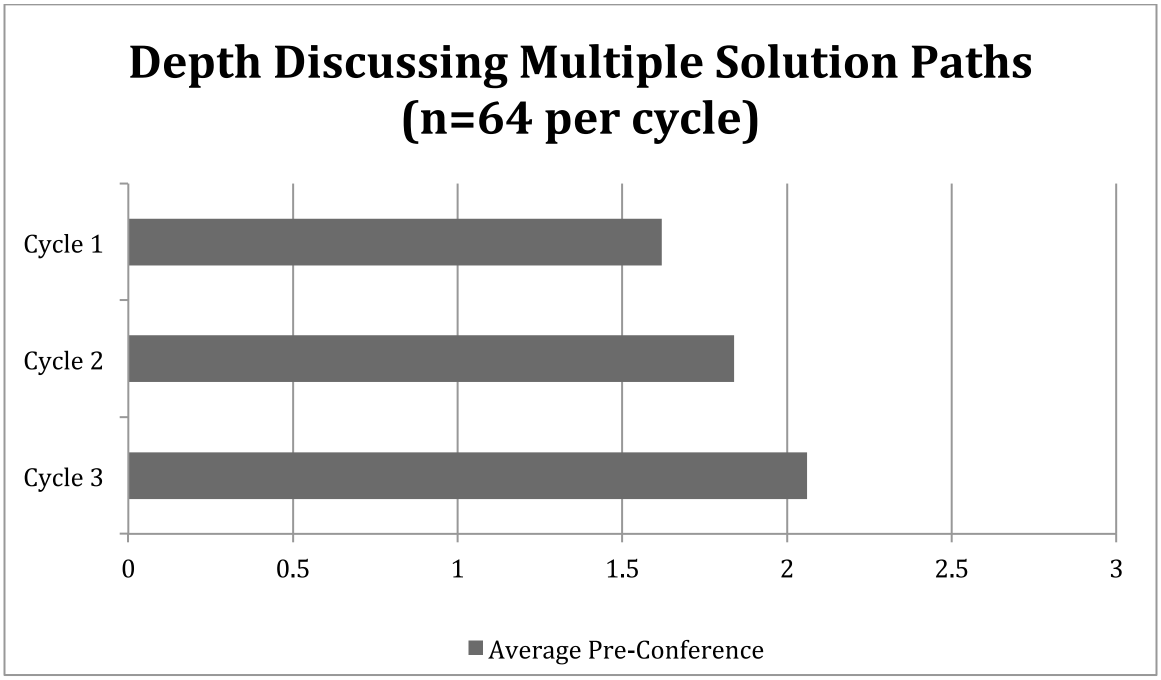


Figure 3. Improvement in teachers' capacity to maintain the cognitive demand of mathematical tasks



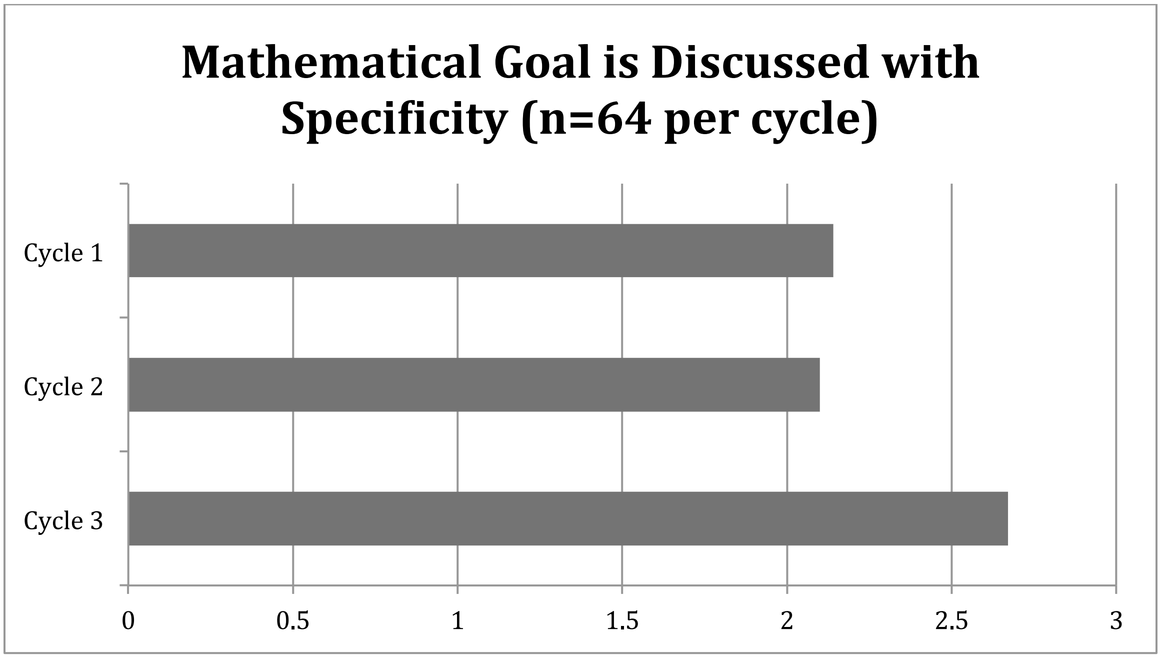
Note: In addition to a mean increase in maintenance of cognitive demand we documented a significant qualitative difference in the degree to which coaches assisted teachers in the classroom in time 3 versus time 1 and 2. In time1, 27 classrooms included coach assistance that significantly improved the teachers' ability to maintain the cognitive demand on the task. At time 2, significant coach assistance was present in 32 teachers' lessons. However, in time 3, there were only 16 cases of significant coach assistance. Together these two pieces of information provide evidence that teachers' capacity to maintain high-level tasks was improving throughout the year.

Figure 4. Improvement in in the quality of coach-teacher pre-observation conference discussions (depth of discussions of mathematical solution paths)



Note: A "1" on the depth scale for discussion of the mathematical solution paths for the task the teacher was planning to teach was assigned when all solution paths were discussed in superficial ways. A "2" was assigned when one solution path was discussed in depth and a "3" was assigned when at least two solution paths were discussed in depth. The change in mean score for the depth of the discussion of mathematical solution paths was equivalent to 0.62 SD.

Figure 5. Improvement in in the quality of coach-teacher pre-observation conference discussions (depth and specificity of discussion of the mathematical goals for the lesson)



Note: This figure displays the change in means over time in the scores assigned to the specificity of discussions of the mathematical goal of the lesson discussed in a coach-teacher pre-observation planning conference. A "1" on the specificity scale for discussions of mathematical goals was assigned when only broad mathematical topics were named. A score of "2" was assigned when a mathematical definition or procedure was named as a goal. A "3" was assigned to discussions of how students acquire the underlying meaning of a mathematical concept. The change in mean score for the specificity of the discussion of mathematical goals from time 1 to time 3 was equivalent to 0.75 SD. Notably, a focus of a face-to-face network meeting with coaches that launched Cycle 3 was discussing the mathematical goal of a lesson with specificity in their pre-conferences with their partner-teachers.

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

	Partner Teacher Report of Press for Mathematical Ideas/ Reasoning (Learning through Discussion)						***Partner Teacher Report of Specificity with Clarity (Coach Prescriptiveness)***					
	Coeff.		**s.e.**	**Coeff.**		**s.e.**	**Coeff.**		**s.e.**	**Coeff.**		**s.e.**
At First Cycle												
Average ($\beta_{100;}\beta_{200}$)	.02		.19	.02		.20	-.09		.14	-.09		.12
Tch. Resp. ($\beta_{101;}\beta_{201}$)‡	-.06		.27	-.06		.28	.34	*	.16	.34	*	.16
PC Effcncy ($\beta_{102;}\beta_{202}$)‡‡				.14		.54				-.84	*	.31
Growth Slope												
Average ($\beta_{110;}\beta_{210}$)	-.04		.21	-.06		.18	.10		.17	.11		.16
Tch. Resp. ($\beta_{111;}\beta_{211}$) ‡	.48	~	.28	.48	~	.28	-.56	*	.21	-.56	*	.20
PC Effcncy ($\beta_{112;}\beta_{212}$) ‡‡				-.54		.59				1.14	*	.51
Task Length (As Implemented)	.01	*	.00	.01	*	.00	.00		.00	.00	*	.00

‡ Coaches' reports of the likelihood teacher follows through on the reported interaction in coach tracker during all reported interactions during the first cycle.

‡‡ Efficiency of the pre-conference rigor score. The sum of items on the pre-conference rubric divided by the length of time in pre-conference. More efficient ratings tended to be short pre-conferences.

Appendix A

Before running the models item sub-scores from the partner-teacher survey were standardized to have a mean of zero and standard deviation of one (see, e.g., Raudenbush, Rowan, & Kang, 1993). The psychometric phase of these models also allowed us to examine differences in teacher perceptions of growth on each of these 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 described below:

$$\text{Item Score}_{mij} = \psi_{1ij}*(Press_{mij}) + \psi_{2ij}*(Prescriptive_{mij}) + \varepsilon_{mij} \quad \text{[equation 2.1]}$$

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; $Prescriptive_{mij}$ is a dummy indicator demarcating three items for the scale Prescriptive; ψ_{1ij} is the average score on the Prescriptive scale at time i for teacher-coach pair j; 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:

$$\psi_{1ij} = \pi_{10j} + \psi_{11j}*(Time) + \psi_{1pj}*(A_{pi}) + e_{1ij}$$
$$\psi_{2ij} = \pi_{20j} + \pi_{21j}*(Time) + \psi_{1pj}*(A_{pi}) + e_{2ij} \quad \text{[equation 2.2]}$$

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; ψ_{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; ψ_{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 Prescriptive 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; ψ_{21j} is the growth slope, in years, for the Prescriptive scale; A_{pi} is a set of (p) time-varying covariates measured at each time point; ψ_{2pj} is the effect of time-varying covariates on the Prescriptive 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:

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

[equation 2.3]

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 *Prescriptive* 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 *Prescriptive* 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 *Prescriptive* 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

Prescriptive scale over the year; r_{21j} is residual error normally distributed with a mean of 0 and a standard deviation of unity.