Instructional Science
https://doi.org/10.1007/s11251-021-09536-7

ORIGINAL RESEARCH

# Coaching that supports teachers' learning to enact ambitious instruction

**Eben B. Witherspoon[1] · Nathaniel B. Ferrer[1] · Richard R. Correnti[1] · Mary Kay Stein[1] · Christian D. Schunn[1]**

Received: 1 April 2020 / Accepted: 12 February 2021


**Abstract**
Teacher learning is a huge challenge in instructional change, but relatively little work has carefully examined the mechanisms by which teachers learn, in contrast to the extensive work on programs that help teachers learn and the high-leverage instructional practices that are strong predictors of student learning. Specifically, relatively little is known about how teachers learn to effectively implement these new instructional practices. Using a mixed-methods, case-comparison design, this study examines specific instructional coaching practices that support 4th–8th grade mathematics teachers in learning to implement ambitious instructional practices. The study leverages a large, state-wide representative dataset in order to purposively select carefully-matched contrasting cases for qualitative analysis from a starting sample of hundreds of teachers, which enabled selecting teachers that began in a very similar place but then progressed at different rates. In-depth qualitative coding was systematically conducted on detailed transcripts of coach-teacher conversations from these carefully selected cases. Finally, these codes were analyzed quantitatively to determine whether the content and form of these conversations predicted improvement in teachers' instructional practices. Results showed that coach-teacher pairs who discuss when and why certain practices should be implemented, and provide more opportunities for teacher input, see larger gains in ambitious instruction in later lessons. Implications for a coaching model based in the cognitive sciences are discussed.

**Keywords** Teacher learning · Coaching · Ambitious instruction · Cognition · Adaptive expertise

## Introduction

There is broad consensus among educational researchers that instructional quality is one of the strongest predictors of student achievement (Aaronson et al. 2007). Further, research on teachers' instructional quality over the past two decades has emphasized the importance of "ambitious instruction", a set of teaching practices broadly characterized

✉ Eben B. Witherspoon
eben.witherspoon@pitt.edu; ebw13@pitt.edu

Extended author information available on the last page of the article

by a focus on students' deep conceptual understanding of content, responsivity to student thinking, and the active involvement of students in the co-construction of knowledge (Lampert 2001; Smith et al. 2001). There is also a growing evidence base for more specific instances of such general practices within disciplines like mathematics (Stein et al. 2008) and literacy (Wray et al. 2000) that have proven to be reliable indicators of instructional quality. Unfortunately, the literature has also consistently found that both novice and in-service teachers often struggle with the implementation of these practices, resulting in many continuing to use traditional "transmission" models of instruction (Wells and Arauz 2006; Kisa and Correnti, 2015; Stroupe and Windschitl 2015).

Professional development workshops, professional learning communities, video clubs, and instructional coaching have all been suggested as potentially powerful interventions that may influence teacher learning to implement such practices (Borko 2004; Sherin and Han 2004; Vescio et al. 2008; Knight 2009). Indeed, recent meta-analyses have shown that such interventions can meaningfully influence teachers' classroom practice. But there is also both large variation in the features of these interventions and the relative strengths of their effects on teachers' practice (Garrett et al. 2019). Even within the context of the large intervention category of instructional coaching, specific structures and protocols can vary widely, which limits knowledge about what components actually matter, and further limits the effect that instructional coaching interventions have on teacher practice (Kraft et al. 2018).

Limits on what is known about effective instructional coaching also stem from the relative lack of attention to mechanisms of teacher learning as the primary focus of inquiry within research studies (Goldsmith et al. 2014). Many studies identify and measure teachers' use of instructional practices that produce desirable student learning outcomes, but relatively few studies relate these findings to different approaches to teacher education in terms of a theory of teacher learning (see Lampert et al. 2013). One approach to bridging these research efforts is to consider teacher learning experiences as important processes which can be both theoretically and empirically related to the development of certain kinds of teacher knowledge. While such an effort does not describe a complete learning mechanism, it can begin to point towards components of a learning mechanisms that lead to potentially desirable outcomes in terms of teacher knowledge and practices. For example, gaining a better understanding of the theoretical underpinnings of features of instructional coaching will help to direct future efforts to re-design and scale specific characteristics of coaching that are known to improve teachers' ability to implement ambitious instruction in complex classroom environments.

Having a useful theory of teacher learning requires careful conceptualization of what is difficult in ambitious instruction. We propose that effective implementation of ambitious instruction requires that teachers not only have sufficient knowledge of specific high-leverage instructional practices, but also the ability to apply that knowledge across a number of complex and varying situations in the classroom (e.g., adapting to the highly varied student ideas that are voiced during rich classroom conversations or across different rich tasks). There are a number of theories from the cognitive sciences regarding how learners adaptively apply skills learned in one context in novel scenarios (see Hatano and Inagaki 1986; Schwartz et al. 2005), but these theories have not been applied to teacher education programs. Therefore, this study aims to contribute to a theory of teacher learning by identifying components of instructional coaching programs that help teachers learn to adaptively implement high-leverage, ambitious instructional practices in the classroom.

## Theoretical background

### Challenges to implementing ambitious, student-centered instruction

The centrality of student ideas in lesson planning and implementation is considered to be a core component in many models of ambitious instruction (Windschitl et al. 2012). This emphasis on the importance of responsivity to students and student thinking is rooted in early work by theorists such as Vygotsky (1978) and Bakhtin (1981) who took a constructivist stance, suggesting that learning is a social act that occurs through the co-construction of understanding and meaning during dialogic interactions. Facilitating this philosophy of teaching and learning often requires a significant shift in teachers' instructional approach, away from more traditional "direct instruction" models centered around the teacher transmitting knowledge to students, and towards a "student-centered" learning environment where knowledge is generated by and with students. Student-centered classrooms include a number of features such as facilitation of group dialogue, tasks that provide opportunities for students to challenge and build their own understandings, and the introduction of formal domain knowledge contingent upon the progress of the learner (Richardson 2003).

Requiring in-the-moment adaptation to live dialogue between multiple actors can understandably present unique challenges to implementation for both novice and experienced teachers (Wells and Arauz 2006). Classroom enactment of a lesson that incorporates in-the-moment generation of student thinking is particularly likely to require that teachers apply instructional practices in extremely variable environments. Thus it is not surprising that studies find wide variation in the implementation of these ambitious reform practices in the classroom, despite over a decade of policy-level pushes towards incorporating these ambitious instructional practices in the United States (see Common Core State Standards Initiative 2010; Kisa and Correnti 2015). For example, in one study of teachers learning to implement ambitious instruction via a common professional development program, only some teachers incorporated student ideas as elements of the lesson design, while others simply identified student ideas as misunderstandings to be corrected (Stroupe and Windschitl 2015). This variation in outcomes can lead to conflicting empirical reports on constructivist instruction, which in part have contributed to a prolonged debate over the benefits of instructional reforms based in constructivist theory over direct instruction (Kirschner et al. 2006; Kuhn 2007; Tobias and Duffy 2009). As a way forward in this debate, some researchers have suggested that it will be essential to better understand the different ways teachers learn these practices, in order to help reduce variation in implementation and student outcomes (Reznitskaya and Gregory 2013).

### Developing adaptive vs. routine expertise for teachers

One potential explanation for differences in the implementation of ambitious instruction, particularly for teachers with similar experience and who engaged in similar professional development opportunities, could be understood as the development of either routine or adaptive expertise (Hatano and Inagaki 1986). Routine expertise is characterized by the application of a set of skills that are enacted semi-autonomously, with relatively little understanding by the actor about when or why a particular solution is applied. This can lead to misapplication of that skill, particularly in situations like teaching, where there is a nearly infinite scope of potential scenarios in which a skill may or may not be applied (e.g., a teacher asking for alternative ideas when no idea has yet emerged). Adaptive expertise, however, includes as part of the learning process the development of an understanding of the *conditions* under which a particular situation is or is not appropriate (Smith et al. 1997). Understanding the necessary and sufficient conditions under which a certain solution will work allows the adaptive expert to recognize when those conditions have changed and then adapt a previously learned procedure to produce a new solution to account for those changes. Interestingly, while adaptive expertise in mathematics has been examined widely in the context of students as learners (Verschaffel et al. 2009), and others have documented adaptive expertise in math and science teaching (Anthony et al. 2015; Bowers et al. 2020), fewer studies apply cognitive theory to understanding how this kind of expertise is developed in teachers as learners (e.g., through different teacher professional development models).

The conditionality required for adaptive expertise in teaching highlights a tension in the literature around how teachers' pedagogical knowledge is conceptualized. For example, while some treat instructional strategies and awareness of possible student solutions as static knowledge acquired and located within an individual teacher (Isiksal and Cakiroglu 2011), others view them as socially situated and constructed within the context of the classroom (Depaepe et al. 2013; Rowland and Ruthven 2011). Instead, recent studies have suggested an approach to professional development that unites these two perspectives (Kaiser et al. 2017; van der Linden and McKenney 2020). It may not be enough for teachers to simply acquire knowledge about how to perform a list of high-leverage practices; they also need to understand *when* these practices can and should be enacted based on the particular task, students and goal of the lesson, which will require some understanding of *why* those specific practices are best implemented at that particular time. Learning such skills only through repetitive practice over many years is not only inefficient, but often fails to ever develop this kind of expertise; heuristics learned in this way have been shown to often be misapplied in novel and complex scenarios (Frensch and Sternberg 1989).

## Cognitive mechanisms for teacher learning: attending to *when* and *why*

Cognitive research suggests that adaptive expertise can be developed through attending to two key outcomes during learning: creating well-organized knowledge structures, and building metacognitive skills (Smith et al. 1997). In the context of teacher learning, attending to multiple conditions under which a teacher practice could be applicable during professional development may offer one mechanism for creating the well-organized knowledge structures important in developing adaptive expertise in ambitious instruction. For example, student learning goals provide a metric through which to evaluate how a particular lesson's activities connect to student thinking, so that activities can be adapted and revised to better align to the big conceptual target of the lesson (Hiebert et al. 2007; Stein and Meikle 2017). Explicitly relating the concrete particulars of the activities students engage in during a lesson to broader conceptual learning goals could help teachers to develop the ability to generalize procedural knowledge that characterizes adaptive expertise, by moving between concrete representations of a problem to more abstract generalizations of that problem (Goldstone and Son 2005). For example, focusing on the specific features of mathematics tasks (e.g., finding the circumference of a wheel to determine the distance a vehicle will travel), and then noting how those pertain to a more general mathematics learning goal (e.g., proportional reasoning), could help a teacher to recognize and identify that underlying concept within new tasks and student contributions during future lessons. Attention to a combination of these conditions during teacher learning could therefore help teachers to better apprehend *when* certain practices may be an appropriate intervention even in novel, in-the-moment teaching situations.

Another critical component of adaptive expertise is that it requires strong executive function and self-regulation skills; the learner must monitor and select from a number of strategies, and revise that selection if it is found to be not applicable in that scenario (Butterfield and Nelson 1989; Salomon and Perkins 1989). Therefore, learning experiences that could lead to adaptive expertise may be more successful when they include these opportunities for metacognitive reflection, such as providing more opportunities for the teacher to independently incorporate their own thinking, or reflect on their decisions when selecting appropriate high-leverage practices (Smith et al. 1997). That is, when teachers are prompted to be metacognitive about their pedagogical decisions, they may develop as part of their instructional schema an understanding of *why* they might select one high-leverage practice over another in any given teaching situation. Understanding the extent to which professional development programs provide teachers with these opportunities to struggle with important pedagogical decisions and allow them to be reflective on their practice could provide important insight into how some professional development opportunities lead to the development of teachers' adaptive expertise, while others do not.

However, few current teacher education programs are structured to explicitly attend to these elements during learning experiences, leaving the development of these high-leverage practices up to chance (Darling-Hammond et al. 2005). For many in-service teachers, this means that adaptive expertise either develops slowly over the course of their career (i.e., across multiple lessons and cohorts of students over many years), or they may never develop this form of expertise. Therefore, one way to better understand how teacher education programs can intentionally develop adaptive expertise in their teachers is not simply to identify a set of high leverage practices to enact or kinds of tools that are commonly used with teachers to help them enact these high leverage practices (such as coaching or lesson study), but also attend to the specific mechanisms through which professional development models can develop teachers' understanding of *when* and *why* they should enact certain high leverage practices.

## The case for instructional coaching

Instructional coaching provides one promising model for efficiently developing teachers' expertise in implementing ambitious instruction. Borrowing from Clarke and Hollingsworth's (2002) interconnected model of teacher professional growth, instructional coaching can be considered a "change sequence" through which teacher learning may occur. In this model, the coach represents an external source of new information introduced into the teacher's professional environment. Interactions between the coach and teacher encompass both enaction (as participation in the instructional coaching cycle) and reflection (as intentional, coach-supported consideration of practice), processes that are hypothesized to be the core mechanisms of teacher change (Clarke and Hollingsworth 2002). Instructional coaching has already been found to incorporate five key aspects of high-quality professional development: a focus on content, active learning, policy coherence, and participation that is sustained and collective (Desimone and Pak 2017). But these foundational elements do not distinguish between models of instructional coaching that would produce routine versus adaptive expertise.

One feature of instructional coaching that may be integral to the development of adaptive expertise for teachers is the extent to which coaches navigate tensions between prescriptive interactions with teachers, and interactions that are more responsive to reflection of the individual teacher on their context (Ippolito 2010; Correnti et al. 2020). These micro-interactional “stances” that coaches take during moment-to-moment interactions with teachers reflect different epistemological perspectives towards teacher learning; one focused on teachers’ acquisition of an established knowledge base of effective teaching practices, and another emphasizing the development of teachers’ tacit awareness of practices that could be effective in response to a particular instructional context (van der Linden and McKenney 2020). Importantly, coaching that encourages teachers to reflect in a way that balances or unites these two perspectives (e.g., both eliciting teacher input on their context and offering expertise) over time could indicate a macro-interactional “orientation” towards coaching that attends to growth in a teacher’s pedagogical knowledge base, while providing the teacher with greater opportunities for the kinds of metacognitive reflection that are related to the development of adaptive expertise (Haneda et al. 2019).

## Study context

The current study was part of a larger project examining the impact of instructional coaching on the practice of 4th–8th grade mathematics teachers throughout a multi-year, statewide coaching project in collaboration with the Tennessee Department of Education. The work of the larger coaching project had two primary goals. First, building on the Content-Focused-Coaching model from the Institute for Learning, a revised model for mathematics instructional coaching, was designed, tested, and iteratively refined to support the transition to teaching that is aligned with rigorous, college-and-career ready mathematics standards. Second, a network of 32 highly-trained coaches were developed throughout the state; these coaches were selected from a pool of 62 applicants through a competitive process. The selected coaches varied in prior experience, district context (e.g., urban, suburban, and rural), and focus (e.g., school-based versus district-based). Coaches were trained across three two-day face-to-face sessions per year with monthly webinars for discussion and reflection in between. Between meetings, coaches were asked to apply what they learned by conducting formal coaching sessions with two partnering teachers, using a Coaching Cycle which included four main stages: Goal and Task Selection, Pre-Observation Conference, Lesson Observation and Post-Observation Conference (see Fig. 1 for an overview, see Appendix A, Fig. 6 for a more detailed version of the model.)

Coaches completed this full Coaching Cycle with each of their partner teachers three times in year one of data collection (2014–15), and twice in year two of data collection (2015–16). Each cycle was documented through audio recordings of pre-lesson planning conferences, videotapes of observed lessons, and audio of post-observation feedback conferences. Within the specific coaching model that was used in our dataset, a strong emphasis was on pre-lesson meetings. Teacher planning (e.g., setting clear goals for student learning; anticipating how students will respond to tasks that embody those goals) is viewed as an essential component of teaching (Stein et al. 2008); as such, coaches’ assistance of teachers’ planning becomes a critical site for coaching efforts. Therefore, while opportunities for teacher learning can certainly exist at each stage in the cycle shown in Fig. 1 (e.g., learning to select a task with the appropriate level of cognitive demand, reflection on prior lesson enactment), we focused on the Pre-Observation Conference as a point where we

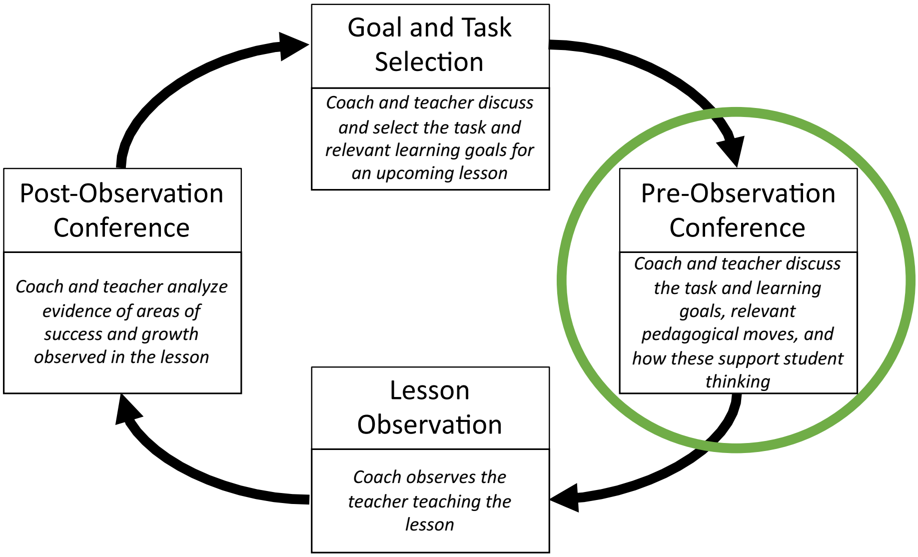



**Fig. 1** An overview of the Coaching Cycle (the circle indicates the focus of the analyses shown in the current study)

anticipated teachers would have especially strong opportunities to learn to enact ambitious instruction on their own.

## Prior measurement development

Each teacher's lessons (one pre-coaching, three coached lessons, one post-coaching) were scored on a 2 to 8 scale using an operationalization of instructional quality focused on the dimension of maintenance of cognitive demand and attention to student thinking, which were emphasized in the coaching model and previously found to be predictive of growth in student achievement. Videos of these lessons were coded by a set of seven mathematics education experts, primarily assistant professors in universities who were trained to utilize the scoring rubric and subsequently scored all classroom videos. Following procedures described in Stein and Kaufman (2010), coders used a rubric to calculate a score for the maintenance of cognitive demand throughout a lesson from the task-as-written to task-as-setup, and from task-as-setup to task-as-enacted (scale 1 to 4). Additionally, raters scored on a rubric the degree to which teachers explored and facilitated the public display of student thinking throughout the lesson (scale from 1 to 4). A mean of the two scales yielded a composite Instructional Quality score on a scale from 2 to 8, with higher scores indicating higher instructional quality. For the overall measure, the intra-class correlation (ICC) was 0.62 (Russell et al. 2020), indicating adequate inter-rater reliability.

Hierarchical linear growth modeling was applied to the Instructional Quality scale data to statistically generate a single estimate of teacher overall growth in Instructional Quality. In these models, time points are nested within subjects in order to better quantify patterns of within-subject change over time. Specifically, rubric scores from each of the scored lessons at every timepoint (up to seven for teachers who participated in year one and year two) were used to produce an estimate for each teachers' overall growth in Instructional

Quality over time, while including a number of covariate adjustments for potential exogenous effects (i.e., the rubric scorer, observed coach assistance). In contrast to simple pre-post approaches to quantifying growth, this approach uses all data points and therefore produces a more robust measure of teacher growth. Tests of model fit using a $c^2$ test of deviance statistics showed that a cubic model was the best fit to the functional form of teachers' improvement (see Russell et al. 2020, for all statistical adjustments used to derive these estimates, and additional details on the cubic growth models). In the full sample, the estimates of teachers' scores in Instructional Quality showed a range of scores at both the initial pre-coaching lesson observation ($M = 5.3$, $SD = 1.8$) and final post-coaching lesson observation ($M = 6.9$, $SD = 0.5$), as well as high variation in growth estimates across the seven time points ($M = 1.6$, $SD = 1.5$).

### The current study

The current study applies a mixed-methods design to understand the particular characteristics of coach-teacher interactions that more commonly occurred with teachers that showed gains in their ability to conduct ambitious instruction in elementary mathematics classrooms. First, we systematically conducted in-depth qualitative coding on detailed transcripts of coach-teacher conversations, to capture the character of coach-teacher interactions around their practice. This allowed us to understand the particular aspects of an instructional coaching model that could produce differences in teacher learning of the adaptive implementation of ambitious instructional practices.

Second, we analyzed these codes using a case comparison design and quantitative analyses to understand the character of interactions that were more likely to be observed among teachers who demonstrated growth on the measure of Instructional Quality described above, which provides a test of the specific aspects of coaching interactions that were theoretically predicted to be associated with teacher learning. Therefore, contrasting high-growth vs. low-growth coach-teacher pairs in Instructional Quality, our main research questions examine: *are there differences in coach-teacher interactions with regard to*:

RQ1: the focus on what the teacher will do (actions), when they would do it (lesson conditions) and/or why they would do it (reasoning).
RQ1b: the inclusion of combinations of conditions (i.e., the specifics of the task, the general learning goal, and/or student thinking).
RQ2: an approach that is more directive (telling information), more reflective (prompting open-ended questions), or a balance of the two.

## Methods

### Sample

The full dataset from the larger project included 32 coaches and 105 partner teachers: 40 partner teachers participated in year one only, 41 participated in year two only, and 24 participated in both years one and two. Situating this study within the context of this larger effort allowed us to select contrasting cases for qualitative analysis from a starting sample of hundreds of teachers, enabling us to select carefully-matched teachers that began in a very similar place but then progressed at different rates. In this way, four teachers within similar baseline scores on our outcome measure were selected for analysis: two teachers showing high growth and two teachers showing low growth (see Case Selection for details). These four individual teachers interacted with four individual coaches (each teacher always with the same coach), all of whom had never worked together before. These steps in the study design was important to control for the impact on teacher learning that could result from a "rich get richer" phenomenon, including from potential variation in prior experiences and coach-teacher relationships (Horn and Kane 2015).

### Case selection

For selection, we focused on the teachers who: (1) started near the middle of the distribution Instructional Quality at baseline; and (2) participated in both years of the coaching cycle. Some coaches worked with teachers that began working on the basics of task selection, and thus were initially focusing on different content in the coaching process than those who had already mastered task selection and were now focusing on classroom enactment. Conversely, some coaches worked with teachers whose scores on our Instructional Quality scale were already at the ceiling and had little room to grow. Luckily the distribution was roughly normal, such that most teachers ($N=63$) began near the mid-point of the Instructional Quality scale ($M=5.3$, SD = 0.1). Of those 63 teachers, only the 24 who had also completed two full years of data collection were considered for selection because the larger amount of data on their teaching creates greater confidence that their teaching had substantially improved.

These 24 teachers on average showed growth in the Instructional Quality scale over time ($M=1.6$), speaking to the value of the coaching model (Russell et al. 2020). However, there was also variation in growth (SD = 0.4), with some teachers making only modest improvements while other teachers made strong improvement. From this subset of 24 teachers, we randomly selected two teachers from the top third of the sample in terms of growth on the Instructional Quality scale ($M=2.0$, SD = 0.3), and another two teachers from the bottom third in terms of growth ($M=1.2$, SD = 0.4), to create our high-growth vs. low-growth four-pair case comparison sample for qualitative coding. This resulted in a final analytic sample of four unique coach-teacher pairs, consisting of four individual teachers who each engaged with a different coach across both years of the coaching model.

Finally, in order to increase the likelihood of identifying richer conversations about practice rather than more rudimentary start-up aspects of the coaching, we ignored the very first round of coach-teacher pre- and post-conferences which tended to be more introductory (i.e., Coaching Cycle A) and instead focused on the second and third rounds of conversations (i.e., Coaching Cycle B and C) where much of the growth in Instructional Quality occurred, leaving us with a total sample of 16 transcripts across the four coach-teacher pairs for coding.

### Measures

#### Coach-teacher interaction codes

Measures of the quality of coach-teacher interactions during pre-conferences were operationalized by a collection of codes describing the particular content and character of the coach-teacher conversations. Codes were applied to transcripts that had been transcribed and divided by a transcription service into individual "utterances" (e.g., something spoken by either the teacher or the coach). Overall, these codes were grouped as measures of the Content of the coach-teacher interactions, or measures of the Form of the coach-teacher interactions. For each measure, Cohen's kappa and raw percent agreement are reported as a measure of reliability.

**Content of interactions** Measures of the content of coach-teacher interactions were operationalized through codes from both coach and teacher utterances. First, the Prompts provided by the coach to elicit additional information from the teacher were coded as either prompts about Actions, or about Reasoning (see Table 1 for details). Actions were primarily focused on the "What" of instruction. That is, it emphasized the pedagogical actions or moves that teachers would perform during the lesson, or a discussion of how to perform these instructional moves (e.g., "So just tell me a little bit about what you're planning on doing with those first two problems or so"). Reasoning, on the other hand, was an indicator that the coach was pressing the teacher to make explicit their rationale for choosing a particular pedagogical move; that is, consider the "Why" of their instruction decisions (e.g., "And why is it that we think it's important for kids to repeat information?"). Finally, coach-teacher interactions were also coded for Conditions, which could be mentioned by either the teacher or the coach. Conditions described the specific context within which the teacher was considering both the particular pedagogical Action they were choosing, and in which they were providing their reasoning; that is, these codes indicated that the coach and teacher were considering "When" a particular pedagogical choice was or was not appropriate (e.g., "So if nobody in you class mentions this answer today, how are we gonna bring that up?"). Conditions were further sub-coded to identify whether the teacher and coach had raised for consideration as context the Goals of the lesson, the Task being used during the lesson, or Student Thinking during the lesson.

**Form of interactions** Codes measuring the Form of the coach-teacher interactions measured the extent to which the coach's questioning during their sessions allowed opportunities for teacher independent input. In particular, coaches asked questions that may have taken an Open-ended form, which allowed for the most independence in the teacher's input, questions that had a form similar to a multiple choice, in which the teacher's answer was limited to a set of Options, or they simply provided the teacher with information by Telling, the form which provided the least amount of opportunity for teacher input. These codes provided an indicator of whether the coaches questioning style adopted a more responsive (e.g., Open) or more directive (e.g., Tell) Orientation during coach-teacher interactions.

**Segmentation** Each transcript was then divided into topical segments: a new segment was created if the Content code (i.e., Actions, Reasoning or Conditions) of the interaction shifted ($\kappa = 0.85$, 93% agreement), and thus each segment had one content code. Collapsing by segment allowed us to analyze the topical co-occurrence of various aspects of the interaction that were included within a single discussion segment but could not have been analyzed at the unit of individual transcript interactions. For example, within a single segment of transcript where the coach was asking the teacher about their Reasoning, this allowed us to also identify what Conditions, if any, that the coach and teacher also mentioned during that segment of the interaction. This reduced 2,005 teacher-coach utterances into 237 segments across the eight transcripts.

**Table 1** Possible sub-codes for each code applied to the content and form of teacher and coach utterances, along with inter-rater reliability (Cohen's Kappa, % agreement), sub-code definitions, and transcript examples (with ***bold-italics*** added to highlight text leading to the selected code)

| | Code | Actor | Sub-code | Definition | Example |
|---|---|---|---|---|---|
| Content of Interaction | Question prompts ($\kappa = .68$, 93%) | Coach | Actions | Prompts teacher to describe *what* move they might do at a particular moment in the lesson | "And when you are having discussions in the classroom, ***what are some moves that you are making*** to engage kids and being purposeful about the way you engage kids?" |
| | | | Reasoning | Prompts teacher to provide rationale for *why* they made a particular pedagogical decision | "So ***why would it have to be that?*** What do you want kids to understand about that? ***Why do the groups have to be what you're saying?"*** |
| | Lesson conditions ($\kappa = .79$, 86%) | Teacher or Coach | Goal | References the teachers' goals for that lesson, either for students or themselves | "So, what are ***the conceptual ideas that you're trying to get kids to learn*** from this task?" |
| | | | Task | References elements of the task, or aspects of the task that could occur within any lesson | "It's a problem about eggs…***I mean, like the counters look like they could represent A.*** So I mean, you can see that." |
| | | | Student Thinking | References thinking of specific students, or that is typical for students of that age group | "But I do think that ***kids naturally will see three out of four, three out of four."*** |
| Form of Interaction | Coaching orientation ($\kappa = .89$, 95%) | Coach | Open | The form of the coach question allows for all possible teacher responses | "So when we look at it here, ***why does it work that this shows three fourths?"*** |
| | | | Options | The form of the coach question limits teachers' response to set of possible answers | "So ***what fraction would represent this part"***? |
| | | | Tell | The form of the coach question does not require or prompt a teacher response | "The amount didn't change. ***That would be a big thing that I'd be listening for kids to say***." |

### Analytic methods

Using the final summarized set of codes collapsed across segments separately for each of the coach-teacher pairs in our case comparison sample, we conducted a series of logistic regression analyses to identify and test whether there were significant differences in the patterns of Content codes identified in transcripts of the coaching conversations with teachers who showed the most growth, and with those who showed the least growth, along the scale of our primary Instructional Quality dependent variable (i.e., maintenance of cognitive demand and student opportunities to engage in conceptual thinking.) The effect sizes of these differences are reported as odd ratios (OR); an odds ratio = 1 is interpreted as the code was just as likely to be observed in the discussion by the high as the low growth coach-teacher pairs. Values greater than 1 mean that the code was more likely to show up for the high growth coach-teacher pairs (e.g., OR = 2.0 means the high growth pair was twice as likely to show that code) and values less than 1 are interpreted as the code was less likely to show up in the high growth teacher pair (e.g., OR = 0.5 means that the high growth teacher was half as likely to show that code). Generally, OR = 1.7, 3.5, and 6.7 are considered "small", "medium" and "large" respectively, roughly corresponding to Cohen's d = 0.2, 0.5, and 0.8 (Chen et al. 2010).

We also used logistic regression to similarly examine differences in the Form of coach teacher interactions between high- and low-growth coach-teacher pairs. Further, we conducted exploratory analyses of the coach-teacher interaction data using survival analysis (see Willett and Singer 2004), to test whether patterns of coach-teacher interactions throughout an entire coaching session differentiated between high-growth and low-growth teachers. Survival analysis examines the amount of time that elapses until a particular event occurs and provides a log-rank test to see if the probability of reaching that event at a particular time is significantly different between two groups. For these analyses, the event was defined as the time elapsed until the first observation of a coach utterance that had been coded as a Tell within each coach-teacher interaction Segment. This method allowed us to examine patterns of Coaching Orientation codes sequentially, to determine whether differences between high growth vs. low growth coach-teacher pairs also appeared in the timing of those questioning styles throughout an entire coaching session.

## Results

### Content of interactions

Our first set of results are related to analyses of the Content codes of the coach-teacher interactions. Across both high- and low-growth coach-teacher pairs, coaches were found to be just as likely to discuss the pedagogical Actions a teacher would perform in their classrooms (OR = 0.75, p = 0.32). That is, discussion about *what* specific teaching moves would take place during the lesson were not a clear differentiator of the coach-teacher pairs that showed high or low growth in Instructional Quality (see Fig. 2).

However, the higher growth coach-teacher pairs were especially likely to consider both the Conditions (i.e., Goals, Tasks and Student Thinking) that would exist in the classroom during the lesson (OR = 5.6, $p < 0.05$), and explicitly elicit the teachers' Reasoning about the decisions that they were making in the classroom (OR = 2.1, $p < 0.05$). That is, coach-teacher pairs who considered both the conditions when a certain move may be performed and had an explicit discussion about why that particular move would be appropriate at that time, were more likely to display observable growth in Instructional Quality during their lessons.

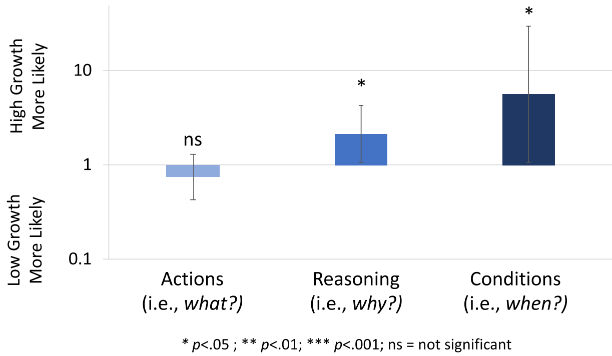
The Conditions category overall showed the largest differences between high and low growth coach-teacher pairs. Theoretically, however, developing the advanced reasoning about conditions for actions required for adaptive expertise should involve attending to multiple types of conditions together (e.g., considering both the task and the goal; Smith et al. 1997; Goldstone and Son 2005). Since combinations of Conditions sub-codes (i.e., Tasks, Goals and Student Thinking) were found to be common within a Segment, follow-up analyses were performed on combinations of these three sub-codes within Conditions. All sub-categories of Conditions were directionally more likely in the high-growth pairs, but only Tasks (specific instantiations of practice) and Goals (abstractions to guide in-the-moment pedagogical decision-making) were significantly higher for high-growth pairs (see Appendix, Fig. 7). All pairwise combinations of these conditions were also directionally more likely in the high growth pairs, but the high-growth coach-teacher pairs showed only one statistically robust difference: they were much more likely to discuss a combination of the specific Tasks for the upcoming lesson and the more general instructional Goals for students (OR = 2.2, $p < 0.05$, see Fig. 3). This finding provides preliminary evidence that there is an association between effective coaching sessions and discussion specifically about both the concrete details of the lesson and the more general instructional goals.

## Form of interactions

Next, we examined differences between high and low growth coach-teacher pairs in the Form of their interactions. Overall, coaches in high growth pairs were more likely to pose Open-ended questions (OR = 3.5, $p < 0.001$) and less likely to pose questions as a limited set of Options to choose from (OR = 0.45, $p < 0.05$; see Fig. 4). This pattern was consistently found when the analyses was conducted separately for each content code (i.e., difference in Form within Actions, Reasoning, or Conditions; see Appendix A, Fig. 8). Surprisingly, there were no differences in the likelihood that a high growth or low growth coach would Tell the teacher information (OR = 1.3, $p = 0.4$), providing contrasting evidence to literature suggesting that any coach-provided direct instruction may limit teacher growth (Heineke 2013). However, in our subsequent analyses the timing of the Tell proved to be important.

To explore the dialogic and interactive nature of coach-teacher interactions, we also analyzed the Form codes as time-series data using survival analysis, to see if there were sequential differences in the patterns of interactions between high- and low-growth coach-teacher pairs. Results showed that within a single discussion segment, high-growth pairs were likely to have significantly more coach-teacher interactions ($\chi^2(1) = 69.85$, $p < 0.001$) take place before the observation of a Tell by the coach (see Fig. 5). This analysis provides evidence suggesting a potential explanation for the null differences found in Telling above; that is, while direct instruction by the coach may not be in itself unproductive for teacher learning from instructional coaches, it is important when a coach chooses to provide that direct instruction to the teacher.

**Fig. 2** Odds ratios of Content codes, for coach-teacher pairs with high and low growth on the Instructional Quality scale, with bars for 95% CI shown. (Coach-teacher pairs are equally likely to have the code when the 95% CI bars include 1. Scale transformed to logarithmic to accurately depict Odds Ratios.)




## Discussion

This study leverages a large, state-wide representative dataset of transcripts of conversations between in-service teachers and instructional coaches, in order to provide a well-matched case-comparison sample of coaches who show different growth in their learning of high-leverage teaching practices. This unique sample allowed an in-depth, mixed methods approach to understanding mechanisms through which the same instructional coaching model can lead to different outcomes in terms of teacher enactment of these practices in their classroom.

Further, this study provides a novel approach to research on teacher professional development by applying a cognitive lens to understanding how specific aspects of coach-teacher interactions can support the development of adaptive expertise in ambitious instruction. For example, our results provide evidence that professional development opportunities may better support this type of teacher learning outcome if they focus on attending to not just *what* the high-leverage pedagogical practices teachers should engage in are, but also specifically if they explicitly discuss the conditions *when* those practices are appropriate to use, and engage teachers in reasoning about *why* those practices may be effective under those conditions. This finding supports recent calls for the design of professional development opportunities that unite two epistemological perspectives: one that emphasizes the deliberate acquisition of knowledge (*what* practices) and another that emphasizes the

**Fig. 3** Odds ratios of combinations of Conditions codes within a segment for coach-teacher pairs with high and low growth on the Instructional Quality scale, with bars for 95% CI shown. (Coach-teacher pairs are equally likely to have the code when the 95% CI bars include 1. Scale transformed to logarithmic to accurately depict Odds Ratios.)

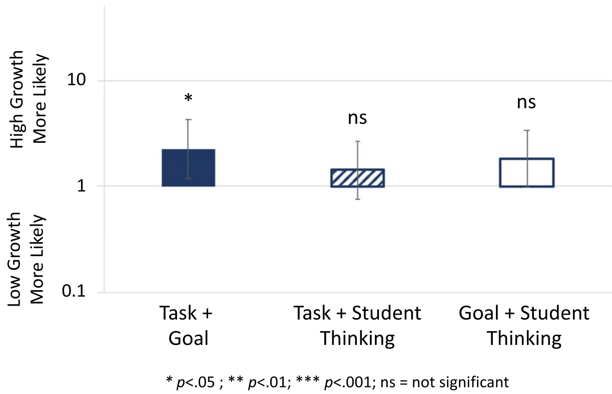


**Fig. 4** Odds ratios of Form codes, for coach-teacher pairs with high and low growth on the Instructional Quality scale, with bars for 95% CI shown. (Coach-teacher pairs are equally likely to have the code when the 95% CI bars include 1. Scale transformed to logarithmic to accurately depict Odds Ratios.)

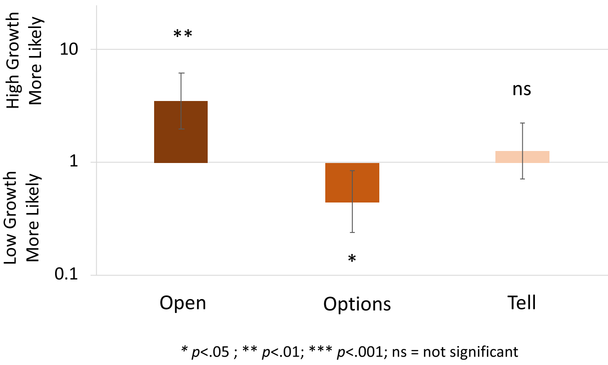
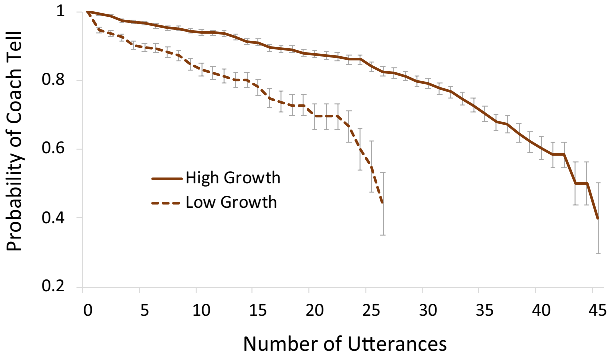



**Fig. 5** Survival analysis of Sequential Form codes showing the number of coach or teacher utterances until the observation of a coach Tell, for coach-teacher pairs with high and low growth on the Instructional Quality scale, with SE bars shown




development of situative awareness (*when* and *why* those practices; van der Linden and McKenney 2020).

Specifically, results from this study showed significant differences in the Content and Form of instructional coaching interactions between coach-teacher pairs who demonstrated more or less growth in their implementation of ambitious instructional practices. Coach-teacher pairs who showed larger gains in maintaining cognitive demand and opportunities for student engagement were more likely to discuss both concrete details of the upcoming lesson as well as overarching learning goals within the same interaction segment. This finding supports earlier work in the cognitive sciences that suggests that drawing connections between specific instantiations of problems (e.g., specific lesson tasks) and more generalized abstractions of those problem (e.g., broader learning goals; Hiebert et al. 2007; Stein and Meikle 2017) can help learners generate adaptive solutions within novel contexts, and extends this line of work to include teachers as learners (Goldstone and Son 2005). In contrast to predictions from the current literature, these data did not show a significant relationship between the discussion of student thinking and ambitious instruction. It is possible that the focus on pre-conference conversations in the current study emphasized the importance of unpacking the concrete details of the task in relation to the learning goals of the lesson, while coach-teacher discussions and reflection on student thinking that appeared during enactment of the lesson were reserved for post-conference conversations. Additional analyses applying a similar analytic approach to post-conference transcripts could test this hypothesis.

In addition, our findings provide some preliminary insights into the Form of coach-teacher interactions that may be most productive for teacher learning. For example, coach-teacher pairs who showed more growth in maintaining cognitive demand and opportunities for student engagement were more likely to pose questions to teachers using open-ended prompts. Interestingly, there were no significant differences between low- and high-growth coach-teacher pairs overall on whether or not coaches provided direct instruction to the teacher by telling them their desired response, a finding that at first appears to suggest no difference between constructivist and direct instruction approaches to coaching (Collet 2012; Kirschner et al. 2006). However, additional analyses examining the transcript data sequentially as it occurs over time across an entire session provides a more nuanced view. Our findings show that in more effective coaching sessions, coaches adopt an Orientation that provides teachers with more opportunities for input and reflection on their practice, by remaining engaged in longer discussions about each pedagogical decision *before* telling the teacher their own interpretation of what a desirable response may be. This finding supports recent work by Haneda et al. (2019) suggesting that instead of a false dichotomy between direct and responsive coaching stances, more effective coaches are able to balance these orientations and determine a "time for telling" (Schwartz and Bransford 1998) that provides the best preparation for teacher to develop adaptive expertise in these practices through their coaching conferences. Coaches that take such an orientation towards professional development are more likely to foster an environment where teachers can engage in the kind of reflection and metacognition that allow them to develop a deeper understanding of the conditions under which high-leverage practices can be performed effectively (Haneda et al. 2017). This also corroborates and extends recent findings showing that variation in levels of adaptive expertise in teaching is linked to varying emphasis on a fixed versus open teaching orientation (Männikkö and Husu 2019); teachers who engage in instructional coaching where this variation in orientation is modeled may be more likely to develop adaptive expertise.

## Limitations

When interpreting these results, it is important to consider the following limitations to the current study. First, as a correlational design, the study cannot make strong causal claims. However, as a novel research approach combined with foundational theories from the cognitive science of learning, the findings provide initial support for areas of support that could be especially important for fostering adaptive expertise in teaching.

Second, while our results suggest that a greater number of utterances between the coach posing questions and providing direct instruction are predictive of teachers who show more growth in maintaining cognitive demand and opportunities for student engagement, it is likely that it is not purely the length, but also the quality of the content of these longer discussions that impacts teacher learning. It is possible that coaches and teachers, after an open-ended question is posed, could have long conversations that are off-topic and not productive. However, at least in this study, the long conversations did not appear to be harmful, and thus were likely important to allow sufficient exploration of concepts.

Finally, while high vs. low teacher group contrast provides insight into the nature and timing of coach actions that may produce differences in teacher learning, our current data does not provide direct insight into the role of the teacher in driving these decisions. The dialogic nature of the coach-teacher interactions means that variation in a teacher's input could lead to both the Content and Form of a coach's input. Indeed, in more varied groups, a coach's Orientation should be contingent on a variety of factors such as teachers' prior knowledge, proficiency with the pedagogical practices, and even affective components of their relationship with the teacher (Ippolito 2010; Lowenhaupt et al. 2014). Future research on this topic could explore variation within a coach across teachers or within a teacher across time and topics.

## Conclusions

Teacher learning is a huge challenge in instructional change, but relatively little work has carefully examined the mechanisms by which teachers learn, in contrast to the extensive work on programs that help teachers learn and the high-leverage instructional practices that are strong predictors of student learning. This study provides a novel approach to understanding the cognitive mechanisms that underlie teachers' learning to implement ambitious instruction. Specifically, it identifies key components of a particular professional development approach, instruction coaching, that are related to larger growth in teachers' maintenance of the rigor of their lessons and in providing students with opportunities to productively struggle with the content. Taken together, the findings here build on and extend prior research on adaptive expertise, by providing evidence for the importance of attending to building well-organized knowledge structures and meta-cognitive skills in developing adaptive expertise, within the context of teachers as learners (Smith et al. 1997). Our findings also provide a potential cognitive explanation for apparent discrepancies in earlier research that suggests making direct recommendations can be a beneficial practice by coaches (Collet 2012), and that it can be detrimental if coaches dominate interactions (Heineke 2013). Instead, it may be that allowing teachers with additional time to "productively struggle" (Stein et al. 2017) to make sense of their own pedagogical decisions before giving a direct recommendation is more important than whether or not the teacher is eventually given the answer by the coach (Schwartz and Bransford 1998).

Further, the results and analytic approach of this study can also provide insight into the design of other models for professional development that focus specifically on building teachers' understanding of when and why certain high leverage practices should be implemented. For example, video-clubs, another popular form of teacher professional development, have also been shown to help teachers develop an ability to recognize and draw out common instructional practices. Recent reviews of research on professional learning communities (PLCs) have shown that while there is evidence of improvement of both teacher practice and student learning, relatively few studies document the actual processes of teachers' change in practice in the classroom (Vescio et al. 2008). Additional work that focuses on uncovering common cognitive mechanisms of teacher learning in these other professional development environments could provide important insights for understanding the most important levers for supporting teachers in the challenge of implementing ambitious instructional practices.

## Appendix A

See Figs. 6, 7, 8.

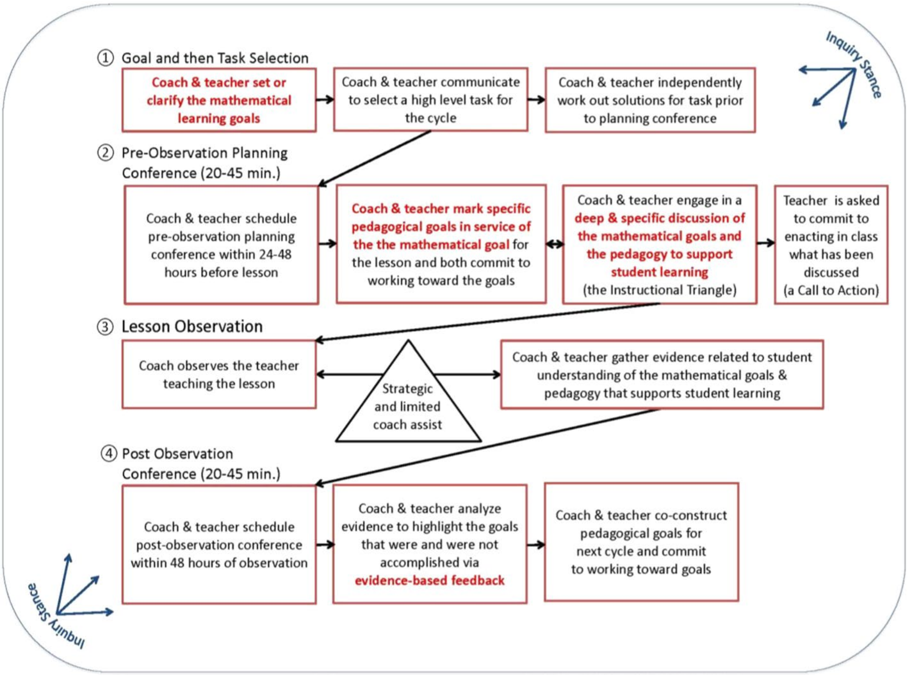
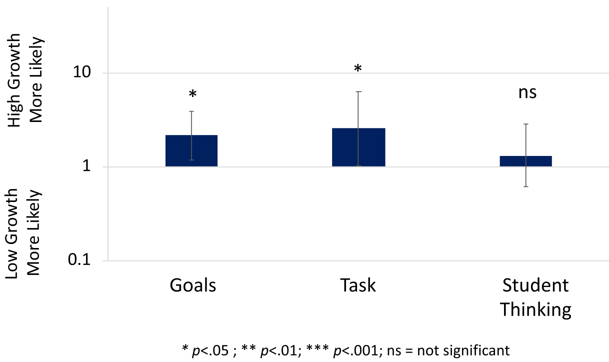



**Fig. 6** A detailed model of the full Coaching Cycle

**Fig. 7** Odds ratios of combinations of Conditions sub-codes within a Segment for coach-teacher pairs with high and low growth on the Instructional Quality scale, with bars for 95% CI shown. (Coach-teacher pairs are equally likely to have the code when the 95% CI bars include 1. Scale transformed to logarithmic to accurately depict Odds Ratios.)



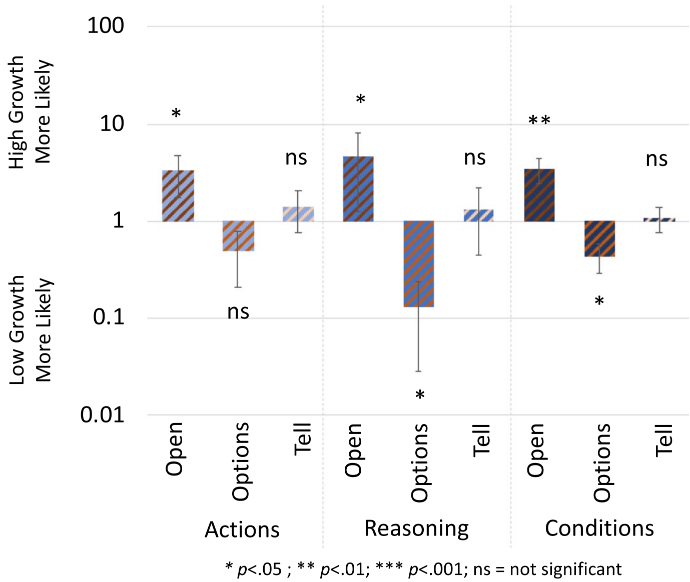



**Fig. 8** Odds ratios of combinations of Form codes by Content codes, within a Segment for coach-teacher pairs with high and low growth on the Instructional Quality scale, with bars for 95% CI shown. (Coach-teacher pairs are equally likely to have the code when the 95% CI bars include 1. Scale transformed to logarithmic to accurately depict Odds Ratios.)

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s11251-021-09536-7.

**Acknowledgements** We would like to thank Lindsay Clare Matsumura and Marguerite E. Walsh for their valuable feedback on earlier drafts of this manuscript.

**Funding** This work was supported by a Teachers as Learners grant from the James S. McDonnell Foundation. The views and conclusions contained herein are those of the authors and should not be represented as official policies of the James S. McDonnell Foundation.

**Data availability** De-identified data and code (Stata 16) is available by request.

## Compliance with ethical standards

**Conflict of interest** The authors declare no conflicts of interest.

## References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics, 25*(1), 95–135. https://doi.org/10.1086/508733

Anthony, G., Hunter, J., & Hunter, R. (2015). Prospective teachers development of adaptive expertise. *Teaching and Teacher Education, 49,* 108–117. https://doi.org/10.1016/j.tate.2015.03.010

Bakhtin, M. (1981). The Dialogic Imagination. In M. Holquist (Ed.), *Hortus Artium Mediaevalium* (Vol. 11). University of Texas Press.

Borko, H. (2004). Professional development and teacher learning: mapping the terrain. *Educational Researcher, 33*(8), 3–15. https://doi.org/10.3102/0013189X033008003

Bowers, N., Merritt, E., & Rimm-Kaufman, S. (2020). Exploring teacher adaptive expertise in the context of elementary school science reforms. *Journal of Science Teacher Education, 31*(1), 34–55. https://doi.org/10.1080/1046560X.2019.1651613

Butterfield, E. C., & Nelson, G. D. (1989). Theory and practice of teaching for transfer. *Educational Technology Research and Development, 37*(3), 5–38. https://doi.org/10.1007/BF02299054

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics: Simulation and Computation, 39*(4), 860–864. https://doi.org/10.1080/03610911003650383

Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education, 18*(8), 947–967. https://doi.org/10.1016/S0742-051X(02)00053-7

Collet, V. S. (2012). The gradual increase of responsibility model: coaching for teacher change. *Literacy Research and Instruction, 51*(1), 27–47. https://doi.org/10.1080/19388071.2010.549548

Common Core State Standards Initiative. (2010). Common Core State StandardS for. *Development*, 31(October), 93. http://www.corestandards.org/

Correnti, R., Matsumura, L. C., Walsh, M., Zook-Howell, D., Bickel, D. D. P., & Yu, B. (2020). Effects of online content-focused coaching on discussion quality and reading achievement: Building theory for how coaching develops teachers' adaptive expertise. *Reading Research Quarterly*, 1–40. https://doi.org/10.1002/rrq.317.

Darling-Hammond, L., Hammerness, K., Grossman, P., Rust, F., & Shulman, L. (2005). The Design of Teacher Education Programs. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing Teachers for a Changing World* (pp. 390–441). Jossey-Bass.

Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education, 34,* 12–25. https://doi.org/10.1016/j.tate.2013.03.001

Desimone, L. M., & Pak, K. (2017). Instructional coaching as high-quality professional development. *Theory into Practice, 56*(1), 3–12. https://doi.org/10.1080/00405841.2016.1241947

Frensch, P. A., & Sternberg, R. J. (1989). Expertise and intelligent thinking: When is it worse to know better? In *Advances in the psychology of human intelligence, Vol. 5.* (pp. 157–188). Lawrence Erlbaum Associates, Inc.

Garrett, R., Citkowicz, M., & Williams, R. (2019). How Responsive is a teacher's classroom practice to intervention? A meta-analysis of randomized field studies. *Review of Research in Education, 43*(1), 106–137. https://doi.org/10.3102/0091732X19830634

Goldsmith, L. T., Doerr, H. M., & Lewis, C. C. (2014). Mathematics teachers' learning: A conceptual framework and synthesis of research. *Journal of Mathematics Teacher Education, 17*(1), 5–36. https://doi.org/10.1007/s10857-013-9245-4

Goldstone, R. L., & Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. *Journal of the Learning Sciences, 14*(1), 69–110. https://doi.org/10.1207/s15327809jls1401_4

Haneda, M., Sherman, B., Nebus Bose, F., & Teemant, A. (2019). Ways of interacting: What underlies instructional coaches' discursive actions. *Teaching and Teacher Education, 78,* 165–173. https://doi.org/10.1016/j.tate.2018.11.017

Haneda, M., Teemant, A., & Sherman, B. (2017). Instructional coaching through dialogic interaction: helping a teacher to become agentive in her practice. *Language and Education, 31*(1), 46–64. https://doi.org/10.1080/09500782.2016.1230127

Hatano, G., & Inagaki, K. (1986). Two Courses of Expertise. In H. Stevenson, H. Azuma, & K. Hakuta (Eds.), *Children development and education in Japan* (pp. 262–272). Freeman.

Heineke, S. F. (2013). Coaching discourse. *The Elementry School Journal, 113*(3), 409–433.

Hiebert, J., Morris, A. K., Berk, D., & Jansen, A. (2007). Preparing teachers to learn from teaching. *Journal of Teacher Education, 58*(1), 47–61. https://doi.org/10.1177/0022487106295726

Horn, I. S., & Kane, B. D. (2015). Opportunities for professional learning in mathematics teacher workgroup conversations: relationships to instructional expertise. *Journal of the Learning Sciences, 24*(3), 373–418. https://doi.org/10.1080/10508406.2015.1034865

Ippolito, J. (2010). Three ways that literacy coaches balance responsive and directive relationships with teachers. *The Elementary School Journal, 111*(1), 164–190.

Isiksal, M., & Cakiroglu, E. (2011). The nature of prospective mathematics teachers' pedagogical content knowledge: The case of multiplication of fractions. *Journal of Mathematics Teacher Education, 14*(3), 213–230. https://doi.org/10.1007/s10857-010-9160-x

Kaiser, G., Blömeke, S., König, J., Busse, A., Döhrmann, M., & Hoth, J. (2017). Professional competencies of (prospective) mathematics teachers—cognitive versus situated approaches. *Educational Studies in Mathematics, 94*(2), 1–22. https://doi.org/10.1007/s10649-016-9713-8

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

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*(4), 437–457. https://doi.org/10.3102/0162373714557519.

Knight, J. (2009). Instructional coaching. *American Educational Research Journal, 46*(2), 532–566. https://doi.org/10.3102/0002831210371497

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: a meta-analysis of the causal evidence. *Review of Educational Research, 88*(4), 547–588. https://doi.org/10.3102/0034654318759268

Kuhn, D. (2007). Is direct instruction an answer to the right question? *Educational Psychologist, 42*(2), 109–113. https://doi.org/10.1080/00461520701263376

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

Lampert, M., Franke, M. L., Kazemi, E., Ghousseini, H., Turrou, A. C., Beasley, H., Cunard, A., & Crowe, K. (2013). Keeping it complex: using rehearsals to support novice teacher learning of ambitious teaching. *Journal of Teacher Education, 64*(3), 226–243. https://doi.org/10.1177/0022487112473837

Lowenhaupt, R., McKinney, S., & Reeves, T. (2014). Coaching in context: the role of relationships in the work of three literacy coaches. *Professional Development in Education, 40*(5), 740–757. https://doi.org/10.1080/19415257.2013.847475

Männikkö, I., & Husu, J. (2019). Examining teachers' adaptive expertise through personal practical theories. *Teaching and Teacher Education, 77,* 126–137. https://doi.org/10.1016/j.tate.2018.09.016

Reznitskaya, A., & Gregory, M. (2013). Student thought and classroom language: Examining the mechanisms of change in dialogic teaching. *Educational Psychologist, 48*(2), 114–133. https://doi.org/10.1080/00461520.2013.775898

Richardson, V. (2003). Constructivist Pedagogy. *Teachers College Record, 105*(9), 1623–1640. https://doi.org/10.1046/j.1467-9620.2003.00303.x

Rowland, T., & Ruthven, K. (2011). Introduction: mathematical knowledge in teaching. In T. Rowland & K. Ruthven (Eds.), *Mathematical Knowledge in Teaching* (pp. 1–5). Springer.

Russell, J. L., Correnti, R., Stein, M. K., Bill, V., Hannan, M., Schwartz, N., et al. (2020). Learning from adaptation to support instructional improvement at scale: Understanding coach adaptation in the TN mathematics coaching project. *American Educational Research Journal*, *57*(1), 148–187. https://doi.org/10.3102/0002831219854050.

Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: rethinking mechanism of a neglected phenomenon. *Educational Psychologist, 24*(2), 113–142. https://doi.org/10.1207/s15326985ep2402

Schwartz, D., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16*(4), 367–398. https://doi.org/10.1207/s1532690xci1604

Schwartz, D., Bransford, J. D., & Sears, D. (2005). Efficiency and Innovation in Transfer. *Transfer of Learning from a Modern Multidisciplinary Perspective, 3,* 1–51. https://doi.org/10.1111/j.1365-2133.2005.06492.x

Sherin, M. G., & Han, S. Y. (2004). Teacher learning in the context of a video club. *Teaching and Teacher Education, 20*(2), 163–183. https://doi.org/10.1016/j.tate.2003.08.001

Smith, E. M., Ford, J. K., & Kozlowski, S. W. J. (1997). Building adaptive expertise: Implications for training design strategies. *Training for a Rapidly Changing Workplace Applications of Psychological Research*. https://doi.org/10.1037/10260-004

Smith, J. B., Lee, V. E., & Newmann, F. M. (2001). *Improving Chicago's Schools: Instruction and Achievement in Chicago Elementary Schools*. Consortium on Chicago School Research.

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), 1–20. https://doi.org/10.1177/2332858416680566.

Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. *Mathematical Thinking and Learning*, *10*(4), 313–340. https://doi.org/10.1080/10986060802229675.

Stein, M. K., & Kaufman, J. H. (2010). Selecting and supporting the use of mathematics curricula at scale. *American Educational Research Journal*, *47*(3), 663–693. https://doi.org/10.3102/0002831209361210.

Stein, M. K., & Meikle, E. (2017). The nature and role of goals in mathematics education. In D. Spangler & J. Wanko (Eds.), *Research companion to principles to action* (pp. 1–11). National Council of Teachers of Mathematics. https://doi.org/10.1007/978-3-319-31450-1.

Stroupe, D., & Windschitl, M. (2015). Supporting ambitious instruction by beginning teachers with specialized tools and practices. *Newly Hired Teachers of Science*. https://doi.org/10.1007/978-94-6300-283-7_13

Tobias, S., & Duffy, T. M. (2009). Constructivist instruction: Success or failure? In S. Tobias & T. M. Duffy (Eds.), *Journal of Educational Technology & Society*. Routledge. https://www.jstor.org/stable/https://doi.org/10.2307/jeductechsoci.13.3.281

van der Linden, S., & McKenney, S. (2020). Uniting epistemological perspectives to support contextualized knowledge development. *Educational Technology Research and Development, 68*(2), 703–727. https://doi.org/10.1007/s11423-020-09772-7

Verschaffel, L., Luwel, K., Torbeyns, J., & Van Dooren, W. (2009). Conceptualizing, investigating, and enhancing adaptive expertise in elementary mathematics education. *European Journal of Psychology of Education, 24*(3), 335–359. https://doi.org/10.1007/BF03174765

Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education, 24*(1), 80–91. https://doi.org/10.1016/j.tate.2007.01.004

Vygotsky, L. S. (1978). *Mind in Society: The Development ofHigher Psychological Processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (eds.)). Harvard University Press. https://doi.org/https://doi.org/10.3928/0048-5713-19850401-09

Wells, G., & Arauz, R. M. (2006). Dialogue in the classroom. *Journal of the Learning Sciences, 15*(3), 379–428. https://doi.org/10.1207/s15327809jls1503_3

Willett, J. B., & Singer, J. D. (2004). How long did it take? Using survival analysis in educational and psychological research. *Best Methods for the Analysis of Change: Recent Advances, Unanswered Questions, Future Directions.*, *April 1989*, 310–327. https://doi.org/https://doi.org/10.1037/10099-018

Windschitl, M., Thompson, J., Braaten, M., & Stroupe, D. (2012). Proposing a core set of instructional practices and tools for teachers of science. *Science Education, 96*(5), 878–903. https://doi.org/10.1002/sce.21027

Wray, D., Medwell, J., Fox, R., & Poulson, L. (2000). The teaching practices of effective teachers of literacy. *Educational Review, 52*(1), 74–84. https://doi.org/10.1080/00131910097432



## Authors and Affiliations

**Eben B. Witherspoon[1] · Nathaniel B. Ferrer[1] · Richard R. Correnti[1] · Mary Kay Stein[1] · Christian D. Schunn[1]**

Nathaniel B. Ferrer
nbf8@pitt.edu

Richard R. Correnti
rcorrent@pitt.edu

Mary Kay Stein
mkstein@pitt.edu

Christian D. Schunn
schunn@pitt.edu

[1] Learning Research and Development Center, University of Pittsburgh, 3939 O'Hara St, Pittsburgh, PA 15217, USA