Coaching That Supports Teachers' Learning to Enact Ambitious Instruction

Eben B. Witherspoon, Nathaniel Ferrer, Richard Correnti, Christian D. Schunn, and Mary Kay Stein
ebw13@pitt.edu, nbf8@pitt.edu, rcorrent@pitt.edu, schunn@pitt.edu, mkstein@pitt.edu
Learning Research and Development Center, University of Pittsburgh

Abstract: Research has identified a number of high-leverage instructional practices that are strong predictors of student learning. However, relatively little is known about how teachers learn to effectively implement these instructional practices. Using a mixed-methods, case-comparison design, this study uses a cognitive lens to examine specific features of instructional coaching that support 3rd–8th grade mathematics teachers in learning to implement ambitious instructional practices. Results showed that coach-teacher pairs who discuss when and why certain practices should be implemented during pre-lesson conferences see larger gains on a measure of ambitious instruction in later lessons. Implications for the design of professional development models based in cognitive and learning science theories are discussed.

Keywords: teacher education, professional development, ambitious instruction, cognitive science

Background and context

There is a broad consensus among educational researchers that instructional quality is one of the strongest predictors of student achievement (Aaronson, Barrow, & Sander, 2007). Further, research on teachers' instructional quality over the past two decades has emphasized the importance of "ambitious instruction", a set of practices broadly characterized 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; J. B. Smith, Lee, & Newmann, 2001). There is a growing evidence base for specific such practices that are reliable indicators of instructional quality in both mathematics (Stein, Engle, Smith, & Hughes, 2008) and literacy (Wray, Medwell, Fox, & Poulson, 2000), but 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 & Arauz, 2006).

Professional development workshops, professional learning communities, video clubs, and instructional coaching have all been suggested as potentially powerful interventions that may influencing teacher learning to implement such practices (Borko, 2004; Knight, 2009; Sherin & Han, 2004; Vescio, Ross, & Adams, 2008). 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, Citkowicz, & Williams, 2019). For example, 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, Blazar, & Hogan, 2018).

Building a theory of teacher learning focused on specific learning mechanisms can guide efforts to better understand variation in implementation, but requires careful conceptualization of what is difficult in ambitious instruction. We propose that ambitious instruction requires teachers not only to 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 unforeseen student contributions). Effective implementation of ambitious instruction therefore requires learning to perform familiar instructional practices accurately and efficiently, as well as the ability to recognize and replicate those performances in a number of similar but non-identical scenarios.

Theoretical framework

Developing adaptive expertise in teaching

In the cognitive sciences, this ability to accurately apply knowledge learned in one context in novel situations has been referred to as "adaptive" as opposed to "routine" expertise (Hatano & Inagaki, 1986; Schwartz, Bransford, & Sears, 2005). 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.

Especially when attempting to implement ambitious instruction, where there are a nearly infinite combination of potential student inputs and teacher responses, this can often lead to misapplication of that learned skill. Adaptive experts, however, develop an understanding of the necessary and sufficient conditions under which a certain solution will work, recognize when those conditions have changed, and adapt the previously-learned procedure to produce a new solution to account for those changes. 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; successful implementation will therefore depend on the development of adaptive expertise. For example, it may not be enough for teachers to simply know how to perform a list of high-leverage practices; they also need to learn *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. Interestingly, cognitive learning theories like adaptive expertise were developed in the context of students as learners; there is little research applying these theories to understand how this kind of expertise is developed in teachers as learners.

Cognitive mechanisms for developing adaptive expertise

Cognitive research suggests that adaptive expertise can be developed in part through creating well-organized knowledge structures during learning (E. M. Smith et al., 1997). In the context of teacher learning, attending to student learning goals during professional development may offer one mechanism for creating these well-organized knowledge structures: goals provide a metric through which to understand *why* 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, Morris, Berk, & Jansen, 2007; Stein & Meikle, 2017).

Further, explicitly grounding these goals in the concrete details of student activities could help teachers develop the ability to generalize their knowledge that characterizes adaptive expertise (Goldstone & 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 this process 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.

However, existing teacher training environments are unlikely to explicitly include these elements (e.g., linking task aspects to broader conceptual goals), and therefore the development of these understandings for teachers may only occur across multiple lessons and cohorts of students over the course of many years, or not at all. In sum, it is important to identify how professional development models can explicitly attend to *when* and *why* teachers should enact certain high leverage practices as well as understand the learning mechanisms through which teachers learn how to implement these practices in an adaptive way.

The current study: Towards a cognitive model of instructional coaching

Instructional coaching provides one promising model for efficiently developing teachers' expertise in implementing ambitious instruction. Coaching has already been found to incorporate aspects of professional development associated with teacher self-reports of learning: a focus on content, active learning, policy coherence, and participation that is sustained and collective (Desimone & Pak, 2017). But these aspects of teacher professional development environments do not distinguish between models of instructional coaching that would predict the development of routine or adaptive expertise. One additional component of instructional coaching that could contribute to developing adaptive expertise is the coaches role in clarifying, synthesizing and simplifying high-leverage teaching practices (Knight, 2009). For example, while a teacher is focusing on certain core mathematical learning goals, coaches can help synthesize across multiple lessons during pre- or post-lesson conferences to direct teachers' attention to connections between concrete aspects of the mathematics task for the upcoming lesson, and broader conceptual learning goals.

In the current study, we apply a mixed-methods design to understand the particular characteristics of coach-teacher interactions that most commonly occurred with teachers that showed gains in their ability to conduct ambitious instruction in elementary mathematics classrooms. We first conduct a qualitative analysis of transcripts of teacher and coach conferences to develop codes that 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 teachers' learning to adaptively implement ambitious instructional practices. We then use a case comparison design and quantitative analyses to understand the character of interactions that were most likely to be observed among teachers' who demonstrated growth in their ambitious instruction, which provides a test of the specific aspects of coaching interactions that were theoretically predicted to be associated with teacher learning. Therefore, our main research questions were as follows:

1.) Are high-growth coach-teacher interactions more likely to focus on the "when" and "why" of teacher practices?
2.) Do high-growth coach-teacher pairs focus specifically on the connections between the task, student thinking, and the learning goal of the lesson?

Methods

The current study was part of a larger project examining the impact of instructional coaching on the practice of K-12 mathematics teachers throughout a multi-year, state-wide coaching project. The work of the larger coaching project had two primary goals. First, a 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 cycles with two partnering teachers which included four main stages: Goal and Task Selection, Pre-Observation Planning Conference, Lesson Observation and Post-Observation Conference. Coaches completed this full Discussion Process with each of their partner teachers for three coaching cycles in year one of data collection (2014-15) and two 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. Partner teachers also videotaped a lesson that represented their typical mathematics instruction prior to beginning to work with their coach (pre-coaching lesson), and a lesson they did not work with a coach to plan or implement at the end of the school year (post-coaching lesson).

Sample and procedures

The full dataset from the larger project included 32 coaches and *N*=105 partner teachers: *n*=40 partner teachers participated in year one only, *n*=41 participated in year two only, and *n*=24 participated in both years one and two. Four teachers who interacted with four partner coaches (one teacher always with the same coach) were selected for analysis representing two showing high growth and two showing low growth (see below for details).

All 105 teachers' lessons (pre, coached, post) were scored on a 2 to 8 scale for instructional quality focused on the dimension of maintenance of cognitive demand that was emphasized in the coaching model and previously found to be predictive of growth in students' learning (see Measures for details). The videos were scored 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.

The growth in cognitive demand was statistically modeled using the scores for all of their observed lessons to produce a growth estimate. Some teachers started very low and struggled with 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. Other teachers already started relatively high and had little room to grow on the cognitive demand scale. Therefore, only 63 teachers who started near the mid-point of the cognitive demand scale (*M*=5.3, *SD*=0.1) were considered for selection. These coach-teacher pairs showed overall growth and variation in growth on the cognitive demand scale (*M*=1.6, *SD*=0.4). From this subset of coach-teacher pairs, we randomly selected 2 coach and teacher pairs who had participated in both years of coaching from the top third of the sample in terms of growth on the cognitive demand scale (*M*=2.0, *SD*=0.3), and another 2 coach and teacher pairs from the bottom third in terms of growth (*M*=1.2, *SD*=0.4), to create our 4-pair case comparison sample for qualitative coding.

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., Cycle A) and instead focused on the second and third rounds of conversations (i.e., Cycle B and C), leaving us with a total sample of *N*=16 transcripts across the 4 pairs for coding.

Measures

Codes of coach-teacher interactions

Measures of the quality of coach-teacher interactions during pre-conferences were operationalized by a collection of codes describing the particular content of the coach-teacher conversations. For each measure, Cohen's kappa and raw percent agreement are reported as a measure of reliability. Content of 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 Reasoning (see Table 1 for details). Actions were an indicator for interactions that were primarily focused on the "What" of instruction. That is, it emphasized the particular pedagogical actions or moves that teachers would perform during the lesson, or a discussion of how to perform a particular instructional move. 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. Finally, the content of coach-teacher interactions was also measured through Conditions, which could be mentioned by either the teacher or the coach. Conditions described the particular settings under 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 appropriate. Conditions were further sub-coded to identify whether the teacher and coach had raised for consideration the Goals of the lesson, the Task being used during the lesson, or Student Thinking during the lesson (1).

Teacher growth in maintenance of cognitive demand

Our primary outcome measure of teachers' growth in ambitious instruction was operationalized as the opportunities that students were given to demonstrate their thinking on tasks with high cognitive demand, coded on a scale of two to eight by expert coders from video of classroom observations (as described in the Procedures

Table 1: Description and examples for each code category and sub-code for teacher and coach utterances, with Cohen's kappa provided as a measure of inter-rater reliability.

	Theme	Actor	Code	Description	Example
Content of Coach-Teacher Interactions	Focus of Prompt ($\kappa = .85$)	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 want kids to understand about that? ***Why do the groups have to be what you're saying?"***
	Lesson Conditions ($\kappa = .75$)	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."***

section above). Following procedures described in Stein (2010), a "score" was calculated for the maintenance of cognitive demand throughout a lesson. Specifically, coders assessed 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. Additionally, raters scored the degree to which teachers explored and facilitated the public display of student thinking throughout the lesson on a scale from 1 to 4. We adjusted the scale of the latter item in order to develop a mean of the two scales – yielding a composite on a scale from 2 to 8. A higher score on this composite represents not only maintenance of cognitive demand of the task during the lesson, but also whether students had the opportunity to engage in and make public their (conceptual) thinking. For this overall measure, an intra-class correlation (ICC) of .62 was calculated, indicating adequate inter-rater reliability. In the overall sample, teachers showed a range of scores at both pre-observation (M=5.3, SD=1.8) and post-observation (M=6.9, SD=0.5), as well as high variation in gains of maintaining cognitive demand from pre- to post-observation (M=1.6, SD=1.5).

Analytic methods

Prior to analysis, codes were summarized as means within a segment of the transcript, using a single Segmentation code (κ = .85, 93% agreement) that indicated a shift in the topic of the coach and teacher interaction. This was coded as "New" if the topic of the interaction shifted, or as "Add" if there was additional interaction with the same topic. Collapsing across this code allowed us to analyze the co-occurrence of various aspects of the interaction that were included within a single discussion segment. 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. Using this 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 codes identified in the transcripts of coaching conversations teachers who showed the most growth, and those who showed the least growth, along the scale of our primary 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), where an odds ratio of 1 would mean that the codes were just as likely to be observed in the discussion by the high and 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 = .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, Cohen, & Chen, 2010). In addition to testing for difference between the high and low coaches on individual codes, we also tested for differences in combinations of Purpose and Conditions codes.

Findings

Across both high and low teacher growth pairs, coaches were just as likely to discuss the pedagogical actions a teacher would perform in their classrooms (OR = 0.75, p = .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 maintenance of cognitive demand (see Figure 1). 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 <.05), and explicitly elicit the teachers' reasoning about the decisions that they were making in the classroom (OR = 2.1, p<.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 the maintenance of cognitive demand during their lessons.

Importantly for understanding the mechanisms through which these differences in interaction might influence teacher learning and subsequent growth in the maintenance of cognitive demand, additional analyses were performed on sub-codes within Conditions, the category with the largest differences between high and low growth coach-teacher pairs.

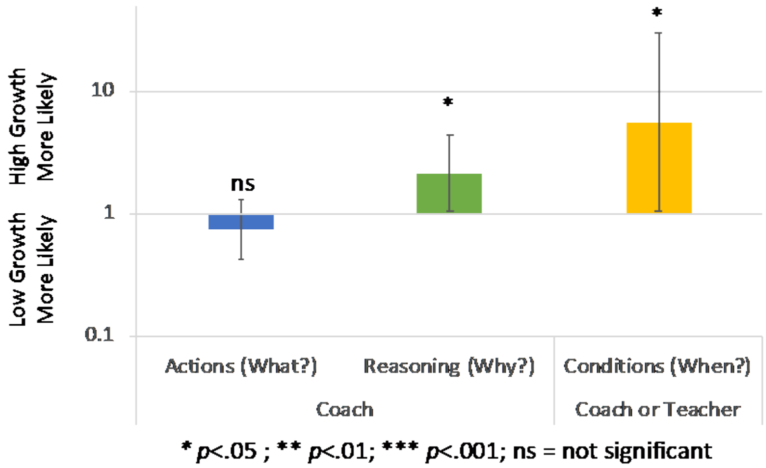

Figure 1. Odds ratios of Content codes, for coach-teacher pairs with high and low growth on the cognitive demand scale, with bars for 95% CI shown. (2)

Results show that within a particular discussion segment, high-growth coach-teacher were much more likely to discuss a combination of both the specific tasks for the upcoming lesson, as well as more general learning goals for students (OR = 7.1, $p<.001$, see Figure 2). This finding provides preliminary evidence that there is an association between effective coaching sessions and discussion about both the concrete specifics of the lesson and more general instructional goals.

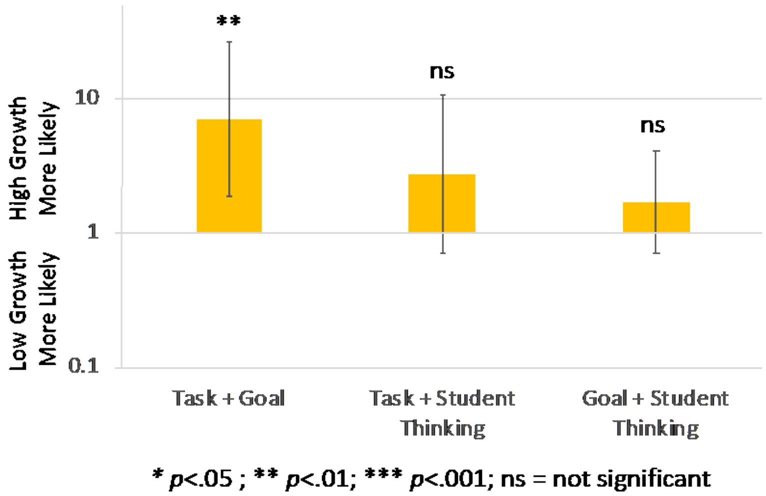

Figure 2. Odds ratios of Conditions codes within a segment, for coach-teacher pairs with high and low growth on the cognitive demand scale, with 95% CI bars shown. (2)

Discussion and implications

Results from this study showed significant differences in the content of instructional coaching interactions between coach-teacher pairs who demonstrated more or less growth in their implementation of ambitious instructional practices. Specifically, coach-teacher pairs who showed larger gains in maintaining cognitive demand and opportunities for student engagement were more likely to discuss within the same interaction segment both concrete details of the upcoming lesson as well as overarching learning goals. This finding supports earlier work in the cognitive sciences that suggests that explicitly drawing connections across specific instantiations of common problems can help develop learning that is generative of solutions within novel contexts, and extends this line of work to include teachers as learners (Alfieri et al., 2013; Goldstone & Son, 2005). Further, it 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 predict observed teacher growth in enactment of ambitious instruction.

This approach can also provide an insight into key features for the design of new 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, has also been shown to help teachers develop an ability to recognize and draw out common instructional practices. Prior work has shown that synthesizing multiple student ideas within a single lesson are thought to be the most expert form of video reflection (Sherin & Han, 2004), and there has been some evidence that engaging in case comparisons between multiple videos can help teachers improve on general practices such as attending to student thinking (Kisa & Stein, 2014). However, these opportunities for generative learning could be made more explicit if the selection of video segments are intentionally coordinated across lesson around a particular high-leverage practice. Recent reviews of research on professional learning communities (PLCs) have also 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). PLCs may offer a particularly rich environment for developing adaptive expertise, by providing a space where multiple practitioners can engage in a process of generalizing the implementation of certain high-leverage practices through sharing specific examples of their teaching experiences that are connected to collective learning goals. Overall, using a cognitive lens to identify a theoretical framework and specific mechanisms through which professional development is thought to produce better teacher gains can contribute to the more intentional design of teacher educational opportunities focused on improving not just an understanding of what practices are effective, but also an understanding of *when* and *why* those practices are best implemented.

While in this particular study, teachers were controlled by study-design to be at relatively similar levels of experience and proficiency, more research will be needed to understand the interaction between teachers' readiness for particular instructional strategies and the approach taken by the coach. For example, another feature of instructional coaching that may be integral to teacher learning is the particular "stance" the coach takes during learning. Studies have identified two primary types of stances that coaches take towards instruction: one that is directive, characterized by coaches acting as the source of knowledge or authority; and another that is responsive, where the coach interacts in a way designed to elicit input and reflection from the teacher (Haneda, Sherman, Nebus Bose, & Teemant, 2019). Research has suggested that making direct recommendations can be a beneficial practice of coaches (Collet, 2012), but also can be detrimental if coaches dominate interactions (Heineke, 2013). While this study purposefully selected a sample of teachers with similar initial levels of proficiency to draw more direct comparisons, it may be that in more varied groups that the openness of coaches questioning should be contingent on teachers prior knowledge of the practices (Ippolito, 2010). Future work in this area would benefit from applying a similar cognitive lens to understand how the relative responsivity of coaches' interactions with teachers may influence teacher learning, and how this may translate to the development of adaptive expertise.

Endnotes

(1) More information about the coding framework can be viewed on the project OSF page: https://osf.io/q9hb8

(2) 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.

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

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning Through Case Comparisons: A Meta-Analytic Review. *Educational Psychologist*, *48*(2), 87–113. https://doi.org/10.1080/00461520.2013.775712

Borko, H. (2004). Professional Development and Teacher Learning: Mapping the Terrain. *Educational Researcher*, *33*(8), 3–15. https://doi.org/10.3102/0013189X033008003

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

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

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

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
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
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). New York, NY: 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
Ippolito, J. (2010). Three ways that literacy coaches balance responsive and directive relationships with teachers. *The Elementary School Journal Volume*, *111*(1), 164–190.
Kisa, M. T., & Stein, M. K. (2014). Using contrasting video cases of the enactment of cognitively demanding science tasks in professional development. *Proceedings of International Conference of the Learning Sciences, ICLS*, *2*(January), 808–815.
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
Lampert, M. (2001). *Teaching problems and the problems of teaching*. New Haven: Yale University Press.
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*, 89–118. https://doi.org/10.1037/10260-004
Smith, J. B., Lee, V. E., & Newmann, F. M. (2001). *Improving hicago s Schools Consortium on*.
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., & 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). https://doi.org/10.1007/978-3-319-31450-1
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
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
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

Acknowledgments

This work was supported by a grant from the James S. McDonnell Foundation.

