EXPLORING THE IMPACT OF DISCUSSION-LEADING PROFESSIONAL DEVELOPMENT ON TEACHERS' PRACTICE

Nicole Garcia
University of Michigan
nmgarcia@umich.edu

Meghan Shaughnessy
Boston University
mshaugh@bu.edu

Merrie Blunk
University of Michigan
mblunk@umich.edu

Jillian P. Mortimer
Albion College
jbpet@umich.edu

D'Anna Pynes
University of Notre Dame
kpynes@nd.edu

Darrius Robinson
University of Michigan
darriusr@umich.edu

This paper examines the impact of an intensive professional development on practicing teachers' mathematics discussion-leading practice. A tool for examining specific discussion-leading moves was used to measure change in teachers' practice as observed through submitted video recordings of mathematics discussions. Participants included 33 teachers from three school districts who submitted a total of 193 videos across the study. The findings reveal that the professional development had differential impact on participating groups. We explore group attributes that could contribute to these differential outcomes.

Keywords: Professional development, classroom discourse

Improving the teaching of mathematics remains a national priority and is critical for improving the learning opportunities for mathematics students. Reforms designed to improve teaching and learning have often centered on curriculum and standards, though studies demonstrate that "no in-school intervention has a greater impact on student learning than an effective teacher" (p.1, NCATE, 2010), indicating that closer attention to teaching quality is essential. This is no small task, given the scale of the teaching force. There are simply not enough teachers with the knowledge and teaching skills needed to reach all students (NCTQ, 2008; Ball & Forzani, 2009). This challenge requires robust professional development that can contribute to improving the teaching and learning of mathematics.

We argue that teachers profit from well-designed opportunities to develop new visions for practice, learn more about students' thinking, or work on specific mathematical topics. However, such opportunities are often insufficient to support teachers with the complexity of classroom teaching. These kinds of professional opportunities focus on critical resources for instruction, but not on the details of practice itself. The professional development (PD) that is the focus of this research was an effort to address the challenge of supporting the learning of practice, specifically mathematics discussion-leading practice. We situated the PD in a "common text" for working on practice, where participants engage in a form of "legitimate peripheral participation," followed by structured opportunities to enhance their capabilities with discussion-leading practices.

Theoretical Framework

Our work is grounded in three bodies of work – conceptions of teaching and learning, teacher learning, and discussion-leading practice. We conceive of quality teaching as critical in advancing mathematics education. Mathematics teaching is something that people do; it is not merely something to know. Teachers must use knowledge flexibly and fluently as they interact in specific contexts with students, with the aim of helping those students become proficient with mathematics. Conceptualizing the work of teaching as interactions among teachers, students, and content, in environments (Cohen et al., 2003) has important implications for the design of PD. It means that simply knowing a lot of math is not enough to be able to help students learn mathematics. Teachers must understand mathematics content and practices in specialized ways that are attuned to their learners. Simultaneously, teachers must be able to engage in teaching practices that enable them to interact with subject matter and students in multiple organizational formats, and support students' interactions with the mathematics being studied. Furthermore, the complex nature of mathematics teaching and the ever-changing nature of the environments in which teaching happens, also require a commitment to, and means of, professional learning in, from, and for practice. Thus, teaching requires the integrated use of specific knowledge and skills in particular contexts of instruction. This emphasis on (a) knowledge use, (b) attention and responsiveness to learners, and (c) context forms the basis of our approach.

Decades of research have demonstrated that many approaches to PD do not adequately support increases in teachers' capabilities (e.g., Coalition for Evidence-Based Policy, 2013; Cohen & Hill, 2001; Garet et al., 2001; Jacob et al., 2017). Often, teachers participate in a patchwork of sessions where presenters share ideas and raise enthusiasm without sufficient focus on teachers' opportunities to develop mathematical knowledge (Little, 2001; Wilson & Berne, 1999), adequate engagement in practice (Darling-Hammond, 1998), or connection with practice (Ball & Cohen, 1999; McLaughlin, 1990; Olson & Craig, 2001). Typically, U.S. teachers' PD opportunities are neither sufficiently "curricular" nor effective for improving practice. In contrast, the structure of our PD reflects the five elements of effective PD identified by Desimone and Garet (2015). First, our PD had a "content focus," --the practice of leading discussions in elementary mathematics, focusing on number and operation, which spans elementary grades. Second, the PD was designed to be "active" with opportunities to engage in leading discussions. Third, the PD was "coherent" in focus, centered on building skill with leading mathematics discussions. Fourth, the 36 hours of PD is aligned with the "sustained duration" associated with positive outcomes for teachers and students (Lynch et al., 2019; Yoon et al., 2007). Finally, "collective participation" was embedded across the PD.

The PD was designed to support the learning of teaching practice by situating learning inside of elementary mathematics classroom instruction. The PD used a classroom as a "common text" for working on practice, where participants were not only watching and discussing, but were engaged in developing and learning practice. This common text creates opportunities for participants to 1) examine lesson plans and student tasks, 2) discuss the plans and tasks with the teacher of record, making adjustments together, 3) observe the enactment of the discussed lesson, including deviations from and adjustments to the planned lesson, 4) collaboratively review students' daily work in response to the teaching, and 5) debrief the teaching of the lesson, unpacking together adjustments and deviations from the plan, the mathematics that unfolded during the lesson, and moves the teacher made to engage students in mathematics discussion (see Shaughnessy et al., 2017).

In addition to engaging in this collective legitimate peripheral participation, the PD involved a focused work designed to improve skill with leading mathematics discussions. Our definition of mathematics discussions entails a period of sustained dialogue among students and the teacher where students respond to and use one another's ideas to develop collective understanding. This requires teachers to engage in areas of work, including launching and concluding the discussion, recording and representing content, and orchestrating the discussion by eliciting and probing student thinking, orienting students to one another's thinking, and making contributions. For a detailed description of this work, see Shaughnessy et al. (2021). Our study investigates the impact of the described PD on teachers' enacted discussion-leading practice.

Methods for Studying Professional Development Impact on Practice

Participants

The full study spanned three iterations of the professional learning involving variations of the setting of the PD (on site or remote) and the components of the PD (peripheral participation with and without the focused discussion-leading session). This paper focuses on three participant groups from iterations 1 and 2 and reports on the impact of the full PD (both peripheral participation and the focused discussion-leading session) on their discussion-leading practice.

Group H1 consisted of 12 teachers from one school district in the midwestern United States, 10 of whom had completed more than three years of mathematics teaching. This group attended on site. Two researchers co-facilitated the practice-based discussion session.

Group H2 consisted of 10 teachers from one school district in the midwestern United States (different from the district in Group H1), seven of whom had completed more than three years of mathematics teaching. This group attended on site and one of the facilitators from Group H1 led the practice-based discussion session.

Group A2 consisted of 11 teachers from a large urban district in the northeastern United States, nine of whom had completed more than three years of mathematics teaching. This group gathered in one location with the other Group H1 facilitator to live-stream and engaged via a chat function in the peripheral participation portion before engaging in the practice-based discussion session with the facilitator.

Data Sources

The PD goals were multi-faceted, taking on many aspects critical to the teaching of mathematics. These goals included teaching practice, mathematical knowledge for teaching, language for talking about teaching, and skill in close observation of teaching and students. However, the primary goal was to increase participants' skills with leading mathematics discussions. The additional foci of enhancing language for talking about teaching, building skill with close observation, and building mathematical content knowledge were in service of this primary goal of improved discussion-leading practice.

This focus on improving mathematics discussion-leading practice necessitated the collection of participants' mathematics discussion-leading practice before and after the intervention. Six videos were collected from each of the 33 participants in the months immediately preceding and immediately following the PD, three pre-PD (in the last two months of school prior to the PD) and three post-PD (in the first two months of school). The set of three videos consisted of two lessons identified by participants as examples of mathematics discussions and one lesson provided by the research team. The selection of mathematical tasks used in a discussion and the content focus impacts the moves that teachers can make when leading discussions as well as the overall trajectory of the discussion, so the third video record of a mathematics lesson used a provided task accompanied by a detailed lesson plan. The video recording of the provided lesson also allowed for a more direct comparison across participants. The topic, task, and lesson structure were constant across participants and across time periods (pre and post). A small number of videos were not included in the analysis. Four videos from Group H1 and one video from A2 were not included in the analysis due to the format of the lesson (lecture only, completing practice problems, or small group instruction only).

Data was also collected around the supporting foci, both to measure incoming skill in these areas and to measure potential changes resulting from the intervention. One such measure was utilized for examining changes in teachers' mathematical knowledge for teaching. The Learning Mathematics for Teaching survey (LMT, 2008), was selected as a pre- and post-intervention measure of teachers' mathematical knowledge for teaching. The PD itself had multiple content strand foci as the PD encompassed both peripheral participation in a grade 5 mathematics program focused on fractional reasoning, operations with integers, and mathematical argumentation as well as a practice-based discussion session focused on number and operations. The trajectory of the classroom content is naturally dependent on students' progression and uptake of the content alongside the teachers' decisions about the trajectory of the work. For that reason, an LMT form focused on the Number and Operations was used. The LMT was administered in the same time frame as the video collection.

Selection of Tools for Data Analysis

Many measures exist for evaluating student outcomes of mathematics lessons, but fewer measures exist for measuring mathematics teaching practice, specifically mathematics discussion-leading practice. To determine changes in both the overall quality of the mathematics and in the specific moves that teachers were using in their discussion-leading practice, tools that are oriented on these qualities were needed.

Mathematical Quality of Instruction. The Mathematical Quality of Instruction instrument was selected to measure the overall mathematical quality of the lessons (Hill et al., 2008). This tool is not aligned with any particular orientation to the teaching of mathematics (e.g., reform-oriented instruction) or to a particular lesson structure, but rather to the depth and quality of the mathematics content available to students during instruction allowing us to discern, for each lesson, the richness of the mathematics, the precision and accuracy of the mathematics instruction, students' involvement with content through Common Core-aligned mathematical practices, and the teacher's ability to interpret and respond to students' mathematical ideas.

Discussion Leading Checklist. A complimentary tool was designed to examine teachers' specific discussion-leading moves. In prior work, a discussion-focused checklist tool was developed to formatively assess beginning teachers' practice. Using the tool entailed recording the presence or absence of particular discussion-leading moves to support the noticing of the work done by teachers in a specific discussion. A previous study showed that the tool was able to reveal variations in practice across teachers and provide fine-grained detail about the skill of teachers in leading mathematics discussion while also accounting for existing classroom norms (Shaughnessy et al., 2021). Because our PD involved practicing teachers, we expanded the list of possible moves in the checklist to more accurately capture the range of moves made by experienced teachers. These additions included advanced areas of work, such as supporting students to make connections and extending and revising student ideas. The checklist tool was designed to capture the presence or absence of particular moves. The tool was not designed to judge the skill with which the move was enacted or students' responses to the move. This is in tension with our desire to capture the quality of teaching. However, the decision was made to account for teachers' attempts to enact different discussion leading moves as we are considering the learning of teachers. Our main goal is to determine whether teachers are trying out new work to improve their practice. We know that trying out new work often results in less than perfect outcomes, particularly in early implementation. We did, however, want to capture problematic implementation to determine whether problematic enactment increased, decreased, or stayed the same before and after the intervention. To that end, the checklist tool also included a section for "issues" such as consistently ineffective probes, as well as an overarching rating indicating whether the instruction in the video was aligned with our definition of a discussion. This overarching rating is meant to serve as an acknowledgement that discussion-leading is more than the sum of the parts that can be captured in a checklist tool.

The tool necessarily focuses on specific observable moves that teachers make while leading a discussion, categorized both by the phase of the work (e.g., task set-up, discussion launch, orchestration, conclusion) and by the buckets of work that teachers must do (e.g., eliciting and probing student thinking, orienting students to one another's thinking, supporting students to make connections). This organization allowed for observations about the density of work in a particular area (e.g., the number of different moves a teacher used to orient students to one another's thinking) and the spread of work across a discussion (e.g., whether a teacher was doing work in all possible areas or if the work was concentrated in one or two areas). Each observable move was marked as present, not present, or not applicable (NA). The code of NA was used judiciously in cases where the move was unnecessary. For example, NA could be used if the task selected precluded a range of responses or methods to be shared or if the teacher did not need to further probe student contributions due to the level of detail of the contribution. Many moves had possible available codes of "once," "more than once," and "student initiated" to support the ability to analyze the frequency and the extent to which established norms seemed to be influencing the work of the teacher. For our analysis, these three codes were collapsed to present.

Data Analysis Methods

Mathematical Quality of Instruction. The 193 videos in this study were coded using the discussion checklist tool and the MQI. MQI coders attended MQI training and were certified as raters. Videos were double-coded for both instruments using associated coded books. Coders met to reconcile discrepancies for each video prior to entering final codes and resolved differences via discussion and referencing the codebooks. MQI coding was applied to equal-length chapters of the full video (approximately 7.5-minute segments), with the number of chapters determined by the length of the video. Each chapter was scored on five domains. The full video was scored on nine factors and assigned one overall score for the lesson. Chapter scores were used to create aggregate video scores for the chapter-specific points. In addition, aggregated pre- and post-intervention MQI scores were created for each participant using mean scores.

Discussion Checklist. The discussion checklist coding was applied to videos. Composite scores were created for each area of work represented in the discussion checklist tool. For those areas where work was consistently possible, sum scores or scaled mean composite scores were used to represent the amount of possible work that occurred relative to the available work. For example, an Orienting composite variable was created using the sum of the "present" moves. There were five possible moves - posing questions to students about others' ideas, asking students to restate another's idea, responding to another student's idea, adding to an idea, and interpreting the strategies of others – that could be observed. For eliciting, a scaled mean score was used to represent that there were four possible moves - eliciting multiple ideas, eliciting a range of responses, engaging several students in sharing ideas, and eliciting students' mathematical processes. A scaled mean was used to enable comparison between teachers who had all four moves available and those who only had three moves available due to task selection (e.g., a range of student responses is unlikely given the task). Composite variables included task setup, discussion launch, eliciting, probing, orienting, generic orienting, connecting/ extending/ revising, concluding, and issues. A detailed description of the checklist items associated with each composite score can be found in Table 1.

Table 1: Composite Variable Descriptions

Variable	*Description*
Task Set Up	Composite score takes into account the work the teacher does to help students make sense of the task (including reading, restating, and unpacking context), whether the teacher maintains cognitive demand of the task, and whether the students are sufficiently prepared to begin the mathematical work.
Launch	Sum score takes into account whether the teacher had the attention of all students before launching the discussion, is concise and on-task during the launch, and explicitly states the goal of the discussion.
Eliciting	Scaled mean score takes into account the work of eliciting multiple ideas, eliciting a range of responses, engages several students, asks about processes. Because not all tasks allow for a range of responses, a conditional mean was calculated based on whether a range was possible, then the mean was scaled.
Probing	Score takes into account the work of posing questions to get students to explain their understanding of relevant mathematical content or processes, or follow-up questions focused on why a student did particular work.
Orienting	Sum score included posing questions to students about others' ideas, asking to restate another's idea, responding to another's idea, adding to an idea, and interpreting the strategies of others.
Generic Orienting	Sum score accounted for generic orienting moves such as encourages the class to attend/listen/respond, uses turn-and-talk to encourage discourse, elicits student to student discourse, and uses moves that require all to respond to others' work.
Revising	Score focuses on moves to supports students to revise their work.
Connecting/ Extending/ Revising	Composite mean takes into account moves including asking students to identify similarities/differences, supporting students to connect to past work, supporting students to connect to the problem context, supporting students to connect between representations, examining efficiency, generating hypothetical situations to extend thinking, and supporting students in revising their thinking.
Concluding	Sum score accounting for the teacher making a closing statement, supporting students in remembering a key idea, and taking stock of the discussion.
Issues	Sum score accounting for problematic moves such as inaccurate revoicing, mathematically problematic contributions, over-scaffolding, consistently using ineffective probes.

Aggregated pre- and post-intervention discussion checklist scores were created for each participant using mean scores. Kruskal-Wallis tests and independent samples tests were performed on pre-data to determine whether differences between participants by study group existed. Significantly different distributions were found for study groups, indicating that separate analysis of these groups is necessary. For each group of teachers, independent samples tests were examined for the discussion checklist score pre and post PD. These tests included Levene's test for equality of variances and *t*-tests for equality of means. Paired samples *t*-tests for aggregated case data were examined by study.

Findings

Examining the data disaggregated by study group, we noticed differences in outcomes of three groups who received the full PD. Next, we unpack the features of these groups, their experiences, and their outcomes to investigate factors that influence participant outcomes.

Group H1. The first group of participants received PD in the first iteration of the study. The content of the peripheral participation experience focused squarely on setting up tasks for productive work, eliciting and probing student thinking, and supporting students to engage with others' ideas, the foundation for orienting students to one another's thinking. The participants observed minimal full-class mathematics discussion-leading work. The student group was not accustomed to engaging in discussions and the teacher had extensive groundwork to do to move toward productive discussions. The focused discussion-leading session supported participants in building their knowledge and skill with launching discussions, orienting students to one another's thinking, representing and recording content, and concluding discussions.

This group entered the professional development with a group LMT score of approximately 0.51 standard deviations above the mean. Overall, few discussions were led prior to the PD, with approximately 20% of the submitted pre-videos rated as aligned to our definition of discussion. Despite a low number of discussions being led, the mean score of the submitted lessons (M = 3.222, SD = 0.43) on the MQI Whole Lesson Mathematical Quality of Instruction was slightly above the middle rating of 3.0 on a 1-5 scale.

Following the PD, this group showed a nearly significant (p = 0.053) increase on paired t-tests in the number of discussions led, with approximately 36% of submitted post-videos demonstrating a discussion. This group also showed significant increases in task set-up (p = 0.012), eliciting (p = 0.009), and concluding (p = 0.016). Nearly significant increases were shown in probing (p = 0.077), generic orienting (p = 0.084), and revising (p = 0.056).

Group H2. The second group of participants received PD in the second iteration of the study. The content of the peripheral participation focused on the full spectrum of discussion-leading practice, beginning with work on setting up tasks for discussion and eliciting student thinking, followed by orienting students to the thinking of others, with the sophistication of orienting moves increasing across the experience. The content of the focused discussion-leading session was identical to the first iteration.

This group entered the professional development with a group LMT score of approximately 0.27 standard deviations below the mean. Overall, few discussions were led prior to the PD, with approximately 15% of the submitted pre-videos rated as aligned to our definition of discussion, a slightly lower percentage than Group H1. The mean score of the submitted lessons (M = 2.7407, SD = 0.49) on the MQI Whole Lesson Mathematical Quality of Instruction was slightly below the middle rating of 3.0.

Following the PD, the group showed no significant difference in the number of discussions (p = 0.347) or any category captured by the checklist tool. Near significant increases were seen in generic orienting (p = 0.069).

Group A2. The third group of participants received PD in the second iteration of the study. The peripheral participation was completed via live-stream and the focused discussion-leading session was delivered by the facilitator who co-facilitated the session in the first iteration of the study. The content of the peripheral participation was identical to that of Group H2.

This group entered the PD with a group LMT score of approximately 0.65 standard deviations above the mean. A high number of discussions were led prior to the PD, with approximately 89% of the submitted pre-videos rated as aligned to our definition of a discussion.

Additionally, the mean score of the submitted lessons ($M = 3.8611$, $SD = 0.54$) on the MQI Whole Lesson Mathematical Quality of Instruction was well above the middle rating of 3.0.

Following the PD, the group showed no significant difference in the number of discussions or categories captured by the tool. Nearly significant increases were seen in connecting/extending ($p = 0.055$) and revising ($p = 0.080$).

Discussion and Implications

There were clear differences in the outcomes of the three study groups, with study group H1 showing the most gains from the PD. We hypothesize three potential reasons for these differences. The first potential hypothesis focuses on the content of the peripheral participation. Group H1 experienced peripheral participation where the discussion-leading work of the teacher was slower in pace, more repetitive, more deliberate, and more challenging than that of the peripheral participation of Group H2 and A2. It is possible that the pacing and repetition made this work more accessible to participants than the smooth, quickly progressing work of iteration 2, thereby allowing participants to consider the specific moves the teacher was making and how they might be incorporated into their own classrooms. They may also have identified directly with the challenges of this work and were ready for the same challenges in their own classrooms.

A second possible hypothesis for the differences is related to incoming knowledge and skills. Groups H1 and A2 both entered the PD with above average mathematical knowledge for teaching and above average mathematical quality of instruction as measured by the whole lesson MQI score. Group H1, however, was not engaged in consistent discussion-leading, unlike Group A2. One explanation for the changes seen in Group H1 is that they had the mathematical knowledge to lead discussions and strong general mathematical instructional skill, but did not yet have discussion-specific skills, but improved in this area following the PD. In contrast, Group A2 brought mathematical knowledge, general mathematical instructional skill, and high discussion-leading skill and so they had less room for improvement. Group H2, on the other hand, entered with low mathematical knowledge, below average mathematical quality of instruction as measured by the whole lesson MQI score, and the lowest level of discussion leading practice of the three groups. It is possible that low mathematical knowledge combined with below average mathematical quality of instruction meant that these teachers were not yet ready for the pace of the PD or to focus on discussion-leading practice.

One final hypothesis is related to contextual factors, which were not measured by the instruments described in this paper. The teachers in Groups H1 and A2 were in consistent school environments where their grade level assignments were stable and their administrators appeared to support their involvement in the PD. Participants in Group H2, on the other hand, discovered during the PD that they would all be changing grade-level assignments due to administrators' decisions based on student outcome data. Learning about this change likely distracted their focus and a grade change can be challenging for implementing new techniques.

Further analysis is needed to determine which of these factors most impacted the PD outcomes. However, the current results do point to the importance of teachers' incoming knowledge and skills to the outcomes of mathematics professional learning.

Acknowledgements

This work was supported by the National Science Foundation under Grant No. 1621104. Any opinions, findings, and recommendations expressed are those of the authors and do not reflect the views of the National Science Foundation.

References

Ball, D. & Cohen, D. (1999). Developing practice, developing practitioners. In L. Darling-Hammond and G. Sykes (Eds), *Teaching as the learning profession* (pp. 3–22). Jossey-Bass.

Ball, D. L., & Forzani, F. M. (2009). The work of teaching and the challenge for teacher education. *Journal of Teacher Education, 60*(5), 497–511.

Coalition for Evidence-Based Policy. (2013). Randomized controlled trials commissioned by the Institute of Education Sciences since 2002: How many found positive versus weak or no effects. Retrieved from http://coalition4evidence.org/wp-content/uploads/2013/06/IESCommissioned-RCTs-positive-vs-weak-or-null-findings-7-2013.pdf.

Cohen, D. K., & Hill, H. C. (2001). *Learning Policy: When State Education Reform Works*. Yale University Press.

Cohen, D. K., Raudenbusch, S., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, *25*(2), 119–142.

Darling-Hammond, L. (1998). Teacher learning that supports student learning. *Educational Leadership, 55*(5), 6–11.

Desimone, L. M., & Garet, M. S. (2015). Best practices in teacher's professional development in the United States. *Psychology, Society, & Education*, 7(3), 252-253. http://repositorio.ual.es/bitstream/handle/10835/3930/Desimone%20En%20ingles.pdf?s

Garet, M. S., Porter, A. C., Desimone, L., Birman, B., & Yoon, K. S. (2001). What make professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915–945.

Hill, H.C., Blunk, M., Charalambous, C., Lewis, J., Phelps, G.C., Sleep, L., & Ball, D.L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction*, *26*, 430–511.

Jacob, R., Hill, H., & Corey, D. (2017). The impact of a professional development program on teachers' mathematical knowledge for teaching, instruction, and student achievement. *Journal of Research on Educational Effectiveness, 10*(2), 379–407.

Learning Mathematics for Teaching. *LMT measures of Mathematical Knowledge for Teaching: Elementary Number Concepts and Operations*, 2008. Ann Arbor, MI: Authors.

Little, J. W. (2001). Professional development in pursuit of school reform. *Teachers caught in the action: Professional development that matters*, *3*, 23–44.

Lynch, K., Hill, H.C., Gonzalez, K.E., & Pollard, C. (2019). Strengthening the Research Base That Informs STEM Instructional Improvement Efforts: A Meta-Analysis. *Educational Evaluation and Policy Analysis, 41*(3), 260–293

McLaughlin, M. (1990). Staff development and school change. In A. Lieberman (Ed.), *Schools as collaborative cultures* (pp. 213–232). Teachers College Press**.**

National Council for Accreditation of Teacher Education (NCATE). (2010). *Transforming Teacher Education through Clinical Practice: A National Strategy to Prepare Effective Teachers.* Author.

National Commission on Teacher Quality (2008). *No common denominator: The preparation of elementary teachers in mathematics by America's education schools*. Author.

Olson, M., & Craig, C. (2001). Opportunities and challenges in the development of teachers' knowledge: The development of narrative authority through knowledge communities. *Teaching and Teacher Education, 17*, 667–684.

Shaughnessy, M., Ball, D. L., & Garcia, N. (2017). A laboratory approach to the professional development of elementary mathematics specialists. In M. B. McGatha & N. R. Rigelman (Eds.), *Elementary mathematics specialists: Developing, refining, and examining programs that support mathematics teaching and learning* (pp. 123–131). IAP.

Shaughnessy, M., Garcia, N. M., O'Neill, M. K., Selling, S. K., Willis, A. T., Wilkes II, C. E., Salazar, S. B., & Ball, D. L. (2021). Formatively assessing prospective teachers' skills in leading mathematics discussions. *Educational Studies in Mathematics*, *108,* 451–472.

Wilson, S. M., & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. *Review of research in education*, 173–209.

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). Reviewing the evidence on how teacher professional development affects student achievement (Issues & Answers Report, REL 2007-No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.