STANDARDIZED ASSESSMENTS OF BEGINNING TEACHERS' DISCUSSION LEADING PRACTICE: IS IT POSSIBLE AND WHAT CAN WE LEARN?

Sarah Kate Selling
University of Michigan
sselling@umich.edu

Meghan Shaughnessy
University of Michigan
mshaugh@umich.edu

Amber Willis
University of Michigan
atwillis@umich.edu

Nicole Garcia
University of Michigan
nmgarcia@umich.edu

Michaela Krug O'Neill
University of Michigan
mkoneill@umich.edu

Deborah Loewenberg Ball
University of Michigan
dball@umich.edu

This paper examines the possibilities of using standardized assessments to assess elementary and secondary novice teachers' skills with leading problem-based mathematics discussion. The findings reveal that our standardized assessments were able to elicit and reveal variations in skills across teachers; provide fine-grained detail about the performance of particular teachers for feedback purposes; and account for existing classroom norms. This suggests that such tools could be useful in assessing and supporting beginning teachers.

Keywords: Teacher Education-Preservice; Instructional Activities and Practices; Assessment and Evaluation; Classroom Discourse

Theoretical Framing

The preparation of beginning teachers has received increased attention in recent years with calls to ensure that beginners are prepared to take responsibility for all students' learning upon entry to the teaching force. This call has spurred changes towards preparing beginners to do core tasks of teaching. In a number of teacher preparation programs, course content is shifting to focus on high-leverage teaching practices such as eliciting and interpreting student's thinking and leading discussions (Ball, Sleep, Boerst, & Bass, 2009; Davis & Boerst, 2014; McDonald, Kazemi, & Kavanagh, 2013). This shift means that teacher education programs must also develop ways to assess the doing of teaching. Many current assessments of teaching, including observation tools (e.g., Danielson, 2011) and portfolios (e.g., Darling-Hammond, 2010), offer information about beginning teachers' skills with respect to broad domains of work like planning, instruction, or assessment. But we need additional tools to improve estimates of beginners' skill with particular practices of teaching.

One practice that has received considerable attention in teacher education is the practice of leading a discussion in mathematics. Mathematics discussions are important for supporting students in developing conceptual understanding (e.g., Michaels, O'Connor, & Resnick, 2008) and learning disciplinary norms and practices (e.g., Lampert, 2003; Yackel & Cobb, 1996).For these reasons, a number of current efforts are focused on helping novice teachers learn to enact this complex practice (Boerst, Sleep, Ball, & Bass, 2011; Lampert et al., 2013). Given this, teacher educators will also need ways to assess novices' skill with leading mathematics discussions, beyond relying on plans for leading a discussion, reflections, or analysis of others' enactment (e.g., through video analysis).

Many factors beyond sheer skill influence novices' enactment of specific practices. As teachers lead whole class discussions, for example, factors related to students, the content, and the environment, matter for the unfolding of the discussion (Cohen, Raudenbush, & Ball, 2003). Teachers' knowledge of the content, the "discussability" of the mathematics task, students' prior experiences participating in discussions, and teachers' knowledge of and relationships with their students all shape how discussions play out. All of this can make it difficult to appraise novices' skills in ways that are comparable and fair. We sought to investigate whether it is feasible to design a

Bartell, T. G., Bieda, K. N., Putnam, R. T., Bradfield, K., & Dominguez, H. (Eds.). (2015). *Proceedings of the 37th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. East Lansing, MI: Michigan State University.

standardized assessment that is capable of: (a) eliciting and revealing variations in skills across teachers; (b) providing fine-grained detail about the performance of particular teachers for feedback purposes; and (c) accounting for existing classroom norms. Below we begin by articulating and specifying what we mean by a "mathematics discussion" then turn to the describing the assessment development.

Work of leading a discussion

We grounded our assessment development in a particular decomposition (Grossman et al., 2009) of the practice of leading a mathematics discussion. We define classroom discussion as "a period of relatively sustained dialogue among the teacher and multiple members of the class" in which students respond to and use one another's ideas to develop collective understanding (TeachingWorks, 2015).Our decomposition is informed by research on practices for orchestrating productive discussions (Smith & Stein, 2011), the concept of talk moves (Chapin, O'Connor, & Anderson, 2013), and research on decomposing practices so that novices can learn them (Boerst et al., 2011). We differentiate *discussion-enabling practices* (Boerst, Moss, & Blunk, 2009), such as anticipating student thinking and monitoring student work, and *discussion- leading practices* used to manage a discussion, such as eliciting student thinking. We distinguish between three stages of the discussion: launch, orchestration, and conclusion. The launch comprises the work a teacher does to frame the discussion (Engle, 2006) and is separate from the actual setting up of the task. Within each stage, teachers engage in particular discussion-leading practices. For example, orchestration includes practices such as *eliciting contributions* and *probing student thinking*. Figure 1 shows our decomposition of discussion-leading, as well as examples of discussion-enabling practices. Within each discussion-leading practice, we further identify techniques (Boerst et al., 2011), including particular talk moves.

<table>
<tr><th>Discussion Enabling</th><th colspan="3">Discussion Leading</th></tr>
<tr><td rowspan="2">Anticipating student thinking
Setting up the problem
Monitoring student work</td><td>Launching</td><td>Orchestrating
- Eliciting
- Orienting
- Probing
- Making Contributions</td><td>Concluding</td></tr>
<tr><td colspan="3">Recording</td></tr>
</table>

Figure 1: Decomposition of Leading a Mathematics Discussion

Assessing beginning mathematics teachers' discussion leading practice

Because beginning teachers are more often assessed on their ability to plan, we set out to design an assessment focused squarely on the interactive work of leading a discussion. We also wanted the assessment to be classroom-based, since so much of the work of leading a discussion involves responding to student thinking in the moment. We wanted our assessment to elicit and capture both a range of and variation in performance, as well as provide fine-grained detail about the demonstrated skills of individual teachers. Because we intended to use the assessment in real classrooms, it was important that it account for variations in classroom norms and grade levels.

With these goals in mind, we developed parallel elementary and secondary assessments to be implemented in novice teachers' classrooms. Simultaneously, we sought to design the assessments with as much standardization as possible to allow for comparing teachers' performances. Because discussion-enabling practices are critical to the work of preparing for a discussion (e.g., Jackson, Garrison, Wilson, Gibbons, & Shahan, 2013), we provided supports for teachers to prepare for the discussion so that our assessment could focus squarely on their discussion-leading skill. Both the elementary and secondary level assessments required teachers to lead problem-based mathematics discussions, in which students work on a mathematical task and then participate in a discussion about their work on the task. We selected mathematical tasks that have been extensively piloted in K-12 classrooms and can be found in existing materials (Ball & Shaughnessy, 2014; Mathematics Assessment Resource Service, 2012). Table1 provides an overview of each of the tasks. These tasks offer opportunities to compare and connect a range of solutions, strategies, or approaches, which represents one discussion structure (Kazemi& Hintz, 2014). The tasks were also selected because they focus on high-leverage mathematics content (Ball & Forzani, 2011), can be used across grade bands, and because we conjectured that they could be implemented as a "drop-in lesson" without necessarily being connected to the current instructional focus of a classroom.

Table 1: Elementary and Secondary Mathematics Tasks

Elementary: You will lead a discussion of solutions to the problem, "Make number sentences for 10," a problem with multiple solutions. The students will first work independently or in pairs on the problem. Then, you will lead a discussion of the task. The goal of the discussion will minimally be to elicit several solutions to the problem and to have students explain why they are or are not solutions and to notice similarities and differences among the solutions. The focus for your discussion will vary depending on student solutions. A table of possible student solutions is included to help you organize your discussion.	*Secondary:* You will lead a discussion of solutions to the problem "Proofs of the Pythagorean Theorem?" The students will first work on the task, then you will lead a discussion of the task. The task asks students to evaluate several attempts at proving the Pythagorean Theorem. The goals of the discussion include eliciting several solutions to the problem, ensuring that the merits and limitations of each solution are explained, and identifying features of a strong mathematical proof. It is not necessary to complete a standard proof of the Pythagorean Theorem.

By specifying the mathematics tasks, we sought to control for the variation that would arise if teachers led discussions on different topics. This was important for achieving our goal of designing a tool that could compare performances. Prior experiences teaching methods courses led us to believe that providing a discussable task was a support that could enable the assessment to focus on assessing teachers' discussion-leading skills.

Grades 2-3	
Generate number sentences for 10 using multiple operations determine whether the value of a proposed expression is 10 listen to classmates and provide justification for agreeing or disagreeing with specific explanations begin considering the idea of infinitely many solutions make a claim regarding the number of possible solutions with justification	Students will likely generate number sentences for ten that move beyond integer pairs that sum to 10 (e.g. 1 + 9). Their expressions are likely to contain multiple terms and are likely to incorporate subtraction, although some students may need prompting The types of expressions that students might generate include: 1 + 1 + 1… = 10, 2 + 3 + 5 = 10, 11 – 1 = 10, 83 – 73 = 10 Students might notice number patterns that allow them to generate many expressions quickly, e.g. 12 - 2, 22 - 12, 32 - 22…. Students may also incorporate multiplication and division. If students use subtraction, multiplication and/or division, they are likely to think that there is a very large finite number of solutions to the problem . Students may state that the number of solutions to the problem is infinite.

Figure 2: Support for Differentiation of Learning Goals and Anticipating Student Thinking

We also built in supports around other discussion-enabling practices. For example, the assessments included supports for understanding the mathematics, anticipating student thinking (Smith & Stein, 2011), and adapting the task for different grades. Figure 2 shows an excerpt of the support materials

Bartell, T. G., Bieda, K. N., Putnam, R. T., Bradfield, K., & Dominguez, H. (Eds.). (2015). *Proceedings of the 37th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. East Lansing, MI: Michigan State University.

for differentiating the goals of the task and anticipating student thinking. Both assessments also included lesson plans with suggested timing, participation structures, and guidance for setting up the task.

We also developed a scoring tool that could be used to analyze and assess videos of teachers' leading these discussions. This tool was designed to focus on the discussion-leading practices core to our decomposition of the work: launching a discussion, eliciting responses, probing students' thinking, orienting students to one another's contributions, making contributions to the discussion, recording/representing content, and concluding the discussion. Our tools also focused on specific techniques. These techniques were selected for their importance to skillful beginning teaching. For example, "orienting students to the contributions of peers" had five associated techniques: (1) students are prompted as needed to talk to the whole class; (2) the teacher poses questions to students about others' ideas and contributions; (3) students are asked to comment on, add to, or restate other students' ideas; (4) listening is supported by moves that ask all students to respond to one another's work; and (5) students are encouraged to listen, and respond to maintain productive and focused interactions. We included this level of detail in the scoring tool to reflect our decomposition of practice and to meet the goal of developing an assessment that would allow us to capture differences in performance across and within performances. Importantly, we were assessing teachers' skills in responding to student contributions, not whether students responded in a particular way (e.g., students might not respond to a well-formulated question intended to probe their thinking).

In developing the scoring tool, we recognized that there are many ways to score or assess instruction, including rubrics that differentiate levels of performance in particular domains. We chose to use threshold statements to record the presence or absence of particular techniques with respect to a defined threshold. This choice motivated the goal of using of the tool for identifying whether teachers were enacting techniques in the different practices. This format could be particularly useful in identifying patterns across groups of teachers as well as within an individual teacher's performance. Additionally, we built in a "not applicable (N/A)" choice because some specific techniques may not be needed in particular cases. Finally, we developed a section that captures common issues that may arise in discussions. A codebook was developed with definitions of each technique, examples of what it was, and examples of what it was not.

Methods

We piloted the assessment with 17 first-year teachers (9 elementary teachers across grades 1-5 and the 8 secondary teachers across grades 7-12). We recruited a diverse sample of teachers with respect to grade level, school district, and teacher preparation program. The purpose of the pilot was to gather data from a range of first-year teachers and was not intended to be representative of all first year teachers. We provided the plan for the discussion and gave participants 45 minutes to prepare.

The discussion was video-recorded by the research team. One camera was set up in the back of the classroom and focused on teacher and student interactions. Discussions ranged from approximately 15 to 45 minutes in length, although the full lessons were longer. Participants also completed a background survey indicating their prior experience and training in discussion leading practices. All teacher names used in the subsequent sections are pseudonyms.

The analysis of the discussions was conducted by the research team through independently watching and scoring each video using the tool described previously. Then the team, comprised of members with expertise in both elementary and secondary mathematics teaching, met together to discuss the scoring and reach a final consensus. When there were discrepancies across scores, the full team examined the video and resolved the issue, referencing the codebook as needed. A subset of the videos (>20%) was additionally coded by a trained rater, yielding an inter-rater agreement of 85% (Miles & Huberman, 1994).

Bartell, T. G., Bieda, K. N., Putnam, R. T., Bradfield, K., & Dominguez, H. (Eds.). (2015). *Proceedings of the 37th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. East Lansing, MI: Michigan State University.

Findings

We analyzed the 17 discussions to examine how well the assessment was able to: (a) elicit and reveal variations in demonstrated skills across teachers; (b) provide fine-grained detail about the performance of particular teachers for feedback purposes; and (c) account for existing classroom norms. We address each of these in the following sections.

Eliciting and revealing variation in demonstrated skills

In looking across all performances elicited by the assessment, we found that the scoring tool was able to reveal a range of skill with respect to all of the discussion leading practices. For example, within the domain of orchestration (including the practices of eliciting, probing, orienting, and making contributions), the assessment and tool revealed skills that varied from a minimum of three demonstrated techniques to performances that revealed use of all 15 techniques associated with orchestration. Additionally, the assessment and tool was able to elicit and reveal a range of performance within each discussion leading practice. For example, Table 2 shows a range of performance on each of the four discussion leading practices used during the orchestration stage of the discussions.

Table 2: Range of performance within discussion leading practices

	Eliciting	*Probing*	*Orienting*	*Contributing*
Range of techniques	1/3 to 3/3	0/4 to 4/4	0/5 to 5/5	1/3 to 3/3

Our study suggests that the assessment and scoring tool are capable of revealing variation in performance. We found that the tool was capable of revealing variation in skills at multiple levels, including higher-level variations between discussion leading practices and more fine-grained variations in performance within particular discussion leading practices. As an example of this variation, within this sample, teachers demonstrated greater skill at eliciting student thinking than probing student thinking or orienting students to the contributions of peers. All of the elementary teachers demonstrated skill with every eliciting technique/move, and six out of eight secondary teachers did as well. In contrast, only two out of 17 teachers engaged in all aspects of orienting, many of them engaged in only some aspects, and a few teachers engaged in almost no orienting. This means that the tool enabled us to discern differences in the skills of groups of novice teachers with respect to discussion-leading practices.

The assessment and scoring tool were also able to capture fine-grained variations of performance within particular discussion-leading practices. For example, with the practice of *probing student thinking*, which included four different techniques, 14 out of 17 teachers engaged in probing students' mathematical processes, while only nine teachers probed students' understanding of key mathematical ideas. For example, one secondary teacher, Ms. Mason, first elicited student ideas about which of three attempted proofs of the Pythagorean Theorem is most valid and complete. A student replied "In attempt number one, you have to find the area of the squares in order to get the triangle in the middle". Ms. Mason responded by probing the student's process, "Okay, how do I find the area of a square?" The tool showed that almost all of her probing focused on process. In contrast, another secondary teacher, Mr. Jacobs, probed both for student process and understanding. In one case, as students were defending their choice of one potential proof as being the best, a student said "I said that number two…it explains things. There's a lot of information". Mr. Jacobs responded, "What do you mean by the most information?" probing the students' understanding of what "information" meant in the context of a proof. This suggests that the structure of the scoring tool used with the discussion assessment data was able to highlight differences in teachers' demonstrated skill with particular practices.

Providing fine-grained detail about the performance of particular teachers

The assessment and scoring tool were able to elicit and capture variations of performance within teachers' discussions. To illustrate this capability, we describe what the scoring tool revealed about the performance of one fourth grade teacher. Mr. Weber launched his discussion efficiently and framed the mathematics to be discussed, although he did not have all students' attention before beginning. Within the practice of eliciting, Mr. Weber enacted all of the techniques that we sought to capture, including eliciting multiple solutions/strategies. He demonstrated less skill in the practices of probing and orienting, posing no questions that probed students' processes or understanding. Mr. Weber also did little to orient students to each other's contributions. He occasionally reminded students to listen to each other and asked for signs of agreement/disagreement with a contribution, but he did not ask students to comment on, add to, or restate others' ideas. Mr. Weber did enact techniques within the practice of making contributions. He revoiced and asked questions to engage students in substantive discussion of the key ideas; however, the mathematical contributions he made did not enrich the core ideas of the discussion nor did they keep the discussion focused on the learning goals. Throughout he demonstrated skill with recording and representing content as he recorded students' number sentences on the board, although he did not always attend to the mathematical accuracy of his records. Mr. Weber concluded by taking stock of the discussion and worked to support students in remembering that they had determined that there could be infinitely many solutions to the task. This highlights how the assessment and scoring tool allowed us to capture detail and nuance within individual teachers' practice.

Accounting for classroom norms

Across the performances the scoring tool appeared to be able to account for variations in existing classroom norms. In some classrooms, students added on to another's thinking and spoke to the whole class without prompting from the teacher. The designation of N/A allowed us to recognize that students were doing these thing while still acknowledging that the teacher did not visibly employ moves that supported students in participating in this way. Yet, only two techniques received a substantive number (>3) of N/As. This indicates that the assessment and tool were able to elicit and capture demonstrated skill with the majority of the techniques. The two techniques that were frequently coded as N/A were (1) supporting students to speak to the whole class and (2) supporting students to listen to contributions of peers. Both techniques are used in the service of *orienting students to the contributions of peers*. The high frequency of N/A suggests that the standardized assessment and scoring tool may not be suitable for systematically eliciting and capturing teachers' skill in this area.

Discussion and Implications

We examined the utility of using a standardized assessment to assess novices' discussion-leading practice. The assessment and tool prompted and captured a range of performance and revealed variations across groups and within individual discussions. We also found that it was able to account for existing classroom norms.

To consider the implications of these findings, we begin by acknowledging the limitations of how we designed and piloted the assessment and scoring tool. One limitation is that the assessment and tool are grounded in a particular decomposition of the practice of leading a discussion. A different decomposition might reveal different capabilities of the assessment and scoring tool because if teachers are more familiar with other decompositions of the practice, they may perform

Bartell, T. G., Bieda, K. N., Putnam, R. T., Bradfield, K., & Dominguez, H. (Eds.). (2015). *Proceedings of the 37th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. East Lansing, MI: Michigan State University.

differently. A second limitation concerns the design of our scoring tool, which used threshold statements, yielding information about the presence or absence of each technique within a practice. This meant that the scores did not always distinguish the quality or quantity of teachers' enactments above the threshold. Finally, these analyses did not consider the validity of the assessment with respect to whether it accurately reflects a teacher's typical practice. Despite these limitations, our findings suggest that a standardized assessment of discussion-leading practice can reveal important information about teachers' skill. As we were not intending to make claims about this sample of teachers or a larger population of novice teachers, the small sample was appropriate for our goals. However, the variations in performance that are used to illustrate the capabilities of the assessment and tool cannot be interpreted as representing the skills of a larger population of teachers.

This standardized assessment accomplished many of our design goals. Its ability to capture a range of skill and to distinguish patterns across groups and within individual teachers' performances could make the assessment and scoring tool useful in teacher education. Teacher educators and programs could use such assessments to track teacher candidates' growth over time and to identify areas of strength and weakness with respect to the practice, which would allow for targeted support and program-level curriculum design. The use of scaffolds for particular parts of a lesson also showed promise for allowing the assessment to focus on a single instructional practice.

Additionally, the subject matter standardization allowed for comparison across performances, as well focusing the assessment squarely on discussion leading practices through providing standardized supports. Standardized mathematics content supported the efficiency and manageability in using both the assessment and scoring tool, as scorers did not need to familiarize themselves with the content and task of each discussion. Even with standardization of content and content supports built into the materials, there were instances in which subject matter knowledge appeared to be a factor in teachers' discussion-leading practice. In these cases it was unclear whether we were accurately capturing a teacher's skill with discussion leading or whether we measured indirectly his/her use of content knowledge in teaching. Subsequent work will need to take this into consideration.

We found that the practice of recording and representing content was challenging to assess using this assessment. Although the assessment materials asked teachers to record student contributions and other relevant mathematics, a number of elementary and secondary teachers involved students in the recording or relied completely on student-generated records. This may have been a reasonable choice given existing norms and routines and in some cases it appeared to support students in providing explanations to the class, but this choice made it difficult to see evidence of how these teachers record content. The materials may not have specified clearly enough that while students often contribute to recording, the assessment was asking teachers to record contributions. Future versions of the assessment will need to keep this challenge in mind; we see several different possibilities for addressing this challenge.

Future research could consider how well the assessment of novices' discussion-leading practice corresponds to their typical practice when leading discussions. Another important direction will be to further take up the questions about the role of subject matter knowledge in assessing teachers' skill with discussion-leading practices. Finally, further research could also investigate the impact of scaffolding materials for novices with different conceptions of what makes a mathematics discussion.

References

Ball, D. L., & Forzani, F. M. (2011). Building a common core for learning to teach: And connecting professional learning to practice. *American Educator*, *35*(2), 17.

Ball, D. L., & Shaughnessy, M. (2014). *Elementary Mathematics Laboratory Lesson Plan: Day One*. Mathematics Teaching and Learning to Teach, University of Michigan: Ann Arbor, MI.

Ball, D. L., Sleep, L., Boerst, T. A., & Bass, H. (2009). Combining the development of practice and the practice of development in teacher education. *The Elementary School Journal*, *109*(5), 458-474.

Boerst, T., Moss, P., & Blunk, M. (2009).*Unpacking core teaching practices in elementary mathematics to support teacher learning and assessment*. Principal Investigators meeting for the National Science Foundation Discovery Research-K12 Program, Washington, DC.

Boerst, T., Sleep, L., Ball, D. L., & Bass, H. (2011). Preparing teachers to lead mathematics discussions. *Teachers College Record*, *113*(12), 2844-2877.

Chapin, S., O'Connor, C., & Anderson, N. (2013) *Classroom discussions in math: A teacher's guide for using talk moves to support the Common Core and more, Grades K–6, 3rd Edition*. Math Solutions.

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

Danielson, C. (2011). *Enhancing professional practice: A framework for teaching*. ASCD.

Darling-Hammond, L. (2010). *Evaluating teacher effectiveness: How teacher performance assessments can measure and improve teaching*. Center for American Progress.

Davis, E. A., & Boerst, T. (2014). *Designing elementary teacher education to prepare well-started beginners*. (Working Paper). TeachingWorks: University of Michigan.

Engle, R. A. (2006). Framing interactions to foster generative learning: A situative explanation of transfer in a Community of Learners classroom. *Journal of the Learning Sciences*, *15*(4), 451-498.

Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. (2009). Teaching practice: A cross-professional perspective. *The Teachers College Record*, *111*(9), 2055-2100.

Jackson, K., Garrison, A., Wilson, J., Gibbons, L., & Shahan, E. (2013). Exploring relationships between setting up complex tasks and opportunities to learn in concluding whole-class discussions in middle-grades mathematics instruction. *Journal for Research in Mathematics Education*, *44*(4), 646–682.

Kazemi, E., & Hintz, A. (2014). *Intentional talk: How to structure and lead productive mathematical discussions*. Stenhouse Publishers.

Lampert, M. (2003). *Teaching problems and the problems of teaching*. Yale University Press.

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

Mathematics Assessment Resource Service (2012). *Proofs of the Pythagorean Theorem*. Shell Center: University of Nottingham. http://map.mathshell.org .

McDonald, M., Kazemi, E., & Kavanagh, S. S. (2013). Core practices and pedagogies of teacher education: A call for a common language and collective activity. *Journal of Teacher Education*, *64*(5), 378–386.

Michaels, S., O'Connor, C., & Resnick, L. B. (2008). Deliberative discourse idealized and realized: Accountable talk in the classroom and in civic life. *Studies in philosophy and education*, *27*(4), 283-297.

Miles, M. B. & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage.

Smith, M. & Stein, M. K. (2011). *5 practices for orchestrating productive mathematics discussions*. National Council of Teachers of Mathematics: Reston, VA.

TeachingWorks. (2015). *Leading a group discussion*. (Unpublished technical report). University of Michigan, Ann Arbor, MI.

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education*, *27*(4), 458–477.

Bartell, T. G., Bieda, K. N., Putnam, R. T., Bradfield, K., & Dominguez, H. (Eds.). (2015). *Proceedings of the 37th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. East Lansing, MI: Michigan State University.