2015

Using video to examine formative assessment practices as measures of expertise for mathematics and science teachers

Amelia Wenk Gotwals
Michigan State University

Joanne Philhower
Michigan State University

Dante Cisterna
Pontificia Universidad Catolica de Chile

Steven Bennett
Michigan State University

Using Video to Examine Formative Assessment Practices as Measures of Expertise for Mathematics and Science Teachers

Article *in* International Journal of Science and Mathematics Education · April 2015

DOI: 10.1007/s10763-015-9623-8

Amelia Wenk Gotwals
Michigan State University

Dante Cisterna
Educational Testing Service

Int J of Sci and Math Educ (2015) 13:405–423
DOI 10.1007/s10763-015-9623-8

Using Video to Examine Formative Assessment Practices as Measures of Expertise for Mathematics and Science Teachers

Amelia Wenk Gotwals · Joanne Philhower · Dante Cisterna · Steven Bennett

Received: 2 May 2014 / Accepted: 12 January 2015 / Published online: 12 March 2015

Abstract Formative assessment practices, including eliciting a broad range of student ideas, noticing the nuances in students' ideas, using these ideas to guide instruction, and promoting student self-regulation of learning are key components of expert teaching. Given the inherent dialogical nature of formative assessment in the classroom, video can provide a powerful tool for capturing and analyzing teachers' formative assessment interactions with students. In this study, we provide a framework for examining expertise in formative assessment and use this framework to quantitatively and qualitatively analyze the practices of 13 mathematics and science teachers. While we only saw a few instances of true expertise in formative assessment practices in our examination of videos, our findings indicate that teachers with more expertise in formative assessment let students' ideas guide their teaching. This leads to higher correlations among the dimensions of practice that we articulate in our framework for expert teachers. However, because many of the instructional decisions that teachers make are not visible on video, video alone may not provide enough information to judge expertise in formative assessment.

Keywords Formative assessment · Video analysis · Mathematics · Science

Teacher expertise is considered the most important element in student achievement (McCaffrey, Lockwood, Koretz & Hamilton, 2003; Rivkin, Hanushek & Kain, 2005; Wright, Horn & Sanders, 1997), and the influences of teachers on student learning have been found to be cumulative and long-lasting (McCaffrey et al., 2003; Sanders & Rivers, 1996). Despite the importance of teachers in supporting student learning, the field is still grappling with how to understand what expertise in teaching looks like and how to help teachers progress toward ambitious (i.e. expert-level) teaching.

A. W. Gotwals (✉) · J. Philhower · S. Bennett
Michigan State University, 620 Farm Lane, East Lansing, MI 48843, USA
e-mail: gotwals@msu.edu

D. Cisterna
Facultad de Educación, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, Macul, Santiago 7820436, Chile

In mathematics and science, researchers have defined ambitious teaching as the use of high-leverage practices that help all students understand and use knowledge to solve authentic problems (e.g. Lampert & Graziani, 2009). Teachers who use ambitious practices elicit students' ideas, make student thinking visible, and encourage students to reflect upon their current level of understanding (Windschitl, Thompson, Braaten & Stroupe, 2012). We have chosen to examine one aspect of ambitious practice: formative assessment (FA; i.e. assessment *for* learning). Formative assessment is an essential aspect of teaching in which teachers gather evidence of what their students know and use this information to modify their teaching practices to improve student learning (Black, Harrison, Lee, Marshall & Wiliam, 2004).

Ausubel (1968) wrote that "the most important single factor influencing learning is what the learner already knows. Ascertain this and teach him accordingly" (p. vi). This idea is at the essence of formative assessment. Instruction that includes FA (i.e. using students' ideas to guide instruction) has been empirically shown to have strong impacts on student learning in both science and mathematics (Bell & Cowie, 2001; Black & Wiliam, 1998; Ruiz-Primo & Furtak, 2006).

In this paper, we explore mathematics and science teachers' FA practices. Formative assessment is well researched as a domain general practice (e.g. Heritage, 2007), and we posit that there are many aspects of expert FA practices that are applicable across domains (e.g. regardless of discipline, ambitious teachers provide opportunities for students to share their thinking and use that evidence to guide their instruction). However, research also tells us that FA practices are highly embedded in disciplinary structures (i.e. attending to students' ideas requires different expertise depending on the content being taught) (Bennett, 2011; Wiliam, 2006). Despite this, there is less research on FA within specific disciplines (Coffey, Hammer, Levin & Grant, 2011).

Because of the importance of the dialogic interactions in eliciting and responding to student ideas, capturing FA in the classroom requires careful observation of interactions between teachers and students. Thus, video may provide a useful tool to operationalize FA practice in which expertise has been defined theoretically, but few concrete examples are available. In addition, examining video may provide insights into the particularities in expertise based on discipline. In this paper, we first provide a framework for conceptualizing expertise in FA. We then present video evidence of mathematics and science teachers' practices to explore the ways in which these teachers enact FA and make comparisons between FA practices in mathematics and science. The specific goal of this study is to use video to examine the nature of mathematics and science teachers' enactment of FA.

Literature Review

Formative assessment practices are a component of ambitious teaching in science and mathematics, and teachers with knowledge and skills in FA are able to better organize their instructional and assessment practices in order to promote student learning (Buck & Trauth-Nare, 2009). Students often enter classrooms with alternative ideas of how the world works and ambitious teachers elicit these prior ideas in order to build on and challenge them (Bransford, Brown & Cocking, 1999). If students do not "externalize and articulate their developing knowledge" (Sawyer, 2008, p. 12), misconceptions may persist and prevent students from acquiring deep, conceptual understandings. Teachers who are able to use information gathered from students to adjust the science and mathematics content they cover and methods by which they teach are more likely to help all kinds of students succeed at high-quality work (Franke, Carpenter, Levi & Fennema, 2001; Windschitl, Thompson & Braaten, 2011).

Expertise is often defined on a continuum from novice to expert (Ericsson, 2006). At the novice end of teachers' FA practice, FA may be seen as being done *to* learners with the majority benefit and information for the teachers. Alternatively, more expert forms of FA practice may consider assessment as being done *with* learners (Allal, 2010). Expert FA is a dialogic process that allows students' ideas to guide instruction and learning (e.g. see Crossouard & Pryor, 2012), and in this dialogic process, expert teachers consider students' ideas as nuanced and potentially productive instead of simply right or wrong (Furtak, Thompson, Braaten & Windschitl, 2012; Otero, 2006; Smith & Stein, 2011).

In order to characterize FA more carefully, we utilize work on feedback by Sadler (1989) and Hattie & Timperley (2007)[1] and posit that FA may be characterized into three large dimensions structured around key questions that teachers and students should ask themselves and each other as they move through the learning process (Gotwals, A.W., Cisterna, D., Lane, J., Kintz & Ezzo, D. (2015). Distinguishing observable formative assessment practices: A synthesis of the literature, under review).

1. Use of learning targets and goal setting: Where are we (teachers and students) going?
2. Evidence of student understanding: Where are we now?
3. Closing the gap/responding to students: How do we (teacher and students) get to the learning target?

Formative Assessment in Mathematics and Science

As we know, the specific discipline (e.g. science, mathematics) plays a large role in teachers' implementation of FA (e.g. Coffey et al., 2011; Pryor & Crossouard, 2010; Wiliam, 2009). What expert FA practice in each of the three dimensions looks like may differ by the content being taught. Below, we present research on expert FA practices in mathematics and science.

Where are we going? Identifying clear, student-friendly learning targets and communicating these targets with students is a quality of expert teaching that has been shown to improve student learning in science (White & Fredericksen, 1998) and mathematics (Tanner & Jones, 2003). When teachers share learning expectations verbally or in rubrics and involve students in the construction of learning goals, instruction tends to become more student-centered and students are more likely to take responsibility for

[1] Sadler (1989) delineated three necessary components of feedback: (1) the standard, which is to be achieved, (2) the actual level of performance, and (3) how to go about closing the gap. Building on this, Hattie and Timperley (2007) suggested that "effective feedback must answer three major questions asked by a teacher and/or a student: Where am I going? (What are the goals?), How am I going? (What progress is being made toward the goal?), and Where to next? (What activities need to be undertaken to make better progress?) (p. 86).

their own work (Tell, Bodone & Addie, 2000), making it a key component of ambitious teaching.

Where are we now? Studies have found that expert teachers often enact FA through dialogic conversation about big ideas in science (Duschl & Gitomer, 1997) and mathematics (Hodgen & Wiliam, 2006). These conversations allow multiple opportunities for gathering evidence of student learning and responding to students (Ruiz-Primo, 2011). For example, these conversations may allow teachers and students to identify or "notice" prior knowledge that can be built on (Otero, 2006; Sherin, Russ & Colestock, 2011); identify problematic ideas or misconceptions that can hinder learning (di Sessa & Minstrell, 1998); provide feedback to motivate improvement (Hattie & Timperley, 2007); and adjust teaching and learning strategies in an ongoing manner (Kohler, Henning & Usma-Wilches, 2008).

The types of questions used in class discussions influence the extent to which students can provide evidence of deep understandings that can be used for formative feedback and instructional decisions (Henningson & Stein, 1997). Questions with higher cognitive demand focus on meaning-making and require students to make connections between different representations (Smith & Stein, 2011).

How do we get to the learning target? In both mathematics and science, closing the gap between students' current understandings and the learning target is an expert FA practice that can be accomplished in multiple ways including engaging in feedback cycles and modifying instruction. When teachers are able to gather more accurate evidence about students' ideas, they can provide feedback that moves student thinking forward (Ash & Lewitt, 2003; Hodgen & Wiliam, 2006); are better able to make informed decision plans (Cowie & Bell, 2001), use feedback and make curricular adjustments that improve student learning (Buck & Trauth-Nare, 2009), and use information from embedded assessments to improve student learning (Ruiz-Primo & Furtak, 2006). In mathematics, teachers using ambitious teaching practices monitor and respond to student performance, adjusting both content and methods to promote high quality academic work (Lampert, Beasley, Ghousseini, Kazemi & Franke, 2010). In the science classroom, ambitious teachers draw out students' prior knowledge and elicit their explanations of natural phenomena, which can lead to deeper engagement with the content (Windschitl et al., 2012).

Use of Video in Examining Formative Assessment

Video provides a tool to bridge the complex theory behind FA with what it looks like in the classroom (e.g. Martin & Siry, 2012). Video recordings can capture different components of teachers' practice and allow researchers to analyze lessons from different perspectives (Hatch & Grossman, 2009). Formative assessment involves cultural routines in the classroom; video allows us to slow down, unpack, and critically examine these routines (Santagata, Zannoni & Stigler, 2007). Because interactions between teachers and students help characterize expertise in FA practices, video provides multiple opportunities for researchers to examine the nuances in these interactions. Stigler, Gallimore & Hiebert (2000) identified five research opportunities in using video: (1) It allows researchers to capture events more fully than checklists or other observation tools; (2) it can be analyzed by multiple coders and with multiple lenses; (3) it can be used over time; (4) it can advance the scientific understanding of classroom processes (such as FA); and (5) it allows researchers to integrate qualitative and quantitative methods of analysis.

However, using video in classrooms can present challenges. For example, students may behave differently when they are being taped. Similarly, teachers may alter their methods based on who is in their classroom. However, teachers are unlikely (and likely unable) to improve their methods drastically just because there is a camera in the classroom (Stigler, Gallimore & Hiebert, 2000). In addition, the use of video for analyzing the complexity of teaching may be overwhelming, especially when analyzing more than one lesson (Hatch & Grossman, 2009).

Methods

This study is part of a larger research project: Formative Assessment for Michigan Educators (FAME). FAME is a professional development program designed to enhance practicing teachers' FA practices in support of student learning. Teachers work in professional learning communities with the goal of improving their understanding about FA theory and practices as well as working towards implementing, reflecting on, and refining new instructional and assessment practices.

Participants and Data

This study examines the FA practices of 13 mathematics and science teachers who participated in the FAME program and were in various stages of learning about FA. These teachers, all with more than 8 years of teaching experience, taught classes ranging from early elementary school (grade 2) through upper-level high school (see Table 1). We have video of two teachers (Teachers C and F) teaching both mathematics and science for the same grade level.

When gathering video data, we used a single camera mounted on a tripod at the back of the classroom and the teacher wore a wireless lavalier microphone. This allowed us to capture all of the teacher's moves and audio; however, there were occasions where it was difficult to hear everything students said in whole-class discussions or small-group interactions when the teacher was not close. We did not use specific criteria for selecting the teachers or lessons to videotape. Rather, we videotaped teachers who were a part of FAME and who volunteered to have us in their classrooms.

Coding

In order to code the videos, we used a research-based FA practice progression (Gotwals, A. W., Lane, J., Cisterna, D., Philhower, J., Bennett, S., Kintz, T. & Roeber, E. (2015) Testing hypotheses about teaching: A practice progressions approach to formative

Table 1 Teachers in the study

Teacher	Grade Level	Content	Number of lessons
Teacher A	2	Mathematics	2
Teacher B	4	Mathematics	3
Teacher C	2	Mathematics	1
Teacher D	6	Mathematics	4
Teacher E	8	Mathematics	3
Teacher F	6	Mathematics	1
Teacher G	HS	Algebra	1
Teacher H	HS	Algebra	5
Teacher C	2	Science	1
Teacher F	6	Science	6
Teacher I	6	Science	1
Teacher J	8	Science	1
Teacher K	HS	Chemistry	4
Teacher L	HS	Chemistry	4
Teacher M	HS	Earth science	1

assessment, under review.). The progression is structured around the three main dimensions of FA (discussed above). We used the literature to define specific teacher practices for each dimension and developed a four-level novice-expert progression to capture the nature of teachers' FA practices. We developed guiding principles to articulate each level in the progression (see Table 2).

We used the guiding principles in Table 2 to develop progressions for seven FA practices that fall under the three dimensions (see Table 3).

We segmented each video by classroom activity type (e.g. teacher lecture, whole class discussion, small group work), recording the amount of time devoted to each activity and the number and types of questions (i.e. low or high cognitive demand; Webb, 2009) asked by the teacher and students during the activity. In addition, we characterized each segment by the presence or absence of the practices listed in Table 3. For those practices that were present in the activity segment, coders assigned them a progression level. Using a social moderation process (e.g. Frederiksen, Sipusic, Sherin & Wolfe, 1998), two researchers coded each classroom video and came to a consensus about codes.

Analysis

In order to examine the extent to which video allowed us to capture FA practices in the classroom, we use a mixed-methods analysis of the data (Teddlie & Tashakkori, 2009). Quantitatively, we examined our codes for patterns within and between teachers who teach different disciplines to examine how well we could capture differences in FA practices (and the levels of expertise of these practices). We focused on the frequency of observations for specific FA practices and computed weighted averages (based on the number of lessons we had for the teacher) for levels of each practice. We used

Table 2 Novice-expert continuum for formative assessment

	Level 1: novice	Level 2: intermediate I	Level 3: intermediate II	Level 4: expert
Teacher and student roles in formative assess ment discourse	Assessment discourse is solely guided by the teacher	Assessment discourse is mostly guided by the teacher, with some student direction	Assessment discourse is guided by students' ideas but teacher leads	Assessment discourse is co-lead by students guided by student ideas
Teacher use of students' ideas/ evidence	Teacher use students' ideas to see if they are right/wrong; students do not appear to use in formation	Teachers use students' ideas to see if they are right/wrong; stu dents use informa tion for verification of answers	Teachers use students' ideas to see what/how they understand; students use information for self-correction	Teachers use students' ideas to see what/how they understand; students use information for self-assessment
How is the informa tion used?	Teacher does not appear to use the information to inform instruction	Teacher uses the information to correct student ideas or discuss non-disciplinary proce dures	Teacher uses the information to scaffold student thinking	Teacher uses the information to guide instruction and promote student's development of self-regulation.

independent samples *t* tests to examine differences in the average levels of expertise between participating mathematics and science teachers and ran Pearson correlations to

Table 3 Dimensions of practice progressions

Dimension	Practice	Definition
1: Use of learning targets and goal setting: Where are we (teachers and students) going?	Learning target use	How teacher introduce the learning target to students and connect it to instructional activities
2: Eliciting evidence of student understanding: What does the student understand now?	Type of question/ info elicit ed	The cognitive demand of questions and tasks that teachers use to gather evidence of student understanding
	Elicitation strategies	Strategies to guide questioning and elicit evidence from students
	Self-assessment	Opportunities for students to be examine their own work and regulate their own learning
3: Closing the gap/responding to students: How do we (teacher and students) get to the learn ing target?	Feedback loops	The types of feedback provided to students and the manner in which the teacher provides the feedback
	Peer assessment	Opportunities for peers to examine each other's work (and, potentially, help each other learn)
	Instructional decisions	The extent to which teachers are explicit about the rationale for making instructional choices alterations

look at relationships between FA practices in each discipline. For our qualitative analysis, we used transcriptions from the video that illustrates teachers' FA practices and how practices are similar or different between mathematics and science teachers.

Findings

The organization of this section follows the three guiding questions that describe the dimensions of FA: Where are we going? Where are we now? How do we get to the learning target? We focus on the range of practices on the novice–expert continuum in each of the three dimensions.

Where are we going?

Learning target use. In the videos we observed, we found that the majority of teachers mentioned learning targets to the students at least once, including six out of eight mathematics teachers and four out seven science teachers. However, there were variations in teachers' levels of practice (see Fig. 1). Specifically, we saw only one teacher (Teacher D) reach a Level 4 (expert). Figure 1 provides the frequency and level of teachers' learning target use.

There was variability in the use of learning targets among teachers and between lessons from the same teacher; however, there was no significant difference in the average level of learning target use between mathematics and science teachers. As an example of a Level 3 use of learning targets in science, Teacher K had the following learning target listed on the board and on students' packets: "Identify half-reactions, oxidation and reduction, in a series of experimental reactions." Before the lab, the teacher said, "Don't forget the learning target, the goal of the lab! First of all, when you look at your cells, we know that oxidation occurs at the anode, electrons are lost, that's where the flux of electrons starts … When you go through this [lab] you're trying to figure out the half reactions, what was oxidized and what is reduced …" Here, the video allowed us to capture both the written words that she had on the wall and how she clearly

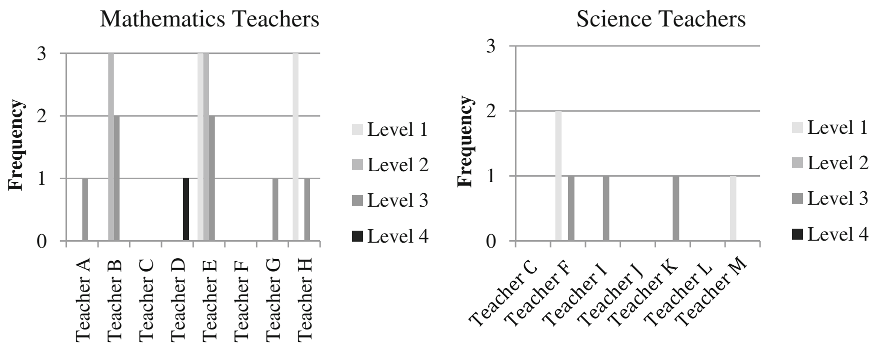

Fig. 1 Frequencies for learning target use by level of expertise

connects the goal of the lab to the learning target. Similarly, in mathematics, Teacher D (who reached Level 4) also had the learning targets posted on the board and read them to her students, "Our learning target that we are going to be spending some days on is, *I can solve real world and mathematical problems involving surface area using nets*." The teacher then held up a box that had been flattened to form a net to show students what a real world net would look like and explained to the students how jobs in packaging would use this type of information to determine the most cost effective box size. Video allowed us to capture the written work, the use of manipulative tools (i.e. the box), and expert use of learning targets.

Where are we now?

Questioning practices. We report on three criteria to examine teachers' questioning practices in the videos: (1) a count of low or high depth of knowledge (DOK) questions asked; (2) average level of the type of question asked; and (3) average level of elicitation strategies. The levels for the type of questions asked ranges from Level 1, with questions that have one correct answer, to a Level 4, with questions that ask students to explain and be metacognitive about their thought process. The levels for elicitation strategies range from teachers using questioning in search of the right answer (Level 1) to a strategy of probing and pressing questions (more dialogic) at a Level 4. The results for the mathematics and science teachers are in Table 4.

Table 4 Mathematics and science teachers' questioning practices

Teacher	Content	Number of lessons	Average number of questions by lesson		Average level of type of question	Average level of elicitation strategy
			Low DOK	High DOK		
A	Math	2	9.6	0.4	1.7	2.5
B	Math	3	28.7	10.0	3.2	3.2
C	Math	1	13.0	10.0	4.0	4.0
C	Science	1	84.0	11.0	1.8	1.8
D	Math	4	19.5	2.2	2.3	2.0
E	Math	3	20.3	1.3	1.2	1.9
F	Math	1	26.0	0.0	1.5	1.5
F	Science	6	15.0	0.0	1.1	1.7
G	Math	1	44.0	2.0	1.8	1.9
H	Math	5	36.4	0.2	1.6	2.0
I	Science	1	40.0	5.0	2.0	2.0
J	Science	1	5.0	0.0	1.0	2.0
K	Science	4	28.5	3.5	1.4	1.6
L	Science	4	14.0	1.2	1.5	2.0
M	Science	1	17.0	0.0	2.0	3.0

All of the teachers asked considerably more low-depth-of-knowledge questions than high depth of knowledge questions. Three teachers (Teachers F, J, and M; all science) did not ask any high-depth-of-knowledge questions during the lessons we observed. Only two teachers (Teachers B & C, in mathematics) reached an average of a Level 3 or higher for the types of questions they asked, and only three teachers (Teachers B, C, and M, two mathematics and one science) used more advanced (Level 3 or higher) elicitation strategies. On average, the mathematics teachers in our study tended to ask more questions than the science teachers; asked higher level questions (M_{math}=2.0; $M_{science}$=1.4), $t(29)=2.4$, $p=0.022$; and had higher elicitation strategies (M_{math}=2.4; $M_{science}$=1.7), $t(35)=2.3$, $p=0.023$.

Table 5 provides a transcript of Teacher C, a second-grade mathematics teacher, engaging in Level 3 questioning and Level 3 feedback (discussed in the next section) around subtraction. In this interaction, she attempts to get the student to explain why she solved the problem in the way that she did. When the student made a mistake, Teacher C did not correct her; rather she asked her to explain her reasoning. In this instance, the student was able to realize her mistake, and the teacher never had to provide the correct answer. The video allowed us to see the student's and teacher's

Table 5 Transcript and notes on a mathematics exchange between teacher C and a student

Transcript	Notes
Student subtracting 505–371 on the front board.	Students given the opportunity to use any method to solve a mathematics problem
Teacher (pointing to the 0): Why did you borrow for this one?	Teacher asks probing question about students' process of thinking about the task
Student: Because you have nothing; you have a zero … so you have to borrow from the neighbor on the left.	Student reasons through why she had to borrow using her own words
Teacher: (pointing to the 5 in the 1's place): You had to borrow to subtract from zero, so why didn't you borrow to subtract this from this?	Teacher probes students' understanding again to better understand
Student: Shrugs shoulders	Student is unsure how to explain her rationale for borrowing
Teacher: Can you take one away from five?	Teacher asks a question that scaffolds student thinking
Student: (smiling) Yes, I can subtract one from five, so it works … so I would take 1 from 5 and that gives me 4 …(she works on the problem) … my answer would have to be 134, but to check my answer I would have to add	Student uses the scaffold in the teachers' question in order to make sense of the problem solving and solves the problem
Teacher: … Hmmmm … Add what?	Teacher asks for clarification
Student: Ummm … 505 and 371?	Student response (incorrect)
Teacher: Okay, so why do you add those numbers?	Teacher asks for a rationale (but does not correct student's mistake)
Student: No … I have to add 134 and 371 to see if I get 505.	Student realizes her mistake and corrects it before moving forward (although does not respond to the *why* part of the question)
Teacher: Okay, let's do it	

gestures during problem solving while learning how Teacher C used the students' understanding to guide her questioning and feedback.

Self-assessment. Helping students develop appropriate self-assessment strategies allows students to take charge of their learning. Figure 2 represents the frequencies and levels of self-assessment opportunities for mathematics and science teachers.

We did not see any teachers provide Level 4 (expert) opportunities for students to self-assess. Almost all of the mathematics teachers in our study (seven out of eight) provided students with some opportunity to self-assess, but only three science teachers offered students opportunities to self-assess. On average, mathematics teachers' use of this practice was at a higher level than science teachers, but none of the teachers reached an expert level ($M_{math}=1.9$; $M_{science}=0.6$), $t(36)=4.6$, $p<.001$.

Teacher L, a high school chemistry teacher, provided Level 3 opportunities for students to self-assess (i.e. she encouraged students to check and adjust their understanding and provided students with opportunities to practice the process of reflection). In one instance, she had students complete an exit slip on red, yellow, or green paper depending on their self-assessed understanding of the content. If students did not understand the learning goal (states of matter), they were to explain what they did not know on the red paper ("use if you really don't understand anything") or the yellow paper, where they would describe what they did not understand. If students were confident in their understanding of the concepts, they were to demonstrate this by writing an explanation or solving a problem on the green paper. The teacher then used these exit slips the next day to follow-up with students.

Teacher H, a high school algebra teacher, also provided multiple Level 3 opportunities for students to self-assess. For example, when reviewing conic sections (i.e. parabola, circle, ellipse, and hyperbola), she had students hold up one to five fingers to represent their confidence in their ability to solve problems involving each of these concepts. When Teacher H noticed students holding up three or less fingers, she asked the students specific questions to determine where they were confused. The use of video allowed us to capture both the students' and teacher's actions and determine the level of the teacher's expertise of implementing this practice.

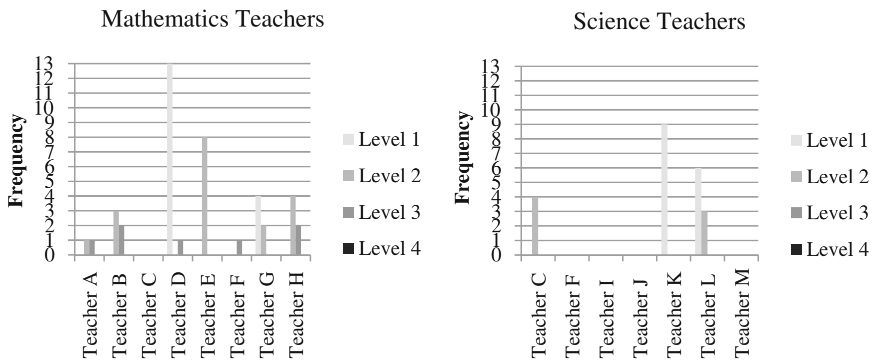

Fig. 2 Frequencies for self-assessment use by level of expertise

How do I close the gap?

Feedback loops. Regardless of discipline, all teachers provided some sort of feedback to their students; however, the level at which teachers implemented these practices was higher in our sample of mathematics teachers than science teachers (M_{math}=1.9; $M_{science}$=1.2), $t(36)=2.6$, $p=.013$ (although, again, the average levels of practice were relatively low). Four mathematics teachers demonstrated Level 3 practices (descriptive feedback focused on students' progress on the task, but the feedback either provides too much or not enough scaffolding), and two of these teachers also demonstrated Level 4 practices (descriptive feedback focused on students' progress on the task that supports the student to determine the next step). The highest achieving science teachers, however, only demonstrated Level 2 practices (evaluative feedback focused on whether student had the right answer). Figure 3 shows the frequency of feedback opportunities provided by the mathematics and science teachers.

As shown in Table 5, Teacher C provided students with non-evaluative feedback that built on their current understandings and that was crafted such that it helped students to close the gap between their current understanding and the learning goal. However, we were unable to find this type of feedback from our science teachers; rather, their feedback tended to be evaluative without providing prompts or scaffolds for closing the gap between students' understanding and the learning target.

Instructional decisions. Similar to other practices, our sample of mathematics teachers had higher levels of instructional decisions (M_{math}=2.1) than our science teachers ($M_{science}$=0.9), $t(36)=2.9$, $p=.007$. Figure 4 shows the number of teachers at each level, as well as the frequency of instructional decision occurrences.

In general, the mathematics teachers tended to make their instructional decisions visible, with all but one of them demonstrating a Level 3 rating and one also achieving a Level 4. For the teachers demonstrating a Level 3 practice, they did not make the use of student evidence explicit in their reasons for making decisions. Fewer science teachers demonstrated observable instructional decision-making. Teachers B (mathematics) and C (science), both elementary teachers, were the only ones who demonstrated expert levels in instructional decision-making. To achieve a Level 4

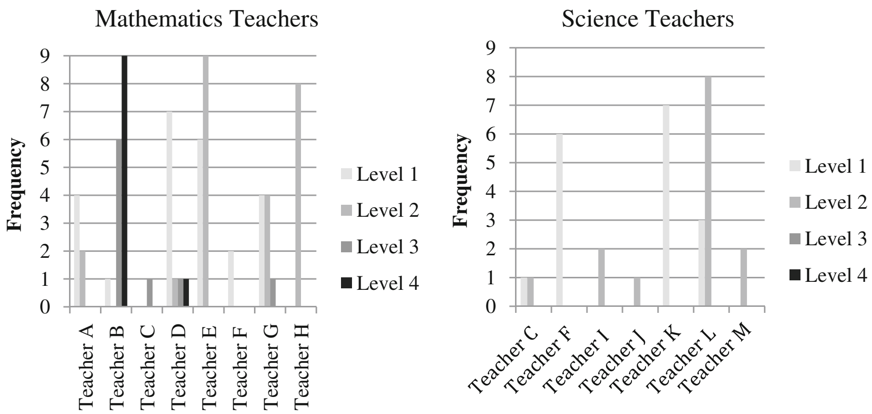

Fig. 3 Frequencies for feedback use by level of expertise

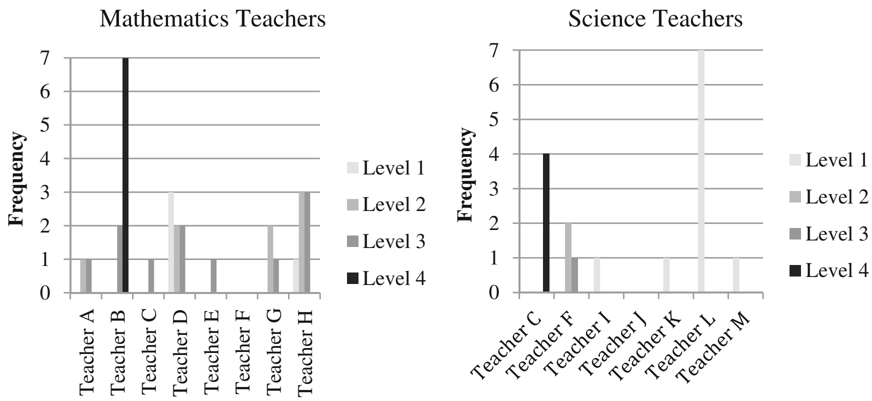

Fig. 4 Frequencies for instructional decisions by level of expertise

rating, Teacher C was explicit about the reasons for her decisions and used evidence of student understanding in making this decision. For example, at one point when discussing plants, she asked students, "What does the seed do?" and gathered multiple answers, which ranged from off-base to partially correct, but not at the level of the learning target. She then asked students to give "Thumbs up if you know what a seeds' job is; thumbs in the middle if you think you do but you are afraid to say so; Thumbs down if you have no idea." Most students had thumbs down, so she said, "It seems we are a little uncertain about the job of the seed," and she used this evidence to guide her next step, which was to ask students scaffolding questions to guide them in remembering their prior learning experiences with seeds. The use of video allowed us to capture this interaction that included both verbal prompts and gestures (e.g. the thumbs).

During a math lesson we observed Teacher B make several instructional decisions while teaching students how to divide:

> Teacher: Okay, some of you are a little stuck, so let's talk through this together. We did the division, so the next step is to multiply (pointing to the letters D, M, S, B on the front board). [Pause to work through the M and S steps. The teacher observes students still struggling] … Okay, I want to show you again how to do this digit by digit because I see some of you are still using expanded notation. Let's continue going through it together.

In this instance, we were able to use the video to see how the teacher walked around and observed students' struggles with doing the work independently and then chose to alter her course of instruction based on her perception of students' understandings.

Correlational Analysis

We ran Pearson correlations to determine the extent to which the FA practices were correlated in each discipline. For Mathematics, Learning Target Use and Self-

Assessment were not significantly correlated with any of the other practices. Elicitation Strategies, Type of Questions, Feedback, and Instructional Decisions were all moderately to strongly correlated with each other (see Table 6).

Similar to the mathematics lessons, in Science, Learning Target Use for the science teachers was not significantly correlated with any other practice. However, the other relationships in science were different from those in math. Specifically, Peer-Assessment was significant and positively correlated with Self-Assessment and Feedback; and Elicitation Strategies was significant and positively correlated with Type of Questions (see Table 7).

Discussion

Despite a small sample size of teachers, our findings provide some suggestions about expertise in FA. In this section, we explain our findings from using video to examine expertise in FA practices and propose implications of these findings.

Mathematics Versus Science Expertise

In our sample of teachers, we found higher levels of FA practices in mathematics teachers across the board. While this does not suggest a higher level of expertise in mathematics teachers overall (in fact, see http://ambitiousscienceteaching.org for examples of ambitious science teaching), these findings may be a reflection of how science is often taught, emphasizing the acquisition of facts (Sawyer, 2008) and valuing "correct" ideas over students' reasoning (Hammer, 2004). In contrast, ambitious science teaching encourages students to develop, share, contemplate, and revise causal explanations, leading to a deep understanding of science concepts (Windschitl et al., 2012). Similarly, in mathematics, ambitious instruction involves using students' ideas to help them acquire, understand, and use complex mathematics knowledge (Lampert et al., 2010).

Table 6 Correlations for mathematics lessons formative assessment dimensions

	Type of questions	Elicitation strategies	Self-assessment	Feedback	Peer assessment	Instructional decisions
Learning target use	−0.126	−0.131	−0.047	−0.092	−0.394	−0.009
Type of questions		0.879[a]	−0.013	0.771 [a]	0.293	0.675 [a]
Elicitation strategies			−0.085	0.681 [a]	0.222	0.684 [a]
Self-assessment				0.348	−0.102	0.316
Feedback					0.268	0.786 [a]
Peer assessment						0.149

[a] Correlation is significant at the 0.01 level (two-tailed)

Table 7 Correlations for science lessons formative assessment dimensions

	Type of questions	Elicitation strategies	Self-assessment	Feedback	Peer assessment	Instructional decisions
Learning target use	0.411	0.073	−0.343	−0.256	−0.299	−0.104
Type of questions		0.518 [a]	0.296	0.452	0.119	−0.031
Elicitation Strategies			0.029	0.462	−0.283	0.283
Self-assessment				0.374	0.569 [a]	0.182
Feedback					0.555 [a]	−0.167
Peer assessment						−0.360

[a] Correlation is significant at the 0.05 level (two-tailed)

When comparing the correlations of the specific FA practices between mathematics and science teachers, we find that FA practices are more highly correlated for the mathematics teachers, who tended to have higher levels of expertise. This suggests that, at more expert levels of teaching, FA is less about separate practices being implemented at different times and more about a singular process of articulating learning goals, eliciting students' ideas, and using these ideas in the course of instruction to guide teaching, provide feedback, and help students regulate their own learning. When FA is done well, it should be seamlessly integrated into instruction and suggests a classroom culture that is welcoming to student ideas (Shepard, 2000) rather than as simply a tool that is used for a given activity.

This finding may also have implications for the strong influence of discipline on the nature of FA practice (Coffey et al., 2011; Pryor & Crossouard, 2010; Wiliam, 2009). While our data do not speak specifically to this, it is possible that mathematics teachers enact FA in ways that are more observable and explicit or that the activities that they enact in the classroom are more suited to certain types of FA that we looked for. Alternatively, it is also likely the differences were due to the types of activities that we capture with mathematics teachers. The mathematics teachers in our study provided students with more opportunities to participate in a whole class discussion than the science teachers in our study (although the difference is not statistically significant). This led to more opportunities for us as researchers to capture interactions between the teacher and students, allowing us to note more examples of questioning and feedback.

Ericsson (2006) stated that,

> …extensive experience of activities in a domain is necessary to reach very high levels of performance. Extensive experience in a domain does not, however, invariably lead to expert levels of achievement… further improvements depend on deliberate efforts to change particular aspects of performance (p. 685).

The teachers in our study, regardless of discipline, did not always (or often) demonstrate high-quality FA practices; however, many of them showed proficiency in specific aspects, such as questioning types, elicitation strategies, and feedback. While we may want teachers to enact more ambitious FA practice, it is very difficult for teachers to change their practice quickly (Bennett, 2011; Webb & Jones, 2009), and embedding FA into their current practice can seem overwhelming (Buck & Trauth-Nare, 2009). The high correlation between practices in expert's classrooms suggests that teachers may be able to choose certain practices (e.g. eliciting students' ideas) and these practices may "bootstrap" improvement in other practices, such as feedback or instructional decisions. This idea is supported by research that shows that when teachers focus on a single aspect of FA, student achievement rises (Shih & Alexander, 2000; Rakoczy, Harks, Klieme, Blum & Hochweber, 2013; Ross, Hogaboam-Gray & Rolheiser, 2002; Ruiz-Primo & Furtak, 2006). Using video to help teachers see what expertise looks like in FA may provide one way for teachers to begin to move their practice forward because they could see how certain types of practices link together in a more seamless embedding of FA and instruction.

Evaluating Expertise in Formative Assessment as a Double Inferential Process

Expertise may be difficult to evaluate solely on observation (with or without video), as much of expertise in teaching resides in the interaction of context-specific knowledge, pedagogical content knowledge, and practice in the classroom. Mathematics teachers seemed more likely to announce their instructional decisions to the class (e.g. "you all seem to be struggling with carrying numbers in subtraction, let's come back together to talk about it"). However, just because they did not announce their decisions, science teachers likely made many instructional decisions based on student evidence that we were unaware of. Formative assessment is an inferential process in that teachers must make a hypothesis about what their students know and use this information to guide their subsequent moves (Bennett, 2011). "Everything students do—such as conversing in groups, completing seatwork, answering and asking questions, … (or) even sitting silently and looking confused—is a potential source of information about how much they understand" (Leahy, Lyon, Thompson & Wiliam, 2005, p. 19). This means that teachers are constantly using verbal and non-verbal information as well as weighing contextual factors (e.g. how much time is left in the class) to make on-the-fly decisions.

In addition to many FA practices being inferential, coding of classroom video is an inferential process. When teachers say or do certain things in the classroom, they often do not explicitly state the evidence or rationale that they are using to guide their moves. Especially at the expert levels of FA, many practices involve both teacher inference about students and observer inference about teachers. The greater the burden on observers to infer about practice, the less observers may be able to "see," the more they have to speculate about what teachers are doing; and as reliance on inference increases so does the threat to reliability among raters (Kennedy, 2010). This suggests that video alone may not allow us to reliably judge teacher expertise in FA; that some components of FA may be more visible than others; and that the way that mathematics tends to be taught may provide more opportunities to make FA practices explicit during classroom instruction.

We recognize the limitations of our study. Although the participant teachers were part of a professional development program to learn about FA, the lessons observed are diverse in terms of grade level, schools, and type of content. Moreover, while we know that there are many mathematics and science teachers with great expertise in FA, our sample did not include extensive evidence of this type of practice. Once we are able to capture more video data, we will likely have more evidence about the nature of teachers' expertise in FA.

Conclusion

Teaching is a diverse and complicated enterprise; and expertise in teaching is difficult to define and even harder to measure (Hill, Ball & Schilling, 2008; Tschannen-Moran, Hoy & Hoy, 1998). Despite this, research needs to continue to identify expert teaching practices, such as FA, embedded within mathematics and science in order to support students' learning. Understanding how teachers develop expertise needs to be considered in order to prepare prospective teachers more appropriately to navigate the rigors of leading a classroom. Our findings suggest there are some aspects of expertise in FA that are applicable and can be analyzed across disciplines, but that it is important to examine teachers' practices in different disciplines to fully characterize expertise in FA. In this vein, Bennett (2011) stated that, "rooting FA in pedagogical skills alone is probably insufficient. Rather, FA would be more profitably conceptualized and instantiated within specific domains" (p. 20). This study presents a first step in characterizing FA practices in mathematics and science and how we might use video to capture the nature of expertise in FA in these different disciplines.

Acknowledgments Work on the Formative Assessment for Michigan Educators (FAME) project was supported by the Michigan Department of Education.

References

Allal, L. (2010). Assessment and the regulation of learning. In P. Peterson, E. Baker & B. McGaw (Eds.), *International Encyclopedia of Education* (Vol. 3, pp. 348–352). Oxford, England: Elsevier.

Ash, D. & Lewitt, K. (2003). Working within the zone of proximal development: formative assessment as professional development. *Journal of Science Teacher Education, 14*(1), 23–48.

Ausubel, D. P. (1968). *Educational psychology: a cognitive view*. New York, NY: Holt, Rinehart & Winston.

Bell, B. & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education, 85*, 536–553.

Bennett, R. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5–25.

Black, P., Harrison, C., Lee, C., Marshall, B. & Wiliam, D. (2004). Working inside the black box: assessment for learning in the classroom. *Phi Delta Kappan, 86*(1), 8–21.

Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice, 5*(1), 7–73.

Bransford, J. D., Brown, A. L. & Cocking, R. R. (1999). *How people learn: brain, mind, experience, and school*. Washington, D.C: National Academy Press.

Buck, G. A. & Trauth-Nare, A. E. (2009). Preparing teachers to make formative assessment process integral to science teaching and learning. *Journal of Science Teacher Education, 20*, 475–494.

Coffey, J. E., Hammer, D., Levin, D. M. & Grant, T. (2011). The missing disciplinary substance of formative assessment. *Journal of Research in Science Teaching, 48*(10), 1109–1136.

Cowie, B. & Bell, B. (2001). A model of formative assessment in science education. *Assessment in Education, 6*(1), 101–116.
Crossouard, B. & Pryor, J. (2012). How theory matters: formative assessment theory and practices and their different relations to education. *Studies in Philosophy and Education, 31*(3), 251–263.
di Sessa, A. A. & Minstrell, J. (1998). Cultivating conceptual change with benchmark lessons. In J. G. Greeno & S. V. Goldman (Eds.), *Thinking practices in mathematics and science learning* (pp. 155–187). New York, NY: Routledge.
Duschl, R. A. & Gitomer, D. H. (1997). Strategies and challenges to changing the focus of assessment and instruction in science classrooms. *Educational Assessment, 4*(1), 37–73.
Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.), *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge, UK: Cambridge University Press.
Franke, M. L., Carpenter, T. P., Levi, L. & Fennema, E. (2001). Capturing teacher's generative growth: a follow-up study of professional development in mathematics. *American Education Research Journal, 38*(3), 653–690.
Frederiksen, J. R., Sipusic, M., Sherin, M. & Wolfe, E. W. (1998). Video portfolio assessment: creating a framework for viewing the functions of teaching. *Educational Assessment, 5*(4), 225–297.
Furtak, E. M., Thompson, J., Braaten, M. & Windschitl, M. (2012). Learning progressions to support ambitious teaching practices. In A. C. Alonzo & A. W. Gotwals (Eds.), *Learning Progressions in Science*. Rotterdam, The Netherlands: Sense Publishers.
Hammer, D. (2004). The variability of student reasoning, lecture 3: manifold cognitive resources. In E. Redish & M. Vicentini (Eds.), *Proceedings of the Enrico Fermi Summer School, Course CLVI* (pp. 321–340). Bologna, Italiana: Italian Physical Society.
Hatch, T. & Grossman, P. L. (2009). Learning to look beyond the boundaries of representation: using technology to examine teaching (overview for a digital exhibition: learning from the practice of teaching). *Journal of Teacher Education, 60*(1), 70–85.
Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.
Henningson, M. & Stein, M. K. (1997). Mathematical tasks and student cognition: classroom based factors that support and inhibit mathematical thinking and reasoning. *Journal for Research in Mathematics Education, 28*(5), 524–549.
Heritage, M. (2007). Formative assessment: what do teachers need to know and do? *Phi Delta Kappan, 89*(2), 140–145.
Hill, H. C., Ball, D. L. & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: conceptualizing and measuring teachers' topic specific knowledge of students. *Journal for Research in Mathematics Education, 39*(4), 372–400.
Hodgen, J. & Wiliam, D. (2006). *Mathematics inside the black box: assessment for learning in the mathematics classroom*. London, England: GL Assessment Limited.
Kennedy, M. M. (2010). Approaches to annual performance assessment. In M. M. Kennedy (Ed.), *Teacher Assessment and the Quest for Teacher Quality.* San Francisco, CA: Jossey-Bass.
Kohler, F., Henning & Usma-Wilches, J. (2008). Preparing preservice teachers to make instructional decisions: an examination of data from the teacher work sample. *Teaching and Teacher Education, 24*(8), 2108–2117.
Lampert, M., Beasley, H., Ghousseini, H., Kazemi, E. & Franke, M. (2010). Using designed instructional activities to enable novices to manage ambitious mathematics teaching. In *Instructional explanations in the disciplines* (pp. 129-141). New York , NY: Springer.
Lampert, M. & Graziani, F. (2009). Instructional activities as a tool for teachers' and teacher educators' learning. *The Elementary School Journal, 109*(5), 491–509.
Leahy, S., Lyon, C., Thompson, M. & Wiliam, D. (2005). Classroom assessment—minute by minute, day by day. *Educational Leadership, 63*, 18–24.
Martin, S. & Siry, C. (2012). An analysis of the utilization of video-based media in science teacher education. In B. Fraser, K. Tobin & C. Campbell (Eds.), *International handbook of science teaching and learning*. (pp. 417–433). Rotterdam: Springer.
McCaffrey, J. R., Lockwood, D. F., Koretz, D. M. & Hamilton, L. S. (2003). *Evaluating value added models for teacher accountability* [Monograph]. Retrieved from http://www.rand.org/pubs/monographs/2004/RAND_MG158.pdf.
Otero, V. (2006). Moving beyond the "Get it or don't" conception of formative assessment. *Journal of Teacher Education, 57*(3), 247–255.
Pryor, J. & Crossouard, B. (2010). Challenging formative assessment: disciplinary spaces and identities. *Assessment & Evaluation in Higher Education, 35*(3), 265–276.

Rakoczy, K., Harks, B., Klieme, E., Blum, W. & Hochweber, J. (2013). Written feedback in mathematics: mediated by students' perception, moderated by goal orientation. *Learning and Instruction, 27*, 63–73.
Rivkin, S. G., Hanushek, E. A. & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrics, 73*(2), 417–458.
Ross, J. A., Hogaboam-Gray, A. & Rolheiser, C. (2002). Student self-evaluation in grade 5−6 mathematics effects on problem-solving achievement. *Educational Assessment, 8*(1), 43–58.
Ruiz-Primo, M. A. (2011). Informal formative assessment: the role of instructional dialogues in assessing students for science learning. *Special issue in assessment for learning, Studies of Educational Evaluation, 37*(1), 15–24.
Ruiz-Primo, M. A. & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: exploring teachers' practices and student learning. *Educational Assessment, 11*(3–4), 205–235.
Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119–144.
Sanders, W. L. & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement (Research Progress Report)*. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.
Santagata, R., Zannoni, C. & Stigler, J. W. (2007). The role of lesson analysis in pre-service teacher education: an empirical investigation of teacher learning from a virtual video-based field experience. *Journal of Mathematics Teacher Education, 10*(2), 123–140.
Sawyer, R. K. (2008). *Optimizing learning: implications of learning sciences research.* Paper presented at the OECD/CERI International Conference, "Learning in the 21st Century: Research, Innovation and Policy," Paris, France. Retrieved from http://www.oecd.org/site/educeri21st/40554221.pdf.
Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4–14.
Sherin, M. G., Russ, R. S. & Colestock, A. A. (2011). Accessing mathematics teachers' in-the-moment noticing. In M. G. Sherin, V. R. Jacobs & R. A. Philipp (Eds.), *Mathematics Teacher Noticing* (pp. 79–94). New York, NY: Routledge.
Shih, S. S. & Alexander, J. M. (2000). Interacting effects of goal setting and self- or other-referenced feedback on children's development of self-efficacy and cognitive skill within the Taiwanese classroom. *Journal of Educational Psychology, 92*(3), 536–543.
Smith, M. S. & Stein, M. K. (2011). *5 Practices for Orchestrating Productive Mathematics Discussions.* Reston, VA: National Council of Teachers of Mathematics.
Stigler, J. W., Gallimore, R. & Hiebert, J. (2000). Using video surveys to compare classrooms and teaching across cultures: examples and lessons from the TIMSS video studies. *Educational Psychologist, 35*(2), 87–100.
Tanner, H. & Jones, S. (2003). Self-efficacy in mathematics and students' use of self-regulated learning strategies during assessment events. In N. A. Pateman, B. J., Dougherty & J. T. Zilliox (Eds.), *Proceedings of the 2003 Joint Meeting of PME and PMENA* (pp. 275–281). Honolulu, HI: CRDG, College of Education, University of Hawai'i.
Teddlie, C. & Tashakkori, A. (2009). *Foundations of mixed methods research: integrating quantitative and qualitative approaches in the social and behavioral sciences*. Thousand Oaks, CA: Sage Publications Inc.
Tell, C. A., Bodone, F. M. & Addie, K. L. (2000). *A framework of teacher knowledge and skills necessary in a standards-based system: lessons from high school and university faculty.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. Retrieved from http://eric.ed.gov/?id=ED447206.
Tschannen-Moran, M., Hoy, A. W. & Hoy, W. K. (1998). Teacher efficacy: its meaning and measure. *Review of Educational Research, 68*(2), 202–248.
Webb, M. & Jones, J. (2009). Exploring tensions in developing assessment for learning. *Assessment in Education: Principles, Policy & Practice, 16*(2), 165–184.
White, B. Y. & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition. Making science accessible to all students. *Cognition and Instruction, 16*(1), 3–118.
Wiliam, D. (2006). Formative assessment: getting the focus right. *Educational Assessment, 11*(3–4), 283–289.
Wiliam, D. (2009). An integrative summary of the research literature and implications for a new theory of formative assessment. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment*. (pp. 18–40). New York, NY: Taylor & Francis.
Windschitl, M., Thompson, J., Braaten, M. & Stroupe, D. (2012). Proposing a core set of instructional practices and tools for teachers of science. *Science Education, 96*(5), 878–903.
Windschitl, M. Thompson, J. & Braaten, M. (2011) Ambitious pedagogy by novice teachers? Who benefits from tool-supported collaborative inquiry into practice and why. *Teachers College Record, 13*(7), 1311–1360.
Wright, S. P., Horn, S. P. & Sanders, W. L. (1997). Teachers and classroom context effects on student achievement: implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*, 57–67.